

Modelos multivariados para procesos estacionarios

Modelos de vectores autoregresivos (VAR)

Fernando Borraz

dECON - FCS - Udelar - 2019

Introducción

- Contamos con una serie de observaciones repetidas en el tiempo de una variable aleatoria y_t (PIB, inflación, M, etc)
- Una serie de tiempo (ST) es un conjunto de observaciones de la misma variable que son consideradas como una sucesión de variables aleatorias. $\{y_t\} = \{y_1, y_2, \dots, y_T\}$ o $\{y_t\}_1^T$
- Consideramos a c/u de las y_t como variable aleatoria. Por ej.
 $\text{PIB}_{2018} = \text{USD } 60000 \text{ millones}$ es solo una realización entre todas las posibles

Introducción

- ¿Por qué no usar el instrumental de econometría clásica que conocemos en un modelo de ST?
- El problema es que la ST que observamos es solo UNA muestra de y_t . La colección de datos observada solo es una posible realización de un proceso estocástico
- Si por ej. tenemos N-coleccciones de $\{y_t\}$ podríamos calcular los momentos de la densidad de y_t (media, varianza, etc)
- Cantidad de parámetros a estimar: T medias, T varianzas y $\frac{T * T - T}{2}$ covarianzas (si $T = 50$ tenemos $50 + 50 + \frac{50^2 - 50}{2} = 1325$ parámetros!!)
- Obviamente son muchos parámetros y no podemos calcularlos en base a una muestra de tamaño T (que es lo que disponemos)

Introducción

- Esto nos fuerza a usar modelos paramétricos y realizar supuestos. Así limitamos la cantidad de parámetros a estimar
- Esto es una limitante de los modelos de ST
- Así limitamos la cantidad de parámetros a estimar
- Un concepto clave será estacionariedad. La estacionariedad de una serie influye fuertemente sobre su evolución y propiedades, por ej. la persistencia de los shocks será infinita si el proceso es no estacionario
- Si las variables son no estacionarias, entonces se puede probar que los supuestos tradicionales del análisis asintótico no son válidos. En otras palabras, los “ratios-t” no tienen una distribución-t, por lo que no podemos realizar pruebas de hipótesis válidos sobre los parámetros de la regresión (idém ratio-F)

Introducción

- 2 tipos de variables: estacionarias y no estacionarias
- 2 tipos de modelos: univariados y multivariados

		Modelos	
		<i>Univariados</i>	<i>Multivariados</i>
Procesos	<i>Estacionarios</i>	ARMA	VAR
	<i>No estacionarios</i>	Raíces unitarias, ARCH, switching	Cointegración, VECM

Introducción

- Razones para desagregar una serie de tiempo económica en sus componentes inobservables (Espasa y Cancelo 1993)
- “El análisis de la coyuntura de un fenómeno económico normalmente no se realiza directamente a través de la evolución de sus datos originales, ya que éstos contienen muchas oscilaciones de escaso interés económico que pueden llevar a conclusiones equivocadas. Por el contrario, dicho análisis se efectúa a través de señales extraídas de los datos, que sirven de estimación de lo que un analista puede considerar como un aspecto esencial del fenómeno en cuestión.”

Introducción

- Los datos originales de una serie económica presentan grandes oscilaciones, muchas de las cuales no son de interés desde el punto de vista del análisis económico. Así, los aspectos esenciales de un fenómeno económico no son directamente observables, con lo cual es necesario disponer de procedimientos que permita cuantificarlos a partir de la información disponible (los datos originales)
- Al analista corresponde definir cuál o cuáles componentes o señales son de interés

Componentes inobservables de una serie económica

- $y_t = I_t + c_t + s_t + \varepsilon_t$
- i) Tendencia (I_t): evolución subyacente de la serie o de largo plazo. Es la evolución firme que hay detrás de la trayectoria
- ii) Estacionalidad (s_t): oscilaciones sistemáticas a lo largo del año. Tiene periodicidad de un múltiplo del año. Son oscilaciones regulares y sistemáticas dentro del año
- iii) Ciclo (c_t): desviaciones sistemáticas que no son estacionales. Son menos sistemáticas y pueden durar entre 2 y 5 años
- iv) Irregular(ε_t): shock aleatorio. Son ruidos de corto plazo. Son no sistemáticas (ruido blanco)

Introducción

- Otra distinción relevante
 - trabajamos sobre primeros momentos (medias)
 - o sobre segundos momentos (varianzas, modelos ARCH)
- Es decir, la series podrán ser no estacionarias en media y/o en varianza

Introducción

- Los procesos de series de tiempo univariados si bien son útiles para realizar predicciones son limitados para entender el funcionamiento de la economía en general pues excluyen del análisis la interacción entre variables económicas
- El instrumental de modelos ARMA puede ser extendido directamente para considerar modelos multivariados
- Esto nos permite incorporar al análisis que las variables económicas se encuentran relacionadas
- Estudiaremos modelos de series de tiempo multivariados para procesos estacionarios, en particular analizaremos los modelos de vectores autoregresivos (VAR)
- Los VAR son una generalización de los modelos AR al caso multivariado (Sims 1980)

Introducción

- Los VAR parten de suponer que no es justificable la separación entre variables endógenas y exógenas que se hace en los modelos de ecuaciones simultáneas (Klein, econometría tradicional). Por ejemplo:

$$\text{Demanda: } Q_D = \alpha_1 P + \alpha_2 Y + \varepsilon_D$$

$$\text{Oferta: } Q_S = \beta_1 P + \varepsilon_S$$

$$\text{Equilibrio: } Q_D = Q_S$$

- Para Sims los datos deben decidir si las variables son endógenas o exógenas y por lo tanto todas las variables se tratan como endógenas. Rechaza la imposición de ceros en los coeficientes de las ecuaciones estructurales (por ej. $\beta_2 = 0$ en Q_S)
- Sims rechaza usar la teoría económica como guía para la modelización. Por esto a los VAR se los llama modelos ateóricos

Introducción

- Estudiaremos los VAR y no los vectores de medias móviles (VMA). Dos justificaciones: i) los modelos VMA son altamente no lineales y difíciles de estimar e interpretar y ii) todo $MA(q)$ invertible puede ser formulado como un $AR(\infty)$, por lo que usando un VAR con un número importante de rezagos se puede aproximar bien un VMA
- Utilidad de los VAR:
 - i) Predicción
 - ii) Causalidad en el sentido de Granger
 - iii) Funciones impulso respuesta (IRF)
 - iv) Descomposición de varianzas (VD)
 - v) Prueba de hipótesis de modelos lineales de expectativas racionales
 - vi) Otros

Estacionariedad

- Como no tenemos varias realizaciones de los procesos juega un rol clave el concepto de estacionariedad (\neq estacionalidad)
- 2 definiciones de estacionariedad (débil, estricta)
- y_t estacionariedad débil (o de segundo orden o en covarianzas) si:
 - 1) $E(y_t) = \mu \quad t = 1, 2, \dots, \infty$
 - 2) $V(y_t) = E[(y_t - \mu)(y_t - \mu)] = \sigma^2 < \infty$ finita y no depende de t
 - 3) $E[(y_t - \mu)(y_{t-s} - \mu)] = \gamma_{t-s-t} = \gamma_s = \gamma_{-s}$ para todo $t, t-s$ ($s \geq 1$). La COV solo depende de la diferencia entre t y $t-s$ y no directamente de t

Estacionariedad débil

- Si el proceso es estacionario débil, todas las medias y varianzas son iguales y todas las covarianzas dependen únicamente de la diferencia entre t y $t-s$
- Las covarianzas γ_s suelen denominarse autocovarianzas
- Como las covarianzas dependen de la distancia " s " (entre 2 observaciones) y no directamente de $t \Rightarrow$ podemos hacer una "muestra" con los datos de solo un universo
- Series económicas estacionarias?
- Como las autocov γ_s dependen de las unidades de medida de y_t es más conveniente trabajar con las autocorrelaciones:
 $\tau_s = \frac{\gamma_s}{\gamma_0}, s = 0, 1, 2, \dots$. Si graficamos τ_s contra s obtenemos la función de autocorrelación o correlograma (ACF)

Estacionariedad estricta

- y_t estacionario estricto si la función de probabilidades conjuntas de las realizaciones de y_t no depende del tiempo
- Esto es: $P[y_{t1} \leq b_1, y_{t2} \leq b_2, \dots, y_{tn} \leq b_n] = P[y_{t1+s} \leq b_1, y_{t2+s} \leq b_2, \dots, y_{tn+s} \leq b_n]$
- Es decir, la probabilidad medida por la secuencia $\{y_t\}$ es la misma que aquella medida por $\{y_{t+s}\}$
- La probabilidad que y pertenezca a cierto intervalo es la misma independientemente en que momento (pasado o futuro) la calculamos
- Ej. la probabilidad que la inflación sea menor a cierto valor por ej 5% es la misma entre y_{2016}, y_{2015} que entre y_{1901}, y_{1900}

