

Modelado Estocástico

Clase 10: Vector Error Correction Models (VECM)

1. Cointegración

- Los modelos VAR que vimos son apropiados para modelar variables que son $I(0)$, como por ejemplo, series de tiempo de retornos de activos financieros o de tasas de crecimiento de variables económicas.
- La teoría económica a veces ofrece relaciones de equilibrio entre variables que son series de tiempo en niveles, en lugar de tasas de crecimiento, y estas series de tiempo están mejor descritas como $I(1)$. En otras ocasiones, relaciones de arbitraje entre precios de activos financieros implican relaciones entre series de tiempo que son $I(1)$.
- El concepto estadístico de cointegración es el requerido para relacionar series de tiempo y modelos VAR con datos que son $I(1)$.

2. 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, las variables) son $I(1)$ y no están cointegradas, 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 $\mathbf{y}_t = (y_{1t} \cdots y_{kt})'$ un vector de $k \times 1$ variables que no están cointegradas. Usando la partición $\mathbf{y}_t = (y_{1t}, \mathbf{y}_{2t}')$, considere la regresión por Mínimos Cuadrados Ordinarios de y_{1t} en \mathbf{y}_{2t} , dada por la ecuación estimada $y_{1t} = \hat{\beta}_2' \mathbf{y}_{2t} + e_t$. Como y_{1t} no está cointegrado

con y_{2t} , el valor verdadero de β_2 es cero. Se puede demostrar (Phillips 1986) que en este caso de regresión espuria (es decir, los residuos e_t son $I(1)$),

- $\widehat{\beta}_2$ 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 β_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>

3. Definición de Cointegración

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

La combinación lineal $\boldsymbol{\beta}'\mathbf{y}_t$ 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.

El vector de cointegración no es único: si $\beta' y_t \sim I(0)$, entonces, $c\beta' y_t \sim I(0)$, con $c \in \mathbb{R}$. Por esta razón, algún tipo de **normalización** suele hacerse, para que el vector de cointegración sea único. Una normalización que suele usarse es $\beta = (1 \quad -\beta_2 \quad -\beta_3 \quad \cdots \quad -\beta_k)'$ de modo que

$$\beta' y_t = y_{1t} - \beta_2 y_{2t} - \beta_3 y_{3t} - \cdots - \beta_k y_{kt} \sim I(0)$$

O bien, podemos escribir esta relación como

$$y_{1t} = \beta_2 y_{2t} + \beta_3 y_{3t} + \cdots + \beta_k y_{kt} + u_t$$

donde u_t suelen llamarse en la literatura con el nombre de “*disequilibrium errors*” o también como “*cointegrating residuals*”.

4. Relaciones de Cointegración Múltiples

Si el vector y_t de dimensión $(k \times 1)$, estuviera cointegrado, entonces pueden haber entre $1 \leq r < k$ vectores de cointegración que sean linealmente independientes. Note que decimos “linealmente independientes” porque si hubiera dos vectores de cointegración linealmente independientes, cualquier combinación lineal entre ellos también va a ser un vector de cointegración.

Por ejemplo, sea $k = 3$ y suponga que existen $r = 2$ vectores de cointegración $\boldsymbol{\beta}_1 = (\beta_{11} \ \beta_{12} \ \beta_{13})'$ y $\boldsymbol{\beta}_2 = (\beta_{21} \ \beta_{22} \ \beta_{23})'$. Entonces,

$$\boldsymbol{\beta}_1' \mathbf{y}_t = \beta_{11}y_{1t} + \beta_{12}y_{2t} + \beta_{13}y_{3t} \sim I(0)$$

$$\boldsymbol{\beta}_2' \mathbf{y}_t = \beta_{21}y_{1t} + \beta_{22}y_{2t} + \beta_{23}y_{3t} \sim I(0)$$

La matriz de (3×2) ,

$$B = \begin{pmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \end{pmatrix}$$

forma una base para el espacio de vectores de cointegración. Los vectores $\boldsymbol{\beta}_1$ y $\boldsymbol{\beta}_2$ no son únicos a menos que sea realice algún tipo de

normalización. Además, sean c_1 y c_2 constantes, $c_1, c_2 \in \mathbb{R}$, entonces se puede mostrar que la combinación lineal entre ellos es otro vector de cointegración, es decir, se puede mostrar que $(c_1 \beta'_1 + c_2 \beta'_2) y_t \sim I(0)$.

