PUBLICACIONES DE 4[er] CURSO

Curso: 4º

Grado: Economía

Asignatura: ECONOMETRÍA III

TRANSPARENCIAS PARTE 3: MODELOS DE SRIES TEMPORALES

TEMA 7: ANÁLISIS MULTIVARIANTE

Profesores: Antonio Aznar

Departamento de ANÁLISIS ECONÓMICO

Curso Académico 2015/16

Facultad de
Economía y Empresa
Universidad Zaragoza

Tema 7: Analisis Multivariante

1. Modelo VAR. Exogeneidad
2. Cointegración
3. Modelo con Mecanismo de Corrección del Error (MCE ó CVAR)

1. Modelo VAR. Exogeneidad: motivación. Del AR(p) al ARD(p,r)

Modelo Autorregresivo con Retardos Distribuidos (ARD)

- En el tema anterior, hemos considerado modelos de predicción que solo utilizan los valores pasados de Y.
- Tiene sentido añadir otras variables (*X*) que pueden ayudar a predecir *Y*, después de incorporar los valores pasados de *Y*:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \ldots + \delta_r X_{t-\mathrm{q}} + u_t$$

- Este es el modelo autorregresivo con retardos distribuidos con p retardos de *Y* y q retardos de *X* … ***ARD(p,q)***.

CC 14.6 Regresión de series temporales con varios predictores (ARD)

El modelo general de regresión de series temporales permite k predictores adicionales, en el que se incluyen q_1 retardos del primer predictor, q_2 retardos del segundo predictor, y así sucesivamente:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + ... + \beta_p Y_{t-p} +$$

$$\delta_{11} X_{1t-1} + ... + \delta_{1q_1} X_{1t-q_1} +$$

$$+ .... + \delta_{k1} X_{kt-1} + ... + \delta_{1q_k} X_{kt-q_k} + u_t$$

Donde los supuestos son

1.

$$E\left(u_t / Y_{t-1}, Y_{t-2}, ...X_{1t-1}, X_{1t-2} .....X_{kt-1}, X_{kt-2}, ...\right) = 0$$

2. (a) Las variables aleatorias $\left(Y_t, X_{1t}, \ldots. X_{kt}\right)$ son conjuntamente estacionarias.

(b) $\left(Y_t, X_{1t}, \ldots. X_{kt}\right)$ y $\left(Y_{t-j}, X_{1t-j}, \ldots. X_{kt-j}\right)$ se hacen independientes cuando j se hace grande.