Estacionariedad

- Notar que estacionariedad estricta \nRightarrow estacionariedad débil pues los segundos momentos pueden no existir
- Ej. distribución Cauchy, $f(y) = \frac{1}{\pi(1+y^2)}$, varianza de y_t es ∞
- Estacionariedad estricta + segundos momentos finitos \Rightarrow estacionariedad débil
- Si $y_t \sim \text{normal}$, estacionariedad estricta \Leftrightarrow estacionariedad débil
- En el curso asumiremos el supuesto de normalidad por lo que simplemente nos referiremos a estacionariedad (que en este caso es débil o estricta)
- ¿Cómo sabemos si la serie y_t es estacionaria? Es un tema fundamental y lo abordaremos en la sección de raíces unitarias. Por el momento utilizaremos análisis gráfico y el correlograma para determinar si las series son estacionarias

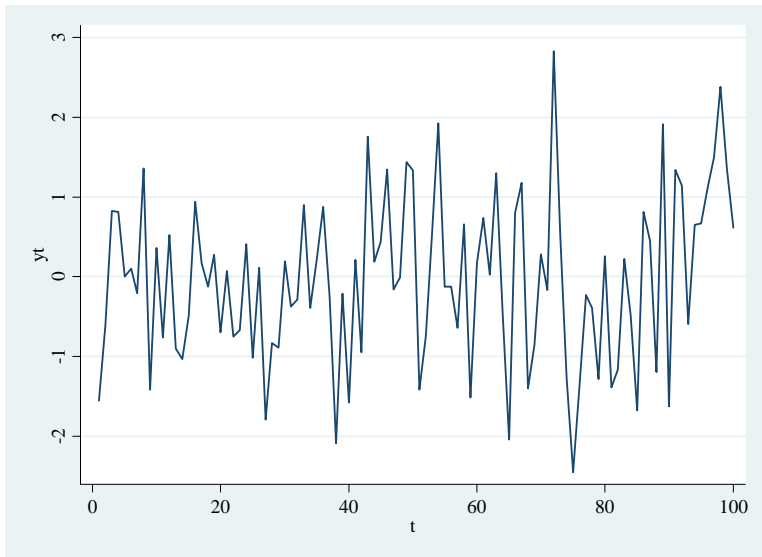
Ergodicidad

- Una serie y_t estacionaria es ergódica en la media si:
 - $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T y_t = E(y_t)$
- Esto es, la media muestral converge a la media del proceso cuando la muestra crece
- En la práctica se requiere que las autocov tiendan a 0
- Intuición. Cuando $s \rightarrow \infty$ el pasado remoto no debe contener información útil para las predicciones actuales del proceso
- Ergodicidad y estacionariedad no siempre son equivalentes
- Un proceso puede ser estacionario pero no ergódico

Ruido blanco

- Un proceso ruido blanco se caracteriza por no tener estructura discernible
- y_t ruido blanco si:
 - 1) $E(y_t) = 0 \quad t = 1, 2, \dots, \infty$
 - 2) $V(y_t) = \sigma^2 < \infty$ finita y no depende de t
 - 3) $E[(y_t - \mu)(y_{t-s} - \mu)] = 0 \quad (s \geq 1)$
- Condicional en la información disponible en $t-1$ no es posible predecir el valor futuro del proceso $\left(\hat{y}_{t+1|t} = 0\right)$
- ¿Un proceso ruido blanco es estacionario?

Proceso de ruido blanco: ejemplo



Pruebas empíricas de raíces unitarias

- Existe un gran variedad de pruebas de raíces unitarias, que difieren en su enfoque, poder y robustez frente a desvíos en las condiciones supuestas
- Considerando la regresión $y_t = \phi y_{t-1} + \varepsilon_t$, la hipótesis nula de raíz unitaria será $H_0) \phi = 1$
- La prueba t exige que tomemos el parámetro estimado, restemos 1 y dividir por la desviación estándar
- Dado que las series económicas son no explosivas no tiene sentido probar $|\phi| > 1$, por lo que es una prueba a una cola
- Probaremos
 - $H_0) |\phi| = 1$ y el proceso no es estacionario
 - $H_1) |\phi| < 1$ y el proceso es estacionario

Prueba de Dickey-Fuller

- Podemos combinar las hipótesis de Trend Stationary y Difference Stationary en un solo modelo con varias hipótesis alternativas, dependiendo del tipo de proceso que caracteriza a la variable cuando esta no tiene una raíz unitaria

- (a) $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$
- (b) $\Delta y_t = \alpha + \gamma y_{t-1} + \varepsilon_t$
- (c) $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t$

- Como todo análisis en econometría se debe partir de lo general a lo particular. Por lo tanto se empieza con la hipótesis más compleja que es (c). Si la tendencia determinística es 0, entonces se realiza el test para el caso (b) en el que solo hay una constante y si esta es 0 se hace el test (a)

Prueba de Dickey-Fuller

- 1. Se empieza por el caso general con tendencia y constante, si se rechaza H_0) \Rightarrow no hay dudas que y_t es estacionario
- 2. Si no se rechaza se va al caso sin tendencia y de nuevo si rechaza H_0) $\Rightarrow y_t$ es estacionario
- 3. Si no se rechaza se va al caso sin constante y de nuevo si rechaza H_0) $\Rightarrow y_t$ es estacionario y sino se rechaza entonces concluimos que y_t no es estacionaria
- Notar que constante y tendencia hacen más difícil de rechazar H_0), por lo tanto si se rechaza H_0), cuando se incluye constante y tendencia entonces también se rechazará cuando no se incluyan C y t
- Notar que como son ecuaciones diferentes, la distribución del estimador de γ va a cambiar en cada tipo de prueba, y por lo tanto los valores críticos son diferentes

Prueba de Dickey-Fuller

- La prueba se basa en la hipótesis nula que $\gamma = 0$, contra $H_1) \gamma < 0$, pues no tiene sentido suponer que el efecto de los shocks en las series económicas crece en el tiempo
- Notar que como en la hipótesis nula existe una raíz unitaria, entonces las pruebas t y F no son válidas
- Dickey Fuller (1979) realizan simulaciones para diferentes tamaños de muestra y obtienen valores críticos, que son menores que los de la t y la F
- Estos valores críticos son válidos ante la presencia de heterocedasticidad y distribución no normal de los errores
- Sin embargo no son válidos bajo autocorrelación de los residuos. Se supuso que la serie sigue solo un AR(1), si la serie presenta correlación a ordenes mayores entonces esta prueba no es válida y se aplica la prueba de Dicker-Fuller aumentada (ADF)

Prueba de Dickey-Fuller aumentada (ADF)

- Igual a la prueba Dickey-Fuller pero controla por autocorrelación de los residuos que en dicho caso se suponían ruido blanco
- Supongamos el caso más simple en que los residuos siguen un AR(1):
 $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ con u_t ruido blanco

- Transformación de Cochrane-Orcutt para $y_t = \alpha + \phi y_{t-1} + \varepsilon_t$

- $y_t - \rho y_{t-1} = \alpha + \phi y_{t-1} + \varepsilon_t - \rho\phi y_{t-2} - \rho\varepsilon_{t-1} \Rightarrow y_t - \rho y_{t-1} = \alpha(1 - \rho) + \phi y_{t-1} - \rho\phi y_{t-2} + u_t$

pues $e_t = \varepsilon_t - \rho\varepsilon_{t-1}$

- Resto y_{t-1} de ambos lados $\Rightarrow y_t - \rho y_{t-1} - y_{t-1} = \alpha(1 - \rho) + \phi y_{t-1} - \rho\phi y_{t-2} + u_t - y_{t-1}$

$$\Rightarrow \Delta y_t = \alpha(1 - \rho) + (\phi - 1 + \rho)y_{t-1} - \rho\phi y_{t-2} + u_t$$

Prueba de Dickey-Fuller aumentada

- $\Delta y_t = \alpha (1 - \rho) + (\phi - 1 + \rho) y_{t-1} - \rho \phi y_{t-2} + u_t$
sumo y resto $\rho \phi y_{t-1}$ del lado derecho
- $\Delta y_t = \alpha (1 - \rho) + (\phi - 1 + \rho + \rho \phi - \rho \phi) y_{t-1} - \rho \phi y_{t-2} + u_t$
- $\Delta y_t = \alpha (1 - \rho) + (\phi - 1 + \rho - \rho \phi) y_{t-1} + \rho \phi (y_{t-1} - y_{t-2}) + u_t$
- $\Delta y_t = \alpha (1 - \rho) + (\phi - 1 + \rho - \rho \phi) y_{t-1} + \rho \phi \Delta y_{t-1} + u_t$
- $\Rightarrow \boxed{\Delta y_t = \alpha_0 + \gamma y_{t-1} + \gamma_1 \Delta y_{t-1} + u_t}$
- con $\alpha_0 = \alpha (1 - \rho)$, $\gamma = (\phi - 1 + \rho - \rho \phi)$ y $\gamma_1 = \rho \phi$
- Notar $\phi = 1 \Rightarrow \gamma = 0$