5. 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.

6. Metodología de Engle y Granger

Vale la pena enfatizar que los estimadores por Mínimos Cuadrados Ordinarios (MCO) son consistentes si la variable dependiente y las independientes son estacionarias. Por eso, generalmente usamos MCO bajo este supuesto. Pero también, cuando la variable dependiente y las independientes están cointegradas, MCO también va a darnos estimadores consistentes de los parámetros.

Sean y_1, y_2, \dots, y_k variables integradas de orden 1. Por ejemplo, podrían ser logaritmos de precios o tasas de interés. Elijamos una de estas

variables como la variable dependiente, digamos, y_1 y corramos una regresión por MCO:

$$y_{1t} = \beta_1 + \beta_2 y_{2t} + \beta_3 y_{3t} + \cdots + \beta_k y_{kt} + u_t \quad (*)$$

Esta regresión se llama la regresión de Engle-Granger y el test de Engle-Granger es un test de raíces unitarias sobre los residuos de esta regresión. Si los residuos de (*) son estacionarios, entonces las variables y_1, y_2, \dots, y_k están cointegradas con vector de cointegración $(1 \quad -\hat{\beta}_2 \quad \cdots \quad -\hat{\beta}_k)$.

En otras palabras,

$$Z = y_1 - \hat{\beta}_1 - \hat{\beta}_2 y_{2t} - \hat{\beta}_3 y_{3t} - \cdots - \hat{\beta}_k y_{kt}$$

es una combinación lineal de variables integradas de orden 1 que resulta ser estacionaria.

La regresión de Engle-Granger no suele usarse con mucha frecuencia porque es la única situación en la que es correcto usar MCO cuando se regresan variables integradas de orden 1. Si y_1, y_2, \dots, y_k no estuvieran cointegradas, entonces el término de error en (*) no sería estacionario y los estimadores por MCO no serían consistentes.

Pero hay otros problemas asociados a usar esta metodología: primero, si $k > 2$, el resultado del test se ve afectado por la elección de la variable dependiente. Es decir, si como variable dependiente usáramos y_2 , en lugar de y_1 , el vector de cointegración va a ser diferente. Y un segundo problema es que con k variables nos permite estimar solamente una relación de cointegración, aunque pueden haber hasta $k - 1$ relaciones de cointegración cuando tenemos k variables integradas de orden 1.

Ejemplos II.5.9 de Carol Alexander (página 232 y comentario PPP en 231)

7. Tests de Johansen-Juselius (1990)

<https://www.stata.com/manuals/tsvecrank.pdf>

Los tests de Johansen buscan cuál es la combinación lineal de las variables que es la más estacionaria mientras que el de Engle-Granger, basado en OLS, busca cuál es la combinación lineal que tiene la menor varianza.

Los tests de Johansen pueden pensarse como una generalización multivariada de los tests de raíces unitarias que vimos.

Recordemos del test de Dickey Fuller del caso univariado:

$$y_t = \rho y_{t-1} + \varepsilon_t; \quad H_0: \rho = 1, \quad H_A: \rho < 1$$

Esto es equivalente a decir que: $\Delta y_t = (\rho - 1)y_{t-1} + \varepsilon_t$

Y recordemos del ADF que en el lado derecho agregábamos rezagos:

$$\Delta y_t = (\rho - 1)y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \dots + \varepsilon_t$$

La generalización para un sistema de k variables es la siguiente: supongamos que las variables y_1, y_2, \dots, y_k las podemos representar en forma de VAR de la siguiente manera, por ejemplo, para un VAR(1):

$$\mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

Restando \mathbf{y}_{t-1} en ambos lados obtenemos

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

donde $\boldsymbol{\Pi} = \mathbf{B} - \mathbf{I}$ y donde \mathbf{I} es la matriz identidad de $k \times k$.