3. Los valores extremos elevados son poco probables

4. No existe multicolinealidad perfecta.

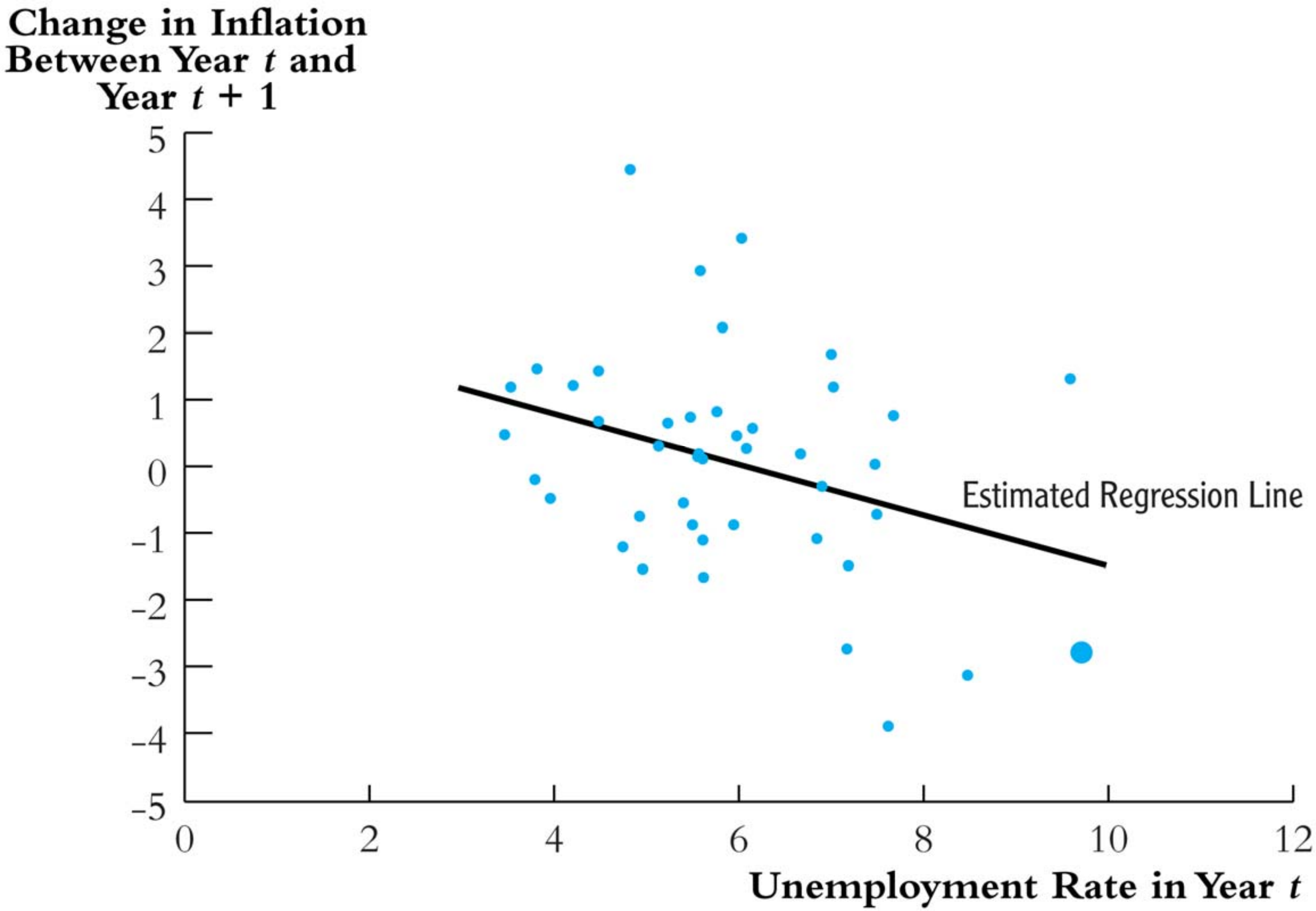
Bajo estos supuestos los estimadores MCO son consistentes.

Ejemplo: Inflación y Desempleo

Según la "Curva de Phillips " si el desempleo está por encima de su nivel de equilibrio, o tipo "natural", entonces la tasa de inflación crecerá. Esto es, el ΔInf_t está relacionado con los valores retardados del tipo de desempleo, con un coeficiente negativo.

- El tipo de desempleo para el que la inflación no cambia se llama el "Tipo de desempleo con inflación constante/Non-Accelerating Inflation Unemployment Rate" (el NAIRU).
- ¿Existe la curva de Phillips en la economía americana?
- ¿Puede ser explotada para predecir la inflación?
- ¿Ha sido la curva de Phillips estable a lo largo del tiempo?

The Empirical U.S. "Phillips Curve," 1962 – 2004 (annual)



Una definición del NAIRU es que es el valor de *unem* para el cual $\Delta Inf = 0$ – la intersección con el eje de abscisas de la línea de regresión.

La curva de Phillips empírica (Mirando hacia atrás).