Prueba de Dickey-Fuller aumentada

- Esto se puede generalizar incluyendo p rezagos de Δy_t si ε_t sigue un $AR(p)$:

$$\begin{aligned} \text{(a)} \quad \Delta y_t &= \gamma y_{t-1} + u_t + \sum_{p=1}^P \gamma_p \Delta y_{t-p} \\ \text{(b)} \quad \Delta y_t &= \alpha + \gamma y_{t-1} + u_t + \sum_{p=1}^P \gamma_p \Delta y_{t-p} \\ \text{(c)} \quad \Delta y_t &= \alpha + \beta t + \gamma y_{t-1} + u_t + \sum_{p=1}^P \gamma_p \Delta y_{t-p} \end{aligned} \tag{8}$$

$H_0) \mid \gamma \mid = 0$ y el proceso no es estacionario

$H_1) \mid \gamma \mid \neq 0$ y el proceso es estacionario

- Nuevamente la distribución del test es diferente a la de la prueba DF por la presencia de los γ_p
- Notar que como en otras ocasiones en econometría incluimos rezagos de la variable endógena para controlar autocorrelación serial de los residuos

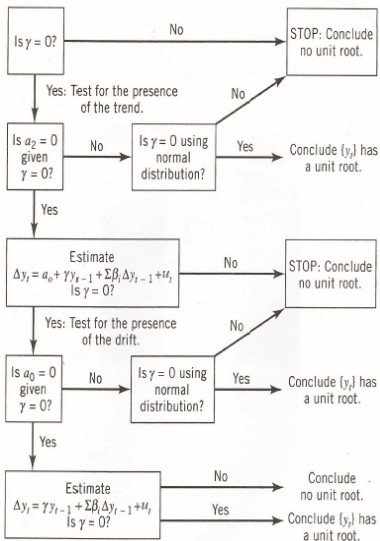
Prueba de Dickey-Fuller aumentada: implementación

- Se debe definir el número de rezagos p . Si $p=0$ tenemos la prueba DF. Definir p según criterio de información de tipo Akaike
- De nuevo tenemos que probar incluyendo primero con constante y tendencia
- Finalmente hay que notar que los valores críticos reportados solo son válidos si se aplican directamente a la series de datos y no a una estimación de los mismos
- Es posible que la serie tenga más de una raíz unitaria por lo que si procedemos de lo general a lo particular deberíamos empezar con H_0) 2 raíces unitarias, $y_t \sim I(2)$, H_1) 1 raíz unitaria ($y_t \sim I(1)$)

$$\Delta^2 y_t = \varphi \Delta y_{t-1} + \varepsilon_t + \sum_{p=1}^P \gamma_p \Delta^2 y_{t-p}$$

Prueba ADF

Estimate $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum \beta_i \Delta y_{t-i} + u_t$



Prueba de Dickey-Fuller aumentada

- 1. Empezar por el modelo más general con constante y tendencia y probar $\gamma = 0$. Si rechazo H_0 listo. No hay raíz unitaria. Pero si no rechazo puede ser por haber incluido demasiados regresores
- 2. Probar la significancia de la variable de tendencia lineal (β)
- 3. Estimar sin tendencia. Solo con constante
- 4. Estimar sin tendencia ni constante

Prueba de Dickey-Fuller aumentada

- Una limitante de la prueba es cuando existe cambio de régimen. Si por ejemplo la serie es estacionaria pero existe cambio de régimen es posible que la prueba la detecte como $I(1)$
- Un problema común de la prueba ADF es el bajo poder (probabilidad de rechazar una hipótesis nula falsa) derivado de la *equivalencia observacional*

Prueba de Dickey-Fuller aumentada

- Supongamos que el verdadero modelo es no estacionario:
 $y_t = y_{t-1} + \varepsilon_t$ pero en una muestra finita de tamaño T
estimamos un AR(1) estacionario $y_t = \phi y_{t-1} + \varepsilon_t$
- Consideremos la predicción y el MSE para ambos modelos
- Para el random walk: $E_t(y_{t+h}) = y_t \Rightarrow$
 $E_t(y_{t+h} - E_t(y_{t+h}))^2 = h\sigma^2$
- Para el proceso AR(1) estacionario: $E_t(y_{t+h}) = \phi^h y_t \Rightarrow$
 $E_t(y_{t+h} - E_t(y_{t+h}))^2 = (1 + \phi^2 + \phi^4 + \dots) \sigma^2$
- Para un valor ϕ próximo a 1 ambos procesos se verán casi iguales en una muestra de tamaño finito (equivalencia observacional)

Prueba Kwiatkowski, Phillips, Schmidt y Shin (KPSS, 1992)

- Bajo poder de la prueba ADF
- Puede que no rechacemos la nula de RU pues si la hay o porque la información no es suficiente para rechazar
- Una solución puede ser usar una prueba con H_0) estacionaria
- KPSS difiere de ADF en que bajo H_0) la serie es estacionaria

Prueba KPSS

- Partimos de un RW:

$$y_t = \alpha + \beta t + \gamma_t + \varepsilon_t \quad \gamma_t = \gamma_{t-1} + v_t \quad v_t \sim iid(0, \sigma_v^2)$$

- Si $\text{VAR}(v_t) = 0$ (γ constante) $\Rightarrow \gamma_t = \gamma_0$ y la serie y_t es estacionaria
- Usamos una regresión lineal para estimar el componente estocástico:
$$y_t = \hat{\alpha} + \hat{\varepsilon}_t$$
- Notar que bajo H_0) (de estacionariedad) $\hat{\varepsilon}_t$ es estacionario
- $H_0) \sigma_v^2 = 0 \quad H_1) \sigma_v^2 > 0$

Prueba KPSS

- $KPSS = \frac{1}{T^2 \hat{\sigma}_u^2} \sum_{t=1}^T S_t^2$
- $S_t = \sum_{i=1}^t \hat{\varepsilon}_i$ (suma parcial de los residuos de y_t contra constante y/o tendencia determinística)
- $\hat{\sigma}_u^2$ estimador asintótico de la varianza de $\hat{\varepsilon}_t$
- Es una prueba de parámetros constantes contra parámetros de un RW
- Valores críticos simulados por KPSS

Comparación ADF-KPSS

- Podemos comparar los resultados de las pruebas ADF y KPSS
- ADF: $H_0) y_t \sim I(1)$. KPSS: $H_0) y_t \sim I(0)$. Cuatro posibles resultados
- 1. ADF rechaza H_0). KPSS no rechaza H_0)
- 2. ADF no rechaza H_0). KPSS rechaza H_0)
- 3. ADF rechaza H_0). KPSS rechaza H_0)
- 4. ADF no rechaza H_0). KPSS no rechaza H_0)

Comparación ADF-KPSS

- En 1 y 2 los resultados son robustos
- En 3 y 4 los resultados son diferentes
- Esto se llama confirmatory data analysis

Pruebas de estacionariedad: implicancias

- Las series en el VAR tienen que ser estacionarias
- Si concluimos que las series no son estacionarias debemos ver si sus primeras diferencias lo son para incluirlas en el VAR

VAR estructural (o en forma primitiva)

- Por simplicidad comenzamos con el VAR(1) con 2 variables endógenas (y_t, z_t)
- $$\begin{cases} y_t = \beta_{10} - \beta_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_t^y \\ z_t = \beta_{20} - \beta_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_t^z \end{cases}$$
- ε_t^Y y ε_t^Z ruidos blancos independientes $\Rightarrow COV(\varepsilon_t^y, \varepsilon_t^z) = 0$

VAR estructural (o en forma primitiva)

- Por simplicidad comenzamos con el VAR(1) con 2 variables endógenas (y_t, z_t)
- $$\begin{cases} y_t = \beta_{10} - \beta_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_t^y \\ z_t = \beta_{20} - \beta_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_t^z \end{cases}$$
- ε_t^Y y ε_t^Z ruidos blancos independientes $\Rightarrow COV(\varepsilon_t^y, \varepsilon_t^z) = 0$

VAR estructural

- Todas las variables (y_t, z_t) son endógenas
- El modelo presenta dinámica pues hay rezagos de las variables endógenas
- Hay shocks para cada variable endógena pues todas son aleatorias y no hay variables determinísticas
- No hay ceros a priori. A priori no restringimos nada
- Hay retroalimentación entre ecuaciones. Un shock a z en t también afecta a y en t
- ¿Qué pasa si estimo cada ecuación por separado, por ej, por MCO?