Análogamente al caso univariado, si agregamos más rezagos de las variables dependientes para remover la autocorrelación en los residuos, el modelo nos quedaría

$$\Delta \mathbf{y}_t = \alpha + \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \cdots + \Gamma_q \Delta \mathbf{y}_{t-q} + \varepsilon_t \quad (**)$$

El número de rezagos de las primeras diferencias se elige de modo que los residuos no estén autocorrelacionados. Ahora bien, como cada una de las variables endógenas y_1, y_2, \dots, y_k es $I(1)$, cada ecuación en (**) tiene una variable dependiente que es $I(0)$ -al tomar primeras diferencias- y consecuentemente el lado derecho de (**) también debe ser estacionario. Y por lo tanto, $\Pi \mathbf{y}_{t-1}$ tiene que ser estacionario.

Si el $\text{rango}(\mathbf{\Pi}) = 0$, entonces la condición que $\mathbf{\Pi}y_{t-1}$ debe ser estacionario no implica nada acerca de las relaciones entre y_1, y_2, \dots, y_k y consecuentemente no hay cointegración.

Ahora si el $\text{rango}(\mathbf{\Pi}) = r > 0$, entonces hay r relaciones linealmente independientes entre y_1, y_2, \dots, y_k que son estacionarias y consecuentemente, las variables van a estar cointegradas.

De modo que el test de cointegración va a ser un test sobre el rango de la matriz $\mathbf{\Pi}$, en donde el rango de esta matriz indica el número de relaciones o vectores de cointegración.

Si hubiera r vectores de cointegración entre y_1, y_2, \dots, y_k , entonces el $\text{rango}(\mathbf{\Pi}) = r$ porque recordemos que el rango de una matriz es igual al número de sus raíces características no nulas.

El procedimiento de Johansen es entonces un test sobre el número de autovalores de la matriz $\mathbf{\Pi}$ que no son cero. Hay dos tests de cointegración que van a encontrar en los paquetes econométricos. Por un lado, Johansen y Juselius (1990) recomiendan usar el **test de la traza**: Bajo H_0 , el número de relaciones de cointegración, r , es a lo sumo R .

$$H_0: r \leq R$$

$$H_A: r > R$$

El estadístico es $Stat_{trace} = -T \sum_{i=R+1}^k \ln(1 - \lambda_i)$ donde T es el tamaño de la muestra, k es la cantidad de variables endógenas en el sistema y los autovalores de la matriz $\mathbf{\Pi}$ están ordenados de modo tal que $1 > \lambda_1 > \dots > \lambda_k \geq 0$, para asegurarse que $Stat_{trace}$ crezca con el número de autovalores que no son cero. Los valores críticos de este test

son provistos por varios software (depende del número de rezagos que se incluyan y si tiene o no constante y tendencia).

El otro test se llama test de máximo autovalor (*maximal eigenvalue test*), la hipótesis nula es la misma y la alternativa es que el número de relaciones de cointegración es $R+1$, es decir, $H_A: r = R + 1$. El estadístico es $Stat_{max} = -T \ln(1 - \lambda_{R+1})$.

Caso Bivariado:

$$\mathbf{\Pi} = \mathbf{B} - \mathbf{I} = \begin{pmatrix} b_{11} - 1 & b_{12} \\ b_{21} & b_{22} - 1 \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}$$

Si $\text{rango}(\mathbf{\Pi}) = 0$, $\Delta \mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t$ es un VAR en primeras diferencias, de modo que cada una de las variables en el sistema es un *random walk con drift* y no va a existir una combinación lineal de las variables que sea estacionaria.

Si $\text{rango}(\mathbf{\Pi}) = 2$, (rango completo) entonces como $\Delta \mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, la solución de largo plazo va a estar dada por las dos ecuaciones independientes:

$$\pi_{11}y_{1t} + \pi_{12}y_{2t} = 0$$

$$\pi_{21}y_{1t} + \pi_{22}y_{2t} = 0$$

En este caso, cada una de las dos variables (“ k ” variables en el caso general) contenidas en el vector \mathbf{y}_t debe ser estacionaria, y consecuentemente no hay relación de cointegración. Para que haya cointegración, la matriz $\mathbf{\Pi}$ no puede tener rango completo.