Este es el modelo ARD(4,4) de inflación (1962 – 2004):

$$\widehat{\Delta Inf_t} = \underset{(.44)}{1.30} - \underset{(.08)}{.42}\Delta Inf_{t-1} - \underset{(.09)}{.37}\Delta Inf_{t-2} + \underset{(.08)}{.06}\Delta Inf_{t-3} - \underset{(.08)}{.04}\Delta Inf_{t-4}$$

$$- \underset{(.46)}{2.64}Unem_{t-1} + \underset{(.86)}{3.04}Unem_{t-2} - \underset{(.89)}{0.38}Unem_{t-3} + \underset{(.45)}{.25}Unemp_{t-4}$$

- $\bar{R}^2 = 0.34$ – una gran mejora respecto al AR(4), para el que $\bar{R}^2 = .18$

Ejemplo: *dinf* and *unem* – GRETL

```
Modelo 10: MCO, usando las observaciones 1962:1-2004:4 (T = 172)
Variable dependiente: d_Inf

                 Coeficiente  Desv. Típica  Estadístico t  Valor p
  -----------------------------------------------------------------
  const           1.30429      0.491550         2.653      0.0088
  Unem_1         -2.63557      0.451408        -5.839      2.77e-08
  Unem_2          3.04309      0.873658         3.483      0.0006
  Unem_3         -0.377371     0.893995        -0.4221     0.6735
  Unem_4         -0.248424     0.477438        -0.5203     0.6035
  d_Inf_1  -0.419822     0.0789031       -5.321      3.37e-07
  d_Inf_2  -0.366630     0.0856460       -4.281      3.17e-05
  d_Inf_3   0.0565683    0.0829877        0.6816     0.4964
  d_Inf_4  -0.0364582    0.0745966       -0.4887     0.6257
Media de la vble. dep.  0.017016   D.T. de la vble. dep.   1.707934
Suma de cuad. residuos  316.1094   D.T. de la regresión    1.392595
R-cuadrado              0.366278   R-cuadrado corregido    0.335175
F(8, 163)               11.77630   Valor p (de F)          3.64e-13
Log-verosimilitud      -296.3965   Criterio de Akaike      610.7930
Criterio de Schwarz     639.1205   Crit. de Hannan-Quinn   622.2862
rho                     0.011320   Durbin-Watson           1.976868
```

Contraste de Causalidad de Granger

Ejemplo: Contraste F.

Contrastamos la hipótesis de que los coeficientes de los retardos de la variable desempleo son cero

$F = 10{,}44 > F(4, 163) = 2{,}43$, por lo que la hipótesis nula se rechaza.

El contraste conjunto de que ninguna de las X’s es un predictor útil, después de tener en cuenta los valores pasados (retardos) de *Y*, se llama **contraste de causalidad de Granger .** *Se utiliza el contraste de la F para contrastar la hipótesis nula de que los coeficientes de una de las variables en X* son cero.

“*Causalidad*” es un término poco apropiado aquí: La causalidad de Granger se identifica con la mejora en la capacidad predictiva añadida por la inclusión de la correspondiente variable.

Selección del número de retardos de los modelos AR y ARD utilizando criterios de información

¿Cómo elegir el número de retardos p en un AR(p)?

- El sesgo de variable omitida es irrelevante para predecir!
- Puedes utilizar los contrastes t o F secuencialmente hacia abajo pero los modelos elegidos suelen tener muchos retardos (¿por qué?)
- Un procedimiento alternativo es utilizar un criterio de información.
- Los criterios de información se basan en un equilibrio entre el sesgo (si se cogen pocos retardos) vs. varianza (si se cogen demasiados)
- Los dos más conocidos son el BIC y el AIC.

El criterio de información de Bayes (BIC)

$$\text{BIC}(p) = \ln\left(\frac{SR(p)}{T}\right) + (p+1)\frac{\ln T}{T}$$

- *Primer Término*: siempre decreciente con respecto a p (cuanto mayor p, mejor es el ajuste)
- *Segundo término*: crece con p.
    - La varianza de la predicción debida al error de estimación crece con p – por lo que no es deseable un modelo de predicción con demasiados coeficientes – pero ¿qué significa "demasiados"?
    - Este término es una "penalización" por utilizar más parámetros – y así hacer mayor la varianza de predicción.