VAR estructural

- Si estimamos cada ecuación por separado (por ej. por MCO) hay endogeneidad que sesga las estimaciones
- Hay sesgo de simultaneidad si β_{12} o β_{21} no son 0
- Muestren que los errores están correlacionados con las variables explicativas y por lo tanto hay endogeneidad
- Si hay un shock a z en t , eso afecta z en t , y por la primera ecuación también se afecta y en t , y entonces en la segunda cambia y_t
- Por lo tanto ε_t^z está correlacionado como y_t que es variable explicativa y existe endogeneidad (en el modelo de regresión lineal $X^T \varepsilon \neq 0$. Y entonces los estimadores MCO son sesgados e inconsistentes)

VAR estructural

- Entonces no podemos estimar adecuadamente el VAR estructural por MCO
- Por lo tanto lo que naturalmente podemos hacer es buscar la forma reducida del VAR

$$\underbrace{\begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix}}_B \underbrace{\begin{bmatrix} y_t \\ z_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix}}_{\Gamma_0} + \underbrace{\begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}}_{\Gamma_1} \underbrace{\begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^z \end{bmatrix}}_{\varepsilon_t}$$

- Esto es: $BX_t = \Gamma_0 + \Gamma_1 X_{t-1} + \varepsilon_t$

- Premultiplicando por $B^{-1} \Rightarrow$

$$B^{-1}BX_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 X_{t-1} + B^{-1}\varepsilon_t$$

VAR en forma reducida

- Esto es:
$$X_t = \underbrace{B^{-1}\Gamma_0}_{A_0} + \underbrace{B^{-1}\Gamma_1}_{A_1} X_{t-1} + \underbrace{B^{-1}\varepsilon_t}_{e_t}$$

- $$X_t = A_0 + A_1 X_{t-1} + e_t$$
 VAR en forma reducida con

- $$A_0 = B^{-1}\Gamma_0 = \frac{1}{1-\beta_{12}\beta_{21}} \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix}_{2 \times 1} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix}$$

- $$A_1 = B^{-1}\Gamma_1 = \frac{1}{1-\beta_{12}\beta_{21}} \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}_{2 \times 2} =$$
$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

- $$e_t = \begin{bmatrix} e_t^y \\ e_t^z \end{bmatrix} = \frac{1}{1-\beta_{12}\beta_{21}} \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^z \end{bmatrix}$$

VAR en forma reducida

- $X_t = A_0 + A_1 X_{t-1} + e_t$ o
- $$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_t^y \\ e_t^z \end{bmatrix}$$
- $E(e_t^y) = E(e_t^z) = 0$
- $$V(e_t^y) = \frac{(\sigma_y^2 + \beta_{12}^2 \sigma_z^2)}{(1 - \beta_{12} \beta_{21})^2} \quad V(e_t^z) = \frac{(\sigma_z^2 + \beta_{21}^2 \sigma_y^2)}{(1 - \beta_{12} \beta_{21})^2}$$
- $$COV(e_t^y, e_t^z) = \frac{1}{(1 - \beta_{12} \beta_{21})^2} (-\beta_{21} \sigma_y^2 - \beta_{12} \sigma_z^2) \neq 0$$
- Estimación?

VAR en forma reducida: estimación

- El VAR en forma reducida sigue siendo un VAR. El modelo $X_t = A_0 + A_1 X_{t-1} + e_t$, el VAR en forma reducida no tiene problemas de simultaneidad por lo cual es posible estimar cada ecuación por separado por MCO
- Notar que ahora si hay covarianza entre e_t^y y e_t^z
- Por lo que podemos estimar A_0 , A_1 , las varianzas y la covarianza
- Sin embargo los parámetros de la forma reducida no son de interés. ¿Cómo los interpretamos? ¿Qué nos dicen?
- No mucho pues los parámetros que nos interesan son los del VAR estructural y no los del VAR reducido

Identificación del VAR

- El problema de identificación del VAR es si podemos recuperar los parámetros de interés del VAR estructural a partir de la estimación por MCO de los parámetros del VAR en forma reducida
- Realizamos las cuentas para el VAR(1) con 2 variables endógenas
- Parámetros estimados del VAR reducido:
 $a_{10}, a_{20}, a_{11}, a_{12}, a_{21}, a_{22}, \sigma_{e_y}^2, \sigma_{e_z}^2, COV(e_t^y, e_t^z)$ (9 parámetros)
- Parámetros desconocidos del VAR estructural:
 $\beta_{10}, \beta_{20}, \beta_{12}, \beta_{21}, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \sigma_z^2, \sigma_y^2$ (10 parámetros)
- Por lo que no podemos identificar el VAR. No se soluciona con un tamaño de muestra mayor

Identificación del VAR

- Con 9 parámetros estimados de la forma reducida no podemos recuperar los 10 parámetros estructurales
- Esto se generaliza para el caso el VAR(1) con n variables endógenas
- Parámetros estructurales:
 - $n^2 - n$ en B (la diagonal es 1)
 - n varianzas
 - n^2 parámetros de rezagos
 - n constantes
- Parámetros forma reducida:
 - n varianzas
 - $\frac{n^2 - n}{2}$ covarianzas
 - n^2 parámetros de rezagos
 - n constantes

Identificación del VAR

- $\# \text{Parámetros estructurales} - \# \text{Parámetros forma reducida} = n^2 - n - \frac{n^2 - n}{2} = \frac{2n^2 - 2n - n^2 + n}{2} = \frac{n^2 - n}{2} = \frac{n(n-1)}{2}$
- Para identificar el VAR estructural precisamos $\frac{n(n-1)}{2}$ restricciones adicionales
- Si por ej. $n=3$ precisamos 3 restricciones
- Esto mismo vale para el VAR con p rezagos

Identificación del VAR: 1) Cholesky

- Como no podemos identificar el modelo de interés (VAR estructural) tenemos que realizar supuestos de identificación
- El supuesto más simple es $\beta_{12} = 0$ o $\beta_{21} = 0$
- Si $\beta_{21} = 0 \implies$ VAR estructural:
- $y_t = \beta_{10} - \beta_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_t^y$
- $z_t = \beta_{20} + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_t^z$
- El modelo es recursivo. El shock ε_t^z impacta a y_t, z_t instantáneamente
- Pero el shock ε_t^y afecta solo a y_t (en el tiempo termina afectando a z a través del efectos de los rezagos de y)

Identificación del VAR: 1) Cholesky

- Y ahora tenemos 9 parámetros conocidos y 9 parámetros desconocidos
- $a_{10} = \beta_{10} - \beta_{12}\beta_{20}$ $a_{22} = \gamma_{22}$ $a_{11} = \gamma_{11} - \beta_{12}\gamma_{21}$
- $\sigma_{e^y}^2 = \sigma_y^2 + \beta_{12}\sigma_z^2$ $a_{12} = \gamma_{12} - \beta_{12}\gamma_{22}$ $\sigma_{e^z}^2 = \sigma_z^2$
- $a_{20} = \beta_{20}$ $COV(e_t^y, e_t^z) = -\beta_{12}\sigma_z^2$ $a_{21} = \gamma_{21}$
- Esta manera de identificar al modelo estructural a partir de la forma reducida se llama **descomposición de Cholesky** (descomposición triangular)
- Se ordena como los shocks se propagan en el sistema dinámico

Identificación del VAR: 1) Cholesky

- Notar que la descomposición de Cholesky no es la única. En el ejemplo podríamos haber supuesto $\beta_{12} = 0$ y así haber obtenido otro modelo estructural a partir de la misma forma reducida
- Por lo tanto toda identificación de esta manera es arbitraria
- Por lo que el análisis VAR debe presentar y estudiar las diferentes posibles descomposiciones
- Es decir, debe ser robusto. Pero un inconveniente es que si el número de variables endógenas $n > 2$ hay muchos ordenes posibles y el análisis se vuelve complejo. Para $n=2$ hay dos ordenes posibles y para $n=3$ hay 6 ordenes posibles, donde n es el número de variables endógenas
- Para un VAR con n variables endógenas hay $n!$ ordenes posibles

Identificación del VAR: 1) Cholesky

- Con $n = 2$ hay 2 ordenes posibles:

- i) $\beta_{21} = 0 \Rightarrow$ Orden: $z \rightarrow y$

$$\begin{cases} y_t = \beta_{10} - \beta_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_t^y \\ z_t = \beta_{20} + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_t^z \end{cases}$$

- ii) $\beta_{12} = 0 \Rightarrow$ Orden: $y \rightarrow z$

$$\begin{cases} y_t = \beta_{10} + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_t^y \\ z_t = \beta_{20} - \beta_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_t^z \end{cases}$$

Identificación del VAR: 1) Cholesky

- Con $n = 3$ hay 6 ordenes posibles
- $w_t \rightarrow z_t \rightarrow y_t$
- $w_t \rightarrow y_t \rightarrow z_t$
- $z_t \rightarrow w_t \rightarrow y_t$
- $z_t \rightarrow y_t \rightarrow w_t$
- $y_t \rightarrow z_t \rightarrow w_t$
- $y_t \rightarrow w_t \rightarrow z_t$