En los casos intermedios en los que si $\text{rango}(\mathbf{\Pi}) = r < k$, existen r vectores de cointegración.

Con r ecuaciones independientes y k variables, existen $k - r$ tendencias estocásticas.

El procedimiento de Johansen es más informativo que el procedimiento de Engle-Granger porque permite encontrar todas las relaciones de cointegración en el sistema. Suele utilizarse cuando hay muchas variables en un sistema y cuando no está claro cuál debiera ser la variable dependiente.

Ejemplo Carol Alexander II.5.10: cointegración en tasas de interés de UK. Usando un nivel de significancia del 1% se rechaza la hipótesis nula para $R=1$, $R=2$ y $R=3$. Por lo tanto, hay 4 vectores de cointegración. Como el número de variables es 6, a lo sumo pudo haber habido 5 relaciones de cointegración. De manera que podemos decir que en este sistema hay un alto grado de cointegración.

Estos son tests **secuenciales**. Miramos $R=0$, luego $R=1$, después $R=2$,...

Johansen tests for cointegration						
Trend: constant			Number of obs = 1898			
Sample: 3 - 1900			Lags = 2			
maximum				trace	5% critical	1% critical
rank	parms	LL	eigenvalue	statistic	value	value
0	42	34216.516		1023.0924	94.15	103.18
1	53	34496.208	0.25526	463.7089	68.52	76.07
2	62	34630.074	0.13156	195.9773	47.21	54.46
3	69	34680.18	0.05143	95.7644	29.68	35.65
4	74	34719.711	0.04080	16.7019*1	15.41	20.04
5	77	34726.44	0.00707	3.2448*5	3.76	6.65
6	78	34728.062	0.00171			
maximum				max	5% critical	1% critical
rank	parms	LL	eigenvalue	statistic	value	value
0	42	34216.516		559.3835	39.37	45.10
1	53	34496.208	0.25526	267.7316	33.46	38.77
2	62	34630.074	0.13156	100.2129	27.07	32.24
3	69	34680.18	0.05143	79.0625	20.97	25.52
4	74	34719.711	0.04080	13.4570	14.07	18.63
5	77	34726.44	0.00707	3.2448	3.76	6.65
6	78	34728.062	0.00171			

Comando:

vecrank mth1 mth2 mth3 mth6 mth9 mth12, trend(constant) max levela

8. Modelos de Corrección de Errores

Esta es la segunda parte del análisis de cointegración. La primera parte consistía en encontrar relaciones de largo plazo. Los modelos de corrección de errores proveen un análisis de la dinámica entre las variables en el corto plazo. Por ejemplo, si los logaritmos de los precios de dos activos estuvieran cointegrados, el correspondiente modelo de corrección de errores va a consistir en un modelo que explica la dinámica entre los retornos logarítmicos de dichos activos.

Supongamos que tenemos dos series de logaritmos de precios, X e Y; el modelo de corrección de errores sería (***):

$$\Delta X_t = c_1 + \sum_{i=1}^m \beta_{11}^i \Delta X_{t-i} + \sum_{i=1}^m \beta_{12}^i \Delta Y_{t-i} + \gamma_1 Z_{t-1} + \varepsilon_{1t}$$

$$\Delta Y_t = c_2 + \sum_{i=1}^m \beta_{21}^i \Delta X_{t-i} + \sum_{i=1}^m \beta_{22}^i \Delta Y_{t-i} + \gamma_2 Z_{t-1} + \varepsilon_{2t}$$

donde $Z = X - \alpha Y$ y los coeficientes se determinan por regresiones usando MCO. ¿Por qué se llama modelo de corrección de errores? Suponga que $\alpha > 0$. Entonces este modelo tendrá un mecanismo de corrección de errores si $\gamma_1 < 0$ y $\gamma_2 > 0$.