- *La minimización del BIC*(p) proporciona un equilibrio entre sesgo y varianza que es mejor en el sentido de que es consistente

$$\hat{p}^{BIC} \xrightarrow{p} p! \quad (SW, App.\ 14.5)$$

Criterio de Información de Akaike *(AIC)*

$$\text{AIC}(p) = \ln\left(\frac{SR(p)}{T}\right) + (p+1)\frac{2}{T}$$

$$\text{BIC}(p) = \ln\left(\frac{SR(p)}{T}\right) + (p+1)\frac{\ln T}{T}$$

La penalización es menor para el *AIC* que para el *BIC* ($2 < \ln T$)

- El *AIC* estima más retardos (mayor *p*) que el *BIC*
- Esto puede ser deseable en algunos casos pero el estimador del número de retardos no es consistente por lo que puede sobreestimar p.

Ejemplo: El modelo AR de inflación, retardos 0 – 6:

# Retardos	BIC	AIC	R^2
0	1.095	1.076	0.000
1	1.067	1.030	0.056
2	**0.955**	0.900	0.181
3	0.957	**0.884**	0.203
4	0.986	0.895	0.204
5	1.016	0.906	0.204
6	1.046	0.918	**0.204**

- El BIC elige 2 retardos, el AIC elige 3 retardos.
- Si se utiliza el R^2 siempre se elige el modelo con más retardos.

Generalización del BIC a modelos multivariantes (ARD)

Sea K = el número total de coeficientes en el modelo (constante, retardos de Y, retardos de X). El BIC es,

$$\mathrm{BIC}(K) = \ln\left(\frac{SR(K)}{T}\right) + K\frac{\ln T}{T}$$

- Calcular considerando todas las combinaciones posibles de retardos de Y y de X (pero esto es mucho)!
- En la práctica tu puedes elegir los retardos de Y con el BIC, y decidir si incluir o no retardos de la X utilizando un contraste de causalidad de Granger con un número fijo de retardos.

DE LA PREDICCIÓN A LA CAUSALIDAD

(Del Capítulo 14 al Capítulo 15)

Efectos Causales Dinámicos y los datos del zumo de naranja

Un efecto causal dinámico es el efecto sobre Y consecuencia de un cambio en X a lo largo del tiempo.

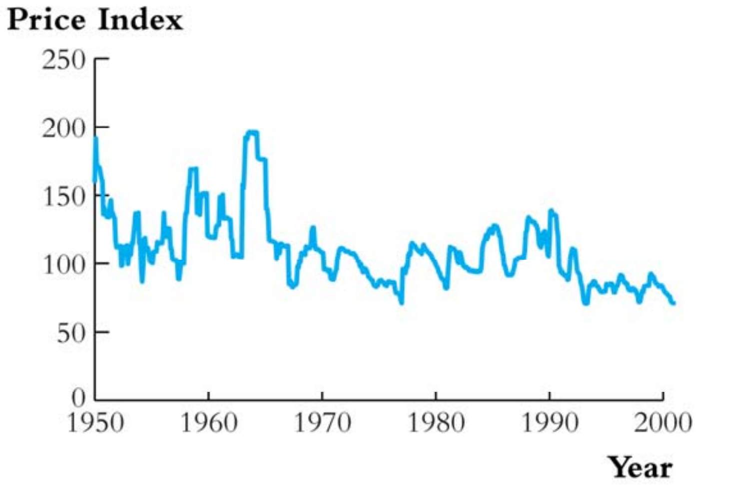
Por ejemplo:

- El efecto de un incremento del impuesto sobre el Tabaco sobre el consumo de cigarrillos este año, el próximo año, dentro de cinco años;
- El efecto del cambio en el tipo de interés del banco central sobre la inflación, este mes, en 6 meses, y en un año,
- El efecto de una helada en Florida sobre el precio del zumo de naranja concentrado en un mes, 2 meses, 3 meses…