Identificación del VAR: 1) Cholesky

- Por ej. $w \rightarrow z \rightarrow y$ hay 3 restricciones. El modelo es

$$\left\{ \begin{array}{l} y_t = \beta_{10} - \beta_{12}z_t - \beta_{13}w_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \gamma_{13}w_{t-1} + \varepsilon_t^y \\ z_t = \beta_{20} - \beta_{23}w_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \gamma_{23}w_{t-1} + \varepsilon_t^z \\ w_t = \beta_{30} + \gamma_{31}y_{t-1} + \gamma_{32}z_{t-1} + \gamma_{33}w_{t-1} + \varepsilon_t^w \end{array} \right\}$$

- Las 3 restricciones son: $\beta_{31} = \beta_{32} = \beta_{21} = 0$

Identificación del VAR: 2) Restricciones de largo plazo (LP)

- Supone que el efecto de la suma de los shocks en el LP es 0
- Ej. Blanchard y Quah, AER 1989
- VAR con dos variables: desempleo y producto
- Shocks de oferta y demanda
- El efecto de LP del shock de demanda en el producto es 0

Identificación del VAR: 3) Restricciones de signo

- Faust (1998, Carnegie-Rochester Conference on Public Policy)
- Información a priori: shocks de política contractivos elevan la tasa de interés y reducen M, P, Y
- Requiere que definamos todo el espacio posible de respuestas
- Resultados no únicos

Identificación del VAR: 4) Otros enfoques

- Restricciones de corto plazo
- Otros a priori (economía pequeña que no afecta las variables internacionales, etc)
- Otros

Estabilidad del VAR

- Misma pregunta y motivación que en los AR
- Comenzamos por un VAR(1) con dos variables en forma reducida:

$$x_t = A_0 + A_1 x_{t-1} + e_t \quad (1)$$

- con

$$x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \quad A_0 = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix}, \quad A_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad e_t = \begin{bmatrix} e_t^y \\ e_t^z \end{bmatrix}$$

- Dado que el termino constante no afecta la estacionariedad asumimos $a_{10} = a_{20} = 0$
- Por lo tanto tenemos las siguientes ecuaciones:

$$y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + e_t^y \quad (2)$$

$$z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + e_t^z \quad (3)$$

Estabilidad del VAR

- Reescribiendo (3):

$$(1 - a_{22}L) z_t = a_{21}y_{t-1} + e_t^z \quad \Rightarrow \quad z_t = \frac{a_{21}y_{t-1} + e_t^z}{(1 - a_{22}L)} \quad (4)$$

- Rezagando (4) un período y sustituyendo en (2) tenemos:

$$y_t = a_{11}y_{t-1} + a_{12} \frac{a_{21}y_{t-2} + e_{t-1}^z}{(1 - a_{22}L)} + e_t^y \quad (5)$$

- Reescribiendo (5):

$$(1 - a_{11}L) y_t = \frac{a_{12}a_{21}}{(1 - a_{22}L)} y_{t-2} + \frac{a_{12}e_{t-1}^z}{(1 - a_{22}L)} + \frac{(1 - a_{22}L)}{(1 - a_{22}L)} e_t^y \quad \Rightarrow$$

Estabilidad del VAR

- $(1 - a_{11}L) y_t = \frac{a_{12}a_{21}y_{t-2} + a_{12}e_{t-1}^z + (1 - a_{22}L)e_t^y}{(1 - a_{22}L)} \Rightarrow$
- $(1 - a_{11}L)(1 - a_{22}L) y_t = a_{12}a_{21}y_{t-2} + a_{12}e_{t-1}^z + (1 - a_{22}L)e_t^y \Rightarrow$
- $$\boxed{[(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2] y_t = a_{12}e_{t-1}^z + (1 - a_{22}L)e_t^y} \quad (6)$$
- (2) en (3) y despejando z_t obtenemos:

$$\boxed{[(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2] z_t = a_{21}e_{t-1}^y + (1 - a_{11}L)e_t^z} \quad (7)$$

Estabilidad del VAR

- En (6) y (7) vemos que como los procesos de medias móviles finitos siempre son estacionarios, la estacionariedad del VAR depende de que el polinomio $\Phi(L) = [(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2]$ sea estacionario
- Es decir, la estacionariedad del VAR depende de que las raíces del polinomio autoregresivo sean en valor absoluto mayores a 1:

$$\Phi(L) = [(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2] = 0 \quad (8)$$

Estabilidad del VAR

- Ahora volvamos a (1) que se puede reescribir como:

$$[I - A_1 L] x_t = e_t \Rightarrow$$

- $$\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} L \right] x_t = e_t \quad (9)$$

- Sea $\Theta(L) = \begin{bmatrix} (1 - a_{11})L & -a_{12}L \\ -a_{21}L & (1 - a_{22})L \end{bmatrix}$

- Determinante

$$\Theta(L) = [(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2] \quad (\Theta(L) = \Phi(L))$$

- Por lo que la condición de estacionariedad del VAR es que el polinomio del determinante $\Theta(L)$ sea estacionario. Esto es, las raíces del determinante

$$\Theta(L) = [(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2] = 0 \text{ deben ser mayor a 1 en valor absoluto}$$

Estabilidad del VAR

- Ahora consideremos un VAR(p):

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + e_t \quad \Rightarrow$$

- $[I - A_1 L - A_2 L^2 - \dots - A_p L^p] x_t = A_0 + e_t \Rightarrow \Phi(L) x_t = A_0 + e_t$
- Por lo que la condición de estacionariedad del VAR es que el polinomio del determinante $\Phi(L)$ sea estacionario
- Esto es, las raíces del determinante $\Phi(L) = [I - A_1 L - A_2 L^2 - \dots - A_p L^p] = 0$ deben ser mayor a 1 en valor absoluto
- Ej. $x_t = \begin{bmatrix} 0,02 \\ 0,03 \end{bmatrix} + \begin{bmatrix} 0,5 & 0,5 \\ 0,5 & 0,5 \end{bmatrix} x_{t-1}$

Estimación del VAR(p)

- $VAR(p) : X_t = A_0 + A_1X_{t-1} + A_2X_{t-2} + \dots + A_pX_{t-p} + e_t$
- Estimamos por MCO pues no hay correlación entre los regresores y el término de error
- El número de parámetros a estimar es enorme. ¿Cuántos?

Estimación del VAR(p)

- $n + n^2p$ parámetros a estimar (A_0 es $n \times 1$ y A_1 es $n \times n$)
- Ej: si $n=5$ y $p=5 \Rightarrow n + n^2p = 5 + 5^2 \cdot 5 = 130$ (y sin contar las varianzas)
- Sobreparametrización relativa a:

→ el número de datos disponibles (por esto se aplica en finanzas pues suele disponerse de muchos datos)

→ ¿qué obtenemos? ¿qué conclusiones sacamos? Lo que obtenemos no es de corto plazo ni es estructural. Como sacamos toda la dinámica de los shocks no es de corto plazo. Y no es estructural pues se trata de una forma reducida

- Podemos decir que estudiamos la dinámica del modelo de largo plazo en una forma reducida

Estimación del VAR(p)

- Es posible incluir variables exógenas (factores climáticos, etc)
- Con una muestra finita el número de rezagos p del VAR debe ser acotado
- En general con un p pequeño se capta bastante del proceso
- Para seleccionar el número óptimo de rezagos hacer pruebas de contenido informacional o restricciones cross-equation
- i) Criterios de información
 - Akaike multivariado: $MAIC = Ln | \hat{\Sigma} | + \frac{2k}{T}$
 - Schwartz multivariado: $MSIC = Ln | \hat{\Sigma} | + \frac{k}{T} Ln(T)$
 - k=# de regresores, $\hat{\Sigma}_{n \times n}$ =matriz de VAR-COV de e

Estimación del VAR(p)

- ii) Restricciones cross-equation

- También se pueden imponer restricciones cross-equation En el espíritu (no restrictivo) de los modelos VAR, cada ecuación debe tener una igual duración del rezago Supongamos que se estima un VAR(8) bivariado usando datos trimestrales y 8 rezagos de las dos variables en cada ecuación, y queremos probar la restricción de que los coeficientes en los rezagos 5 a 8 son 0 en conjunto

- Esto se puede realizar mediante una prueba de razón de verosimilitud

$$LR = T \left[Ln \mid \hat{\Sigma}_R \mid -Ln \mid \hat{\Sigma}_U \mid \right] \sim \chi_j^2$$

- R=restringido, U=sin restringir y j es la cantidad de restricciones

Estimación del VAR(p)

- Los tests t no tienen mucho uso por colinealidad dado que los rezagos están correlacionados. Se usan las pruebas F
- Difícil interpretar tantos parámetros
- Los parámetros tampoco sirven pues son de una forma reducida y colineales

Predicción

- $y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + e_t^y$
- $z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + e_t^z$
- $E[y_{t+1}|t] = E[a_{11}y_{t|t}] + a_{12}E[z_{t|t}] + E[e_{t+1|t}^y] = a_{11}y_t + a_{12}z_t$
- $E[y_{t+2}|t] = ?$