Para ver esto, suponga que Z fuera grande y positivo. Entonces X va a caer porque $\gamma_1 < 0$ e Y va a aumentar porque $\gamma_2 > 0$. Ambos efectos producirán una reducción de Z y de esta manera, los errores se corrigen. Suponga en cambio que Z fuera grande y negativo: X va a aumentar porque $\gamma_1 < 0$ e Y va a caer porque $\gamma_2 > 0$, produciéndose un aumento de Z y de esta manera, los errores también se corrigen.

La magnitud de los estimadores de γ_1 y de γ_2 representarán la velocidad de ajuste hacia el restablecimiento del equilibrio de largo plazo.

Ejercicio: ejemplo II.5.11 de Carol Alexander

9. Más sobre VECM:

En primer lugar, vamos a mostrar que un VAR se puede transformar en un VECM. Sea un VAR(2):

$$\mathbf{y}_t = \Phi \mathbf{D}_t + \Pi_1 \mathbf{y}_{t-1} + \Pi_2 \mathbf{y}_{t-2} + \varepsilon_t$$

\mathbf{D}_t contiene la constante y una tendencia determinística (si quisiéramos).
Restemos a ambos lados del igual \mathbf{y}_{t-1} :

$$\mathbf{y}_t - \mathbf{y}_{t-1} = \Phi \mathbf{D}_t + (\Pi_1 - \mathbf{I}) \mathbf{y}_{t-1} + \Pi_2 \mathbf{y}_{t-2} + \varepsilon_t$$

Ahora sumemos y restemos $\Pi_2 \mathbf{y}_{t-1}$ en el lado derecho,

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + (\Pi_1 + \Pi_2 - \mathbf{I}) \mathbf{y}_{t-1} - \Pi_2 \mathbf{y}_{t-1} + \Pi_2 \mathbf{y}_{t-2} + \varepsilon_t$$

Reescribiendo, $\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + (\Pi_1 + \Pi_2 - \mathbf{I}) \mathbf{y}_{t-1} - \Pi_2 \Delta \mathbf{y}_{t-1} + \varepsilon_t$

Llamando $\Pi = \Pi_1 + \Pi_2 - \mathbf{I}$ y $\Gamma_1 = -\Pi_2$,

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \varepsilon_t$$

Hemos logrado reescribir el VAR(2) como un VECM.

En segundo lugar, vamos a mostrar que un VAR(3) se puede transformar en un VECM. Supongamos que tenemos un VAR(3):

$$\mathbf{y}_t = \Phi \mathbf{D}_t + \Pi_1 \mathbf{y}_{t-1} + \Pi_2 \mathbf{y}_{t-2} + \Pi_3 \mathbf{y}_{t-3} + \varepsilon_t$$

Restemos a ambos lados del igual \mathbf{y}_{t-1} y sumemos y restemos $(\Pi_2 + \Pi_3) \mathbf{y}_{t-1}$ en el lado derecho:

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + (\Pi_1 + \Pi_2 + \Pi_3 - \mathbf{I}) \mathbf{y}_{t-1} - \Pi_2 \Delta \mathbf{y}_{t-1} - \Pi_3 (\mathbf{y}_{t-1} - \mathbf{y}_{t-3}) + \varepsilon_t$$

Ahora sumemos y restemos $\Pi_3 \mathbf{y}_{t-2}$ en ambos lados:

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + \Pi \mathbf{y}_{t-1} - \Pi_2 \Delta \mathbf{y}_{t-1} - \Pi_3 (\mathbf{y}_{t-2} - \mathbf{y}_{t-3}) - \Pi_3 (\mathbf{y}_{t-1} - \mathbf{y}_{t-2}) + \varepsilon_t$$

donde $\Pi = \Pi_1 + \Pi_2 + \Pi_3 - \mathbf{I}$

Por lo tanto, logramos reescribir el VAR(3) como un VECM:

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \Gamma_2 \Delta \mathbf{y}_{t-2} + \varepsilon_t$$

donde $\Pi = \Pi_1 + \Pi_2 + \Pi_3 - I$, $\Gamma_1 = -\Pi_2 - \Pi_3$, y $\Gamma_2 = -\Pi_3$