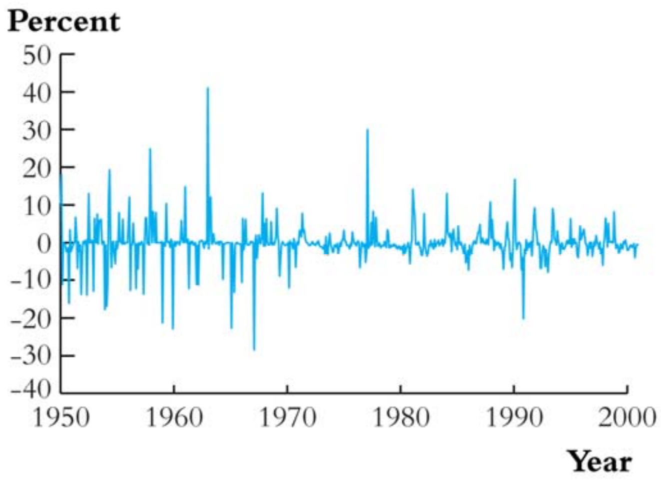
Los datos del zumo de naranja

- Mensuales, Jan. 1950 – Dec. 2000 (T = 612)
- *Precio* = precio de OJ congelado (un subcomponente del índice de precios del productor)
- *%ChgP* = cambio porcentual del precio anualizado, por lo que $\%ChgP_t = 1200\Delta\ln(Price_t)$
- *FDD* = número de grados-día de congelación durante el mes, recogidos en Orlando FL
  - Ejemplo: Si Noviembre tiene dos días con temperaturas bajas < 32^o, una a 30^o y otra a 25^o, entonces FDD_{Nov} = 2 + 7 = 9

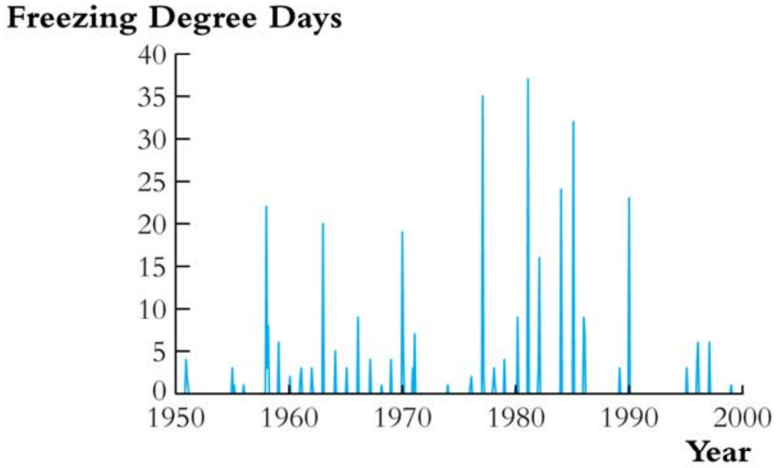
FIGURE 15.1 Orange Juice Prices and Florida Weather, 1950–2000



(a) Price Index for Frozen Concentrated Orange Juice



(b) Percent Change in the Price of Frozen Concentrated Orange Juice



(c) Monthly Freezing Degree Days in Orlando, Florida

Regresión OJ(Zumo de Naranja) inicial

$$\widehat{\%ChgP_t} = -.40 + .47FDD_t$$
$$(.22) \quad (.13)$$

- Relación positiva estadísticamente significativa
- Mayor grados-día de congelación $\Rightarrow$los precios crecen
- Los errores estándar son (HAC) SE's – *luego comentamos*
- Pero ¿cuál es el efecto de *FDD* a lo largo del tiempo?