Evaluación de la predicción

- Min MSE (Mean Square Error) = $\frac{1}{s} \sum_{i=1}^s (y_{t+i} - y_{t+i|t})^2$
- Min MAE (Mean Absolute Error) = $\frac{1}{s} \sum_{i=1}^s |y_{t+i} - y_{t+i|t}|$
- Min MAPE (Mean Absolute Percent Error) = $100 * \frac{1}{s} \sum_{i=1}^s \left| \frac{y_{t+i} - y_{t+i|t}}{y_{t+i}} \right|$
- Nos quedamos con el modelo que minimize cualquiera de estos criterios

Evaluación de la predicción: Prueba de Granger-Newbold (1976)

- Sin embargo lo anterior no nos permite detectar diferencias significativas desde un punto de vista estadístico
- El objetivo es comparar las predicciones de dos modelos
- Se asumen errores de predicción con esperanza 0, no correlacionados y normales
- Sean e_{1i}, e_{2i} los errores de predicción de los dos modelos ($i=1,2,\dots,s$).
s predicciones a un paso
- Definimos $x_i = e_{1i} + e_{2i}$; $z_i = e_{1i} - e_{2i}$
- Bajo H_0) Igual poder predictivo de ambos modelos:
- $\tau_{xz} = E(x_i z_i) - E(x_i) E(z_i) = E(x_i z_i) = E(e_{1i}^2) - E(e_{2i}^2)$

Evaluación de la predicción: Prueba de Granger-Newbold

- τ_{xz} es el coeficiente de correlación entre x_i, z_i
- Si $\tau_{xz} > 0 \Rightarrow$ el modelo 1 tiene mayor MSE
- Si $\tau_{xz} < 0 \Rightarrow$ el modelo 2 tiene mayor MSE
- $\hat{\tau}_{xz}$ es el coeficiente de correlación muestral entre x_i, z_i
- Granger y Newbold probaron:

$$t_s = \frac{\hat{\tau}_{xz}}{\sqrt{\left(1 - \hat{\tau}_{xz}^2\right)/(s-1)}} \sim t_{s-1}$$

- Si $\hat{\tau}_{xz} > 0$ y $t_s > \text{valor crítico de tablas de una } t_{s-1} \Rightarrow$ rechazamos que ambos modelos predicen igualmente bien. El MSE del modelo 1 es estadísticamente mayor que el del modelo 2

Causalidad

- Una tentación es identificar los modelos VAR sobre la base de pruebas de causalidad en el tiempo, lo que es equivocado. Para ver esto claramente, veamos la definición de causalidad
- En los VAR podemos ver si en la predicción de cada una de las variables son significativos o útiles los rezagos de las demás variables
- El enfoque tradicional en econometría es definir el MRL con base a una teoría económica previa que nos dice cual es y , cual es x . Por ejemplo, y es consumo, x es ingreso y el consumo es una función del ingreso. Y regresamos y respecto a x para ver la significación del coeficiente de x . Es claro, que una alta correlación entre dos variables no constituye evidencia por si de que existe una relación de causalidad entre ambas

Causalidad a la Granger (1969)

- La pregunta que nos hacemos es: ¿z puede ayudar a predecir y?
- Si la respuesta es si, se dice que z causa a y en el sentido de Granger
- Si estimamos la siguiente regresión por MCO:

$$y_t = \phi(L)y_{t-1} + \theta(L)z_{t-1} + \varepsilon_t \quad H_0) \text{ z no causa y, } \theta(L) = 0$$

- Si rechazamos que $\theta(L)$ es 0, entonces decimos que z causa a y en el sentido de Granger
- Esto es, z causa a y en el sentido de Granger si contiene información que ayuda a predecir y dicha información no esta contenida en la historia de y

Causalidad a la Granger

- No se trata de una prueba de causalidad en el sentido literal de la palabra, de causa efecto. Es un test de "precedencia temporal" o "predicibilidad". Es de destacar que se plantea como desconocida la mecánica por la que la variable z termina afectando la variable y por lo que se debe tener cuidado con la interpretación de los resultados
- Dada la alta correlación la prueba se hace sobre el conjunto de rezagos de z y no sobre parámetros individuales, por lo que es una prueba F con $H_0) \theta(L) = 0$. Dado que es posible obtener una correlación espúrea se testea $H_0)$ que z no causa en el sentido de Granger a y
- En la práctica es posible que z causa a la Granger a y , que y cause a la Granger a z . Esto puede suceder pues las variables son simultáneas (explicadas por una tercera variable w) como cuando se vuelven simultáneas por problemas de medición, por ejemplo la causalidad es mensual pero los datos son trimestrales

Causalidad a la Granger

- Los VAR son el entorno óptimo para testear causalidad a la Granger
- Cuando se rechaza la hipótesis de no causalidad pues los coeficientes son estadísticamente significativos, pero son bajos en valor absoluto en muchas ocasiones el práctica pueden dejarse de lado sin inconvenientes

- Otra manera de ver el test de que z no causa y a la Granger es:

$$ECM[y_t^f \mid y_{t-1}, \dots] = ECM[y_t^f \mid y_{t-1}, \dots, z_{t-1}, \dots]$$

- Y áca decimos que z no es linealmente informativo sobre el futuro de y

Causalidad a la Granger

- Veamos las implicancias de causalidad a la Granger en un VAR(2) bivariado
- Supongamos que z no causa a y en el sentido de Granger:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_{t-2} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} e_t^y \\ e_t^z \end{bmatrix}$$

- Si hacemos la predicción un período adelante de y :

$$E[y_{t+1} \mid y_t, z_t] = a_{10} + a_{11}y_t + b_{11}y_{t-1}$$

y para s períodos adelante:

$$E[y_{t+s} \mid y_t, z_t] = a_{10} + a_{11}E[y_{t+s-1} \mid y_t, z_t] + b_{11}E[y_{t+s-2} \mid y_t, z_t]$$

- Por lo que solo depende de los valores de y

Función de impulso respuesta (IRF)

- Una manera más general de mirar el tema de predicción es a través de la función impulso respuesta (IRF)
- Damos un shock al proceso y vemos como se disipa
- Uno de los usos más frecuentes de los modelos VAR son las FIR que describen la trayectoria del sistema dinámico frente a shocks

Función de impulso respuesta (IRF)

- Comencemos por un VAR(1) con 2 variables en forma matricial
- $x_t = A_0 + A_1 x_{t-1} + e_t \Rightarrow$ sustituyendo $x_{t-1}, x_{t-2} \dots$
- $x_t = A_0 + A_1 [A_0 + A_1 x_{t-2} + e_{t-1}] + e_t =$
 $A_0 + A_1 A_0 + A_1 e_{t-1} + e_t + A_1^2 x_{t-2} =$
- $= A_0 + A_1 A_0 + A_1 e_{t-1} + e_t + A_1^2 [A_0 + A_1 x_{t-3} + e_{t-2}] =$
- $= A_0 + A_1 A_0 + A_1^2 A_0 + A_1^2 e_{t-2} + A_1 e_{t-1} + e_t + A_1^2 x_{t-3} = \dots$
- $= (I + A_1 + A_1^2 + \dots + A_1^s) A_0 + \sum_{i=0}^s A_1^i e_{t-i} + A_1^{s+1} x_{t-s-1}$

Función de impulso respuesta

- $$x_t = (I + A_1 + A_1^2 + \dots + A_1^s) A_0 + \sum_{i=0}^s A_1^i e_{t-i} + A_1^{s+1} x_{t-s-1}$$

- Bajo estabilidad:

- $$x_t = (I - A_1)^{-1} A_0 + \sum_{i=0}^{\infty} A_1^i e_{t-i}$$

- $$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{t-i}^y \\ e_{t-i}^z \end{bmatrix} \quad (9)$$

- $$\bar{y} = \frac{a_{10}(1-a_{22})+a_{12}a_{20}}{(1-a_{11})(1-a_{22})-a_{12}a_{21}} \quad \bar{z} = \frac{a_{20}(1-a_{11})+a_{21}a_{10}}{(1-a_{11})(1-a_{22})-a_{12}a_{21}}$$

- Esto nos permite calcular cuanto cambia $y_t, y_{t+1}, y_{t+2}, \dots$ cuando cambia e_t^y o e_t^z . Interesante?