Generalizando, si partimos de un VAR(p):

$$\mathbf{y}_t = \Phi \mathbf{D}_t + \Pi_1 \mathbf{y}_{t-1} + \dots + \Pi_p \mathbf{y}_{t-p} + \varepsilon_t \quad (\#1)$$

Recordemos que el VAR(p) es estable si todas las raíces de

$$\det(\mathbf{I}_k - \Pi_1 z - \dots - \Pi_p z^p) = 0$$

están afuera del círculo unitario. Si alguna de las raíces estuviera sobre el círculo unitario, entonces alguna o todas las variables en \mathbf{y}_t serían I(1) y podrían estar cointegradas. El VAR(p) se puede transformar a un VECM:

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{y}_{t-p+1} + \varepsilon_t \quad (\#2)$$

donde $\Pi = \Pi_1 + \dots + \Pi_p - \mathbf{I}_k$, $\Gamma_i = -\sum_{j=i+1}^p \Pi_j$, $i = 1, \dots, p-1$

Π = Matriz de Impacto de Largo Plazo; Γ_i = Matrices de Impacto de Corto Plazo

El teorema de representación de Granger dice que la representación VAR no es la indicada cuando las variables están cointegradas. En estos casos, debe utilizarse la representación VECM.

Notemos que los parámetros Π_i del VAR se pueden recuperar a partir de los parámetros Π y Γ_i del VECM:

$$\Pi_1 = \Gamma_1 + \Pi + I_k$$

$$\Pi_i = \Gamma_i - \Gamma_{i-1}, i = 2, \dots, p$$

En el VECM, $\Delta \mathbf{y}_t$ y sus lags son $I(0)$. El único término que potencialmente incluye términos $I(1)$ es $\Pi \mathbf{y}_{t-1}$, pero como el lado izquierdo y los demás términos en el lado derecho son $I(0)$, $\Pi \mathbf{y}_{t-1}$ tiene que ser $I(0)$.

Entonces, si $\text{rango}(\Pi) = 0$, (#2) se reduce a un VAR(p-1):

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + \Gamma_1 \Delta \mathbf{y}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{y}_{t-p+1} + \varepsilon_t$$

Si $0 < \text{rango}(\Pi) = R < k$, entonces \mathbf{y}_t es $I(1)$ con R vectores de cointegración linealmente independientes y $k - R$ tendencias estocásticas comunes (raíces unitarias). Como Π tiene rango R , se puede demostrar que es posible escribir

$$\Pi = \alpha\beta'$$

donde α y β son matrices de dimensión $k \times R$ (Π es de dimensión $k \times k$)

Las filas de β' forman una base de los R vectores de cointegración y los elementos de α distribuyen el impacto de los vectores de cointegración sobre la evolución de $\Delta\mathbf{y}_t$. El VECM (#2) nos queda

$$\Delta\mathbf{y}_t = \Phi\mathbf{D}_t + \alpha\beta'\mathbf{y}_{t-1} + \Gamma_1\Delta\mathbf{y}_{t-1} + \cdots + \Gamma_{p-1}\Delta\mathbf{y}_{t-p+1} + \varepsilon_t \quad (\#3)$$

donde $\beta'\mathbf{y}_{t-1} \sim I(0)$ ya que β' es la matriz de vectores de cointegración.

10. Pronósticos en Stata

Podemos generar predicciones/pronósticos (forecasts) en modelos ARIMA o modelos VAR o VEC.