Efectos causales dinámicos y Modelo de Retardos Distribuidos

El Modelo de Retardos Distribuidos con Efecto Contemporaneo(CRD):

$$Y_t = \beta_0 + \beta_1 X_t + \ldots + \beta_r X_{t-r} + u$$

- β_1 =efecto impacto de un cambio en ***X*** =efecto de un cambio en X_t sobre Y_t, manteniendo constante el pasado de X_t
- β_2 =multiplicador dinámico 1-periodo = efecto de un cambio en X_{t-1} sobre Y_t, manteniendo constantes X_t, X_{t-2},X_{t-3}, …
- β_3 multiplicador dinámico 2-periodo = efecto de un cambio en X_{t-2} sobre Y_t, manteniendo constantes X_t, X_{t-1}, X_{t-3},…
-

- ***Multiplicadores dinámicos acumulados***
- El multiplicador dinámico acumulado (2 periodos) es $\beta_1 + \beta_2 + \beta_3$ =efecto impacto +efecto 1-periodo +efecto 2-periodos

Exogeneidad en la regresión de series temporales

Exogeneidad (pasada y presente)

X es ***exógena si*** $E(u_t|X_t, X_{t-1}, X_{t-2},\ldots) = 0$.

Exogeneidad Estricta (pasada, *presente*, *y futura*)

X es estrictamente ***exógena si*** $E(u_t|\ldots, X_{t+1}, X_t, X_{t-1}, \ldots) = 0$

- La exogeneidad estricta implica la exogeneidad
- Por el momento suponemos que X es exógena –volveremos luego a la exogeneidad estricta.
- Si X es exógena, entonces podemos usar MCO para estimar el efecto causal dinámico de un cambio en X sobre Y.

CC 15.2 Modelo CRD. Supuestos

$$Y_t = \beta_0 + \beta_1 X_t + \ldots + \beta_{r+1} \mathrm{X}_{\mathrm{t}-r} + u_t$$

Hipótesis:

1. $E(u_t|X_t, X_{t-1}, X_{t-2},\ldots) = 0$ (X es exógena)
2. (a) Y y X son estacionarias;
   (b) (Y_t, X_t) y (Y_{t-j}, X_{t-j}) son independientes conforme j se hace grande.
3. Y y X tienen momentos finitos diferentes de cero de orden 8.
4. No hay multicolinealidad perfecta.

Supuestos del Modelo CRD

Las hipótesis 1 y 4 son habituales

- La hipótesis 3 es familiar excepto lo del orden 8 que tiene que ver con el estimador HAC
- La hipótesis 2 es diferente – anteriormente era que (X_i, Y_i) fuera i.i.d. –pero con datos de series temporales la cuestión se hace más complicada.

2. (a) Y y X tienen distribuciones estacionarias;
    - Si es así, los coeficientes no cambian dentro de la muestra (validez interna);
    - Y los resultados pueden ser extrapolados fuera de la muestra (validez externa).
    - Esto es en series temporales la contrapartida del i.i.d.

Supuestos del Modelo CRD

2. (b) (Y_t, X_t) y (Y_{t-j}, X_{t-j}) se hacen independientes conforme j se hace muy grande

- Intuitivamente, esto dice que hemos separado experimentos en periodos muy distantes
- En datos de sección cruzada, hemos asumido que Y y X eran i.i.d., una consecuencia del muestreo aleatorio simple – esto conduce al CLT.(Teorema Central del Limite)
- Una versión del CLT se da si las variables de series temporales se hacen independientes conforme la separación temporal crece – La hipótesis 2(b) es la contrapartida en series temporales a la parte "distribuidos independientemente" del i.i.d.

Bajo las hipótesis comentadas:

- Los MCO proporcionan estimadores **consistentes** de β_1, $\beta_2,\ldots,\beta_r$ (de los multiplicadores dinámicos)
- En muestras grandes, la distribución muestral de $\hat{\beta}_1$, etc., es normal
- ***PERO*** la formula de la varianza de esta distribución muestral no es la usual de la sección cruzada (i.i.d.) , porque u_t no es i.i.d. – u_t puede estar serialmente correlacionada!
- Esto significa que los errores estándar habituales están equivocados!
- Necesitamos estimaciones de los EE que sean robustas a la autocorrelación y a la heterocedasticidad.