Función de impulso respuesta

- No, pues las IRF vistas anteriormente están hechas sobre los shocks del modelo en forma reducida (los e) y no respecto de las innovaciones de la forma estructural o primitiva (los ε) que es lo que realmente nos interesa
- Recordemos la relación entre ambos shocks ($e_t = B^{-1}\varepsilon_t$):

$$e_{t-i} = \begin{bmatrix} e_{t-i}^y \\ e_{t-i}^z \end{bmatrix} = \frac{1}{1-\beta_{12}\beta_{21}} \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-i}^y \\ \varepsilon_{t-i}^z \end{bmatrix} \quad (10) \quad (10) \text{ en } (9):$$

- $$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \frac{1}{1-\beta_{12}\beta_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-i}^y \\ \varepsilon_{t-i}^z \end{bmatrix}$$

Función de impulso respuesta

- $$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \frac{1}{1-\beta_{12}\beta_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-i}^y \\ \varepsilon_{t-i}^z \end{bmatrix}$$

- Simplificando la notación:

- $$\psi(i) = \frac{1}{1-\beta_{12}\beta_{21}} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix}$$

- $$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \psi_{11}(i) & \psi_{12}(i) \\ \psi_{21}(i) & \psi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{t-i}^y \\ \varepsilon_{t-i}^z \end{bmatrix}$$

Función de impulso respuesta

- Los términos $\psi(i)$ se denominan multiplicadores dinámicos
- La IRF son los gráficos de $\psi_{11}(i), \psi_{12}(i), \psi_{21}(i), \psi_{22}(i)$ contra i
- Por ej. $\psi_{12}(0)$, es el cambio en y en el período t ante un shock a z en t (ε_t^z)
- $\psi_{12}(1)$, es el cambio en y un período adelante ($t+1$) ante un shock a z en t (ε_t^z)
- $\sum_{i=0}^{\infty} \psi_{12}(i)$ es el efecto acumulado en el tiempo en y de un shock a z en t

Función de impulso respuesta

- La respuesta depende de la identificación supuesta, por lo que se deben presentar diferentes descomposiciones
- La forma de las respuestas no dependen del tamaño del shock, pero el nivel si. Por ejemplo un shock de tamaño 1 tiene la misma forma de disipación que uno de tamaño 2 pero su nivel es la mitad
- Si se cumple absolute summability, la respuesta de largo plazo es 0. No hay respuesta en el largo plazo

Función de impulso respuesta: inferencia

- La IRF la construimos sobre variables aleatorias (los coeficientes estimados), y por lo tanto también es una variable aleatoria y queremos conocer su distribución para hacer pruebas de hipótesis
- 2 enfoques:
 - paramétrico: normalidad
 - no paramétrico: bootstrap

Inferencia I: normalidad

- Siendo \hat{A} los parámetros estimados del VAR en forma reducida, por el teorema central del límite:

$$\sqrt{T}(\hat{A} - A) \rightarrow X \quad \text{con} \quad X \sim N(0, \Omega Q^{-1})$$

donde Q es el valor esperado de la matriz de momentos de los regresores (por ej. $E(x_t x_t^T)$)

- Como lo que nos interesa es la IRF que es una función $f(\cdot)$ de los parámetros, podemos usar el teorema de Slutsky y obtenemos:

$$\sqrt{T}(f(\hat{A}) - f(A)) \rightarrow G_s X$$

donde $G_s = \frac{\partial f(\cdot)}{\partial \hat{A}}$. Por lo que $\sqrt{T}(f(\hat{A}) - f(A)) \rightarrow N(0, G_s \Omega Q^{-1} G_s^T)$

Inferencia I: normalidad

- El error estándar que nos interesa viene dado por la raíz cuadrada de la diagonal de la matriz $G_S \Omega Q^{-1} G_S^T$
- Teorema de Slutsky: Sea x_T una secuencia de vectores aleatorios $K \times 1$ tales que $x_T \rightarrow c$ (converge en probabilidad a c), entonces $g(x_t) \rightarrow g(c)$ si $T \rightarrow \infty$.

$$\text{plim } g(x_t) = g(\text{plim } x_T) \quad \text{si } g(.) \text{ es continua en } \text{plim } x_t.$$

- El teorema de Slutsky es el uso más popular del operador plim , muestra que dicho operador pasa por funciones no lineales siempre y cuando sean continuas. El operador esperanza no tiene esta propiedad y esto es una de las cosas que hace difícil el análisis de las propiedades de los estimadores en muestra pequeña

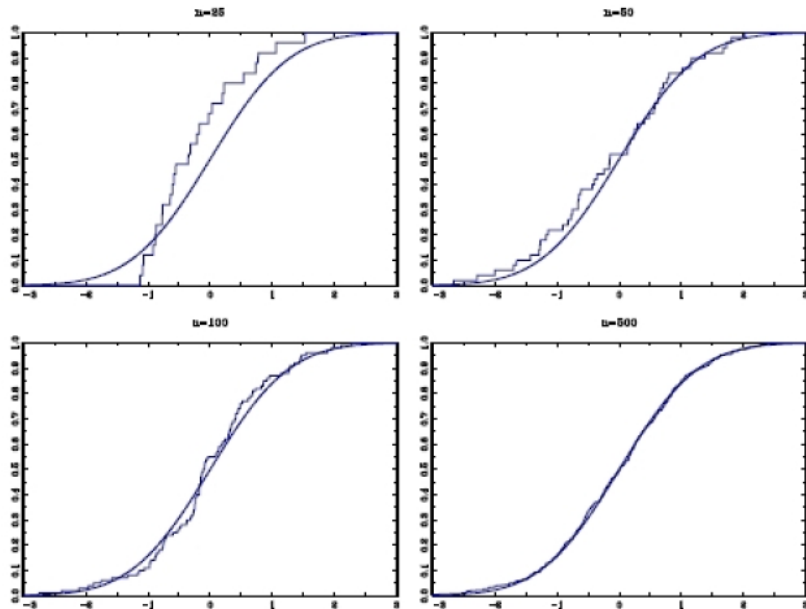
Inferencia II: bootstrap (Efron,1979)

- Inferencia estadística vista previamente se basa en teoría asintótica
- Resultados en muestra finita de pruebas estadísticas vistas no disponibles
- Bootstrap: enfoque alternativo que considera a la muestra observada como la población
- Se basa en teoría asintótica y exactamente válida en grandes muestras

Inferencia II: bootstrap

- Evita realizar supuestos sobre distribución
- Permite realizar inferencia en problemas sin resolución analítica mediante calculos numéricos
- No requiere asumir distribuciones asintóticas
- Generalmente bootstrap equivalente a resultados asintóticos de primer orden

Bootstrap: F empírica para una $N(0,1)$



Bootstrap en el VAR

- 1. Estimar el VAR por MCO y guardar las matrices \hat{A} , $\hat{\varepsilon}$
- 2. Extraer una muestra aleatoria con reemplazo de tamaño T de los $\hat{\varepsilon}$
- 3. Calcular $x_t^s = \hat{A}_0 + \hat{A}_1 x_{t-1} + \dots + \hat{A}_p x_{t-p} + \hat{\varepsilon}_t^s$
- 4. Estimar este nuevo VAR(p) y obtener las *IRF*
- 5. Repetir los pasos 2 a 4 S veces, obteniendo:
 $IRF_{(1)}, IRF_{(2)}, \dots, IRF_{(S)}$ ($S = 1.000, 5.000, \text{ o } 10.000$)
- 6. Aproximar la distribución empírica de la *IRF*. Por ej. si S es 1000, y ordenamos crecientemente las *IRF* : $\left[IRF_{(25)}, IRF_{(975)} \right]$ es un IC al 95%

Descomposición de varianza (VD)

- La descomposición de la varianza ofrece un método ligeramente diferente de examinar la dinámica de los modelos VAR
- Brinda información sobre la proporción de los movimientos en las variables dependientes que son explicados por sus “propios” shocks, versus los shocks de otras variables
- Esto se realiza determinando cuanto de la predicción s pasos hacia adelante de la varianza del error se explica por innovaciones de cada variable exploratoria ($s = 0, 1, 2, \dots$)
- La VD brinda información acerca de la importancia relativa de cada uno de los shocks a las variables que componen el VAR

Descomposición de varianza

- Calculamos los errores de predicción
- $x_t = A_0 + A_1 x_{t-1} + e_t \Rightarrow x_{t+1} = A_0 + A_1 x_t + e_{t+1}$
- $ep_{t+1|t} = x_{t+1} - x_{t+1|t} = A_0 + A_1 x_t + e_{t+1} - E_t (A_0 + A_1 x_t + e_{t+1})$
- $= A_0 + A_1 x_t + e_{t+1} - A_0 - A_1 x_t - 0 \Rightarrow ep_{t+1|t} = e_{t+1}$
- $ep_{t+2|t} = x_{t+2} - x_{t+2|t} =$
 $A_0 + A_1 x_{t+1} + e_{t+2} - E_t (A_0 + A_1 x_{t+1} + e_{t+2})$
- $= A_0 + A_1 (A_0 + A_1 x_t + e_{t+1}) + e_{t+2} -$
 $E_t [A_0 + A_1 (A_0 + A_1 x_t + e_{t+1}) + e_{t+2}]$
- $ep_{t+2|t} = e_{t+2} + A_1 e_{t+1}$