Modelos ARIMA:

```
arima dlws, ar(1)           // Para entender bien qué hace Stata más  
adelante, vamos a estimar un AR(1). Esto puede generalizarse para cualquier ARIMA  
estimates store mi_ar_1     // Con el comando estimates store Stata  
guarda la estimación previa con el nombre mi_ar_1
```

`forecast create mi_arma, replace` // Para hacer pronósticos,
debo empezar con creando el forecast con el comando `forecast create`, y lo
voy a llamar `mi_arma`

`forecast estimates mi_ar_1` // Agrego la información de
la regresión previa al forecast que acabo de crear

`forecast solve, begin(m(2019m1)) end(m(2021m9))` // Con este
comando Stata nos calcula los forecasts, es decir, las predicciones a futuro de la(s)
variable

Sea y_t un proceso AR(1), el cual podemos representar de la siguiente manera

$$y_t = c + \rho y_{t-1} + \varepsilon_t$$

Su media será $\mu = E(y_t) = \frac{c}{1-\rho}$, y a este proceso AR(1) podemos reescribirlo como

$$y_t - \mu = \rho(y_{t-1} - \mu) + \varepsilon_t$$

De modo que,

$$(1 - \rho L)(y_t - \mu) = \varepsilon_t$$

Sea $\rho(L) = \frac{1}{1-\rho L}$,

$$\rho(L)^{-1}(y_t - \mu) = \varepsilon_t$$

Note que para el AR(1), $\rho(L) = \frac{1}{1-\rho L} = 1 + \rho L + \rho^2 L^2 + \rho^3 L^3 + \dots$

Caso General: note que en el caso general, (no necesariamente AR(1)),

$$\frac{\rho(L)}{L^s} = L^{-s} + \rho_1 L^{-s+1} + \rho_2 L^{-s+2} + \rho_{s-1} L^{-1} + \rho_s L^0 + \rho_{s+1} L^1 + \rho_{s+2} L^2 + \dots$$

Definimos el *annihilation operator* (operador de eliminación) como:

$$\left[\frac{\rho(L)}{L^s} \right]_+ = \rho_s + \rho_{s+1} L^1 + \rho_{s+2} L^2 + \dots$$

Este operador reemplaza las potencias negativas de L por ceros.

Si queremos pronosticar para el período de $t+s$ con información hasta el período t , el pronóstico óptimo es

$$\hat{E}(y_{t+s} | \varepsilon_t, \varepsilon_{t-1}, \dots) = \mu + \left[\frac{\rho(L)}{L^s} \right]_+ \varepsilon_t$$

Volvemos al AR(1): en este caso, $\rho(L) = \frac{1}{1-\rho L} = 1 + \rho L + \rho^2 L^2 + \rho^3 L^3 + \dots$

Y además,

$$\left[\frac{\rho(L)}{L^s} \right]_+ = \rho^s + \rho^{s+1} L^1 + \rho^{s+2} L^2 + \dots = \frac{\rho^s}{1 - \rho L}$$

Consecuentemente,

$$\hat{E}(y_{t+s} | y_t, y_{t-1}, \dots) = \mu + \frac{\rho^s}{1 - \rho L} (1 - \rho L)(y_t - \mu) = \mu + \rho^s (y_t - \mu)$$

Esto es lo que calcula Stata cuando ejecutamos `forecast solve`.

11. Apéndice: Descomposición de Cholesky

Sea **A** una matriz simétrica definida positiva.

Entonces, existe una matriz triangular inferior **L** tal que

$$\mathbf{A} = \mathbf{L}\mathbf{L}'$$

Ejemplo: sea $A = \begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix}$

El determinante de esta matriz es 36 y notar que todos los *leading principal minors* son positivos, y consecuentemente, la matriz es definida positiva.

La descomposición de Cholesky nos asegura que existe una matriz triangular inferior L tal que $\mathbf{A}=\mathbf{L}\mathbf{L}'$. En este caso, pueden verificar que esta matriz es:

$$L = \begin{pmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ -8 & 5 & 3 \end{pmatrix}$$

Referencias

- Elliott, G., Rothenberg, T. J., & Stock, J. H. (1992). Efficient tests for an autoregressive unit root.
- Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, 2(2), 111-120.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3), 231-254.
- Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and statistics*, 52(2), 169-210.

Manual de Stata

IRF:

<https://www.stata.com/manuals/tsirfcreate.pdf>

<https://www.stata.com/manuals/tsirf.pdf>

VEC:

<https://www.stata.com/manuals13/tsvecintro.pdf>