Descomposición de varianza

- $x_{t+s} = e_{t+s} + A_1 e_{t+s-1} + A_1^2 e_{t+s-2} + \dots + A_1^{s-1} e_{t+1} + (I + A_1 + A_1^2 + \dots + A_1^{s-1}) A_0 + A_1^s x_t$
- $x_{t+s|t} = (I + A_1 + A_1^2 + \dots + A_1^{s-1}) A_0 + A_1^s x_t$
- $ep_{t+s|t} = e_{t+s} + A_1 e_{t+s-1} + A_1^2 e_{t+s-2} + \dots + A_1^{s-1} e_{t+1} = \sum_{i=0}^{s-1} A_1^i e_{t+s-i}$
- Recordar que nos interesan únicamente los shocks estructurales ($e_t = B^{-1} \varepsilon_t$)
- $ep_{t+s|t} = \sum_{i=0}^{s-1} A_1^i B^{-1} \varepsilon_{t+s-i}$

Descomposición de varianza

- $ep_{t+s|t} = \sum_{i=0}^{s-1} A_1^i B^{-1} \varepsilon_{t+s-i}$
- Error de predicción s períodos adelante solo para y como función de los componentes estructurales:
- $ep_{t+s|t}^y = \varphi_{11}(0) \varepsilon_{t+s}^y + \varphi_{11}(1) \varepsilon_{t+s-1}^y + \dots + \varphi_{11}(s-1) \varepsilon_{t+1}^y + \varphi_{12}(0) \varepsilon_{t+s}^z + \varphi_{12}(1) \varepsilon_{t+s-1}^z + \dots + \varphi_{12}(s-1) \varepsilon_{t+1}^z$
- Varianza de dicho error de predicción:
- $\sigma_y^2(s) = \left[\varphi_{11}(0)^2 + \varphi_{11}(1)^2 + \dots + \varphi_{11}(s-1)^2 \right] \sigma_y^2 + \left[\varphi_{12}(0)^2 + \varphi_{12}(1)^2 + \dots + \varphi_{12}(s-1)^2 \right] \sigma_z^2$

pues los términos cruzados de cancelan

Descomposición de varianza

- $\sigma_y^2(s) = \left[\varphi_{11}(0)^2 + \varphi_{11}(1)^2 + \dots + \varphi_{11}(s-1)^2 \right] \sigma_y^2 + \left[\varphi_{12}(0)^2 + \varphi_{12}(1)^2 + \dots + \varphi_{12}(s-1)^2 \right] \sigma_z^2$
- De igual manera para z
- $\sigma_z^2(s) = \left[\varphi_{21}(0)^2 + \varphi_{21}(1)^2 + \dots + \varphi_{21}(s-1)^2 \right] \sigma_y^2 + \left[\varphi_{22}(0)^2 + \varphi_{22}(1)^2 + \dots + \varphi_{22}(s-1)^2 \right] \sigma_z^2$
- Las proporciones explicadas por los shocks a y e z vienen dadas respectivamente por:

$$\frac{\left[\varphi_{11}(0)^2 + \varphi_{11}(1)^2 + \dots + \varphi_{11}(s-1)^2 \right] \sigma_y^2}{\sigma_y^2(s)} \text{ y } \frac{\left[\varphi_{21}(0)^2 + \varphi_{21}(1)^2 + \dots + \varphi_{21}(s-1)^2 \right] \sigma_y^2}{\sigma_z^2(s)}$$

Descomposición de varianza

- Sea $\left[\varphi_{11} (0)^2 + \varphi_{11} (1)^2 + \dots + \varphi_{11} (s-1)^2 \right] = K$
- $\left[\varphi_{12} (0)^2 + \varphi_{12} (1)^2 + \dots + \varphi_{12} (s-1)^2 \right] = M$
- $\left[\varphi_{21} (0)^2 + \varphi_{21} (1)^2 + \dots + \varphi_{21} (s-1)^2 \right] = N$
- $\left[\varphi_{22} (0)^2 + \varphi_{22} (1)^2 + \dots + \varphi_{22} (s-1)^2 \right] = O$
- $\left\{ \begin{array}{l} \sigma_y^2(s) = K\sigma_y^2 + M\sigma_z^2 \\ \sigma_z^2(s) = N\sigma_y^2 + O\sigma_z^2 \end{array} \right\}$
- $VD^{yy}(s) = \frac{K\sigma_y^2}{K\sigma_y^2 + M\sigma_z^2} * 100 \quad \quad VD^{yz}(s) = \frac{M\sigma_z^2}{K\sigma_y^2 + M\sigma_z^2} * 100$

Descomposición de varianza

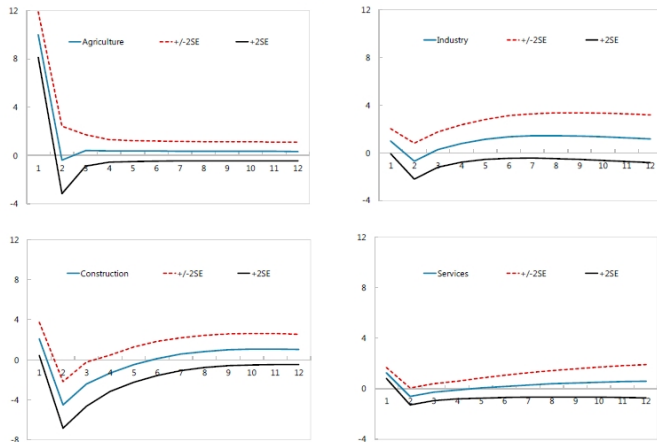
- Notar que como los términos son positivos las varianzas crecen con el horizonte de predicción
- Las VD van a depender de los supuestos de identificación del VAR

VAR: ejemplo

- IMF Uruguay Country Report No. 11/376
- VAR para estimar spillovers del PIB del sector agropecuario en el resto de la economía. 1997Q1 a 2011Q1, $T=61$. $p=2$, $n=4$ (agricultural, industrial, construction, services)
- "The VAR model is specified with agriculture as the most exogenous sector. It is then followed by the industrial, construction, and services sectors"
- "This order seeks to show that a shock to the agricultural sector at period t have a contemporaneous effect on the rest of the sectors included in the model"
- Incluye variables dummy para controlar por factores exógenos como sequías

VAR: ejemplo

Figure 5. Uruguay: Impulse-Response Functions to a 10 Percent Increase in Agricultural Output
(In percent)



Source: IMF staff calculations.

VAR: ejemplo

Table 2. Forecast Error Variance Decomposition Due to an Agricultural Sector Shock
(In percent)

Quarters	Industry	Construction	Services
1	25.7	34.9	53.5
4	15.1	27.9	22.8
8	12.6	23.8	17.7
10	11.5	22.2	16.1

Probando la hipótesis de expectativas racionales

- Sea un VAR(1) bivariado

$$- y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + e_t^y$$

$$- z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + e_t^z$$

- Sean $y_t = (i_t - i_t^*)$ $z_t = [\ln(E_t) - \ln(E_{t-1})]$

- UIP (Uncovered interest parity): $y_t = E_t(z_{t+1})$

- $a_{11}y_{t-1} + a_{12}z_{t-1} + e_t^y = E_t(a_{21}y_t + a_{22}z_t + e_{t+1}^z)$

Probando la hipótesis de expectativas racionales

- $a_{11}y_{t-1} + a_{12}z_{t-1} + e_t^y = E_t (a_{21}y_t + a_{22}z_t + e_{t+1}^z)$
- $E_{t-1} (a_{11}y_{t-1} + a_{12}z_{t-1} + e_t^y) = E_{t-1} (a_{21}y_t + a_{22}z_t + e_{t+1}^z)$
- $a_{11}y_{t-1} + a_{12}z_{t-1} + 0 = a_{21}E_{t-1}y_t + a_{22}E_{t-1}z_t + 0$
- $a_{11}y_{t-1} + a_{12}z_{t-1} =$
 $a_{21} (a_{11}y_{t-1} + a_{12}z_{t-1}) + a_{22} (a_{21}y_{t-1} + a_{22}z_{t-1})$
- $a_{11}y_{t-1} + a_{12}z_{t-1} = (a_{21}a_{11} + a_{22}a_{21}) y_{t-1} + (a_{21}a_{12} + a_{22}^2) z_{t-1}$

Probando la hipótesis de expectativas racionales

- $a_{11}y_{t-1} + a_{12}z_{t-1} = (a_{21}a_{11} + a_{22}a_{21})y_{t-1} + (a_{21}a_{12} + a_{22}^2)z_{t-1}$
- $\left\{ \begin{array}{l} a_{11} = a_{21}a_{11} + a_{22}a_{21} \\ a_{12} = a_{21}a_{12} + a_{22}^2 \end{array} \right\}$
- Tiene que cumplirse:
 - i) $a_{11} = \frac{a_{22}a_{21}}{1-a_{21}}$
 - ii) $a_{12} = \frac{a_{22}^2}{1-a_{21}}$
- Que puede testearse mediante una prueba LR