Errores Estándar consistentes en presencia de heterocedasticidad y Autocorrelación

- Cuando u_t esta correlacionada serialmente, la varianza de la distribución muestral de los MCO depende de esta correlación serial.
- La fórmula habitual para estimar el error estándar robusto a la heterocedasticidad solo vale para errores no correlacionados.
- Necesitamos nuevos estimadores de los EE que sean consistentes y robustos ante la heterocedasticidad y la correlación serial. (HAC)

Errores Estandar *HAC* .

Las mates para un solo regresor X_t:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

El estimador MCO:

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})u_t}{\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

$$\cong \frac{\frac{1}{T}\sum_{t=1}^{T} v_t}{\sigma_X^2} \text{ (en muestras grandes)}$$

donde $v_t = (X_t - \bar{X})u_t$.

Errores Estandar *HAC*

Así, en muestras grandes,

$$\operatorname{var}(\hat{\beta}_1) = \operatorname{var}\left(\frac{1}{T}\sum_{t=1}^{T} v_t\right)/(\sigma_X^2)^2$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\operatorname{cov}(v_t, v_s)/(\sigma_X^2)^2$$

En los datos de sección cruzada i.i.d., $\operatorname{cov}(v_t, v_s) = 0$ for $t \neq s$,

por lo que

$$\operatorname{var}(\hat{\beta}_1) = \frac{1}{T^2}\sum_{t=1}^{T}\operatorname{var}(v_t))/(\sigma_X^2)^2 = \frac{\sigma_v^2}{T(\sigma_x^2)^2}$$

Este es el resultado habitual de sección cruzada.

Errores Estandar *HAC*.

Pero en series temporales, $\mathrm{cov}(v_t, v_s) \neq 0$ en general.

Considerar $T = 2$:

$$\mathrm{var}\left(\frac{1}{T}\sum_{t=1}^{T} v_t\right) = \mathrm{var}[\tfrac{1}{2}(v_1+v_2)]$$

$$= \tfrac{1}{4}[\mathrm{var}(v_1) + \mathrm{var}(v_2) + 2\mathrm{cov}(v_1,v_2)]$$

$$= \tfrac{1}{2}\sigma_v^2 + \tfrac{1}{2}\rho_1\sigma_v^2 \qquad (\rho_1 = \mathrm{corr}(v_1,v_2))$$

$$= \tfrac{1}{2}\sigma_v^2 \times f_2, \text{ donde } f_2 = (1+\rho_1)$$

- En los datos i.i.d. , $\rho_1 = 0$ por lo que $f_2 = 1$, llegándose a la fórmula habitual.
- En series temporales, si $\rho_1 \neq 0$ entonces $\mathrm{var}(\hat{\beta}_1)$ no es la fórmula habitual.

Expresion para var($\hat{\beta}_1$), general *T*

$$\text{var}\left(\frac{1}{T}\sum_{t=1}^{T} v_t\right) = \frac{\sigma_v^2}{T} \times f_T$$

por lo que

$$\boxed{\text{var}(\hat{\beta}_1) = \left[\frac{1}{T}\frac{\sigma_v^2}{(\sigma_X^2)^2}\right] \times f_T}$$

donde

$$\boxed{f_T = 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j} \quad (\textit{SW, eq. (15.13)})$$

- Los convencionales MCO EE's están equivocados cuando u_t está serialmente correlacionado.
- El error de los MCO EE radica en el factor f_T
- Necesitamos un estimador del EE diferente!!!

Errores Estándar *HAC*

- Los errores estándar que usan estimadores consistentes de f_T se llaman **H**eteroskedasticity- and **A**utocorrelation-**C**onsistent (**HAC**) standard errors (Errores Estándar consistentes en presencia de heterocedasticidad y autocorrelación).
- En la literatura hay diferentes procedimientos para estimar f_T.

Ejemplo:Precios OJ heladas(gretl)

Modelo 2: MCO, usando las observaciones 1950:01-2000:12 (T = 612)
Variable dependiente: %ChgP

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	-0.402256	0.198392	-2.028	0.0430	**
FDD	0.466218	0.0587115	7.941	9.69e-015	***

Media de la vble. dep.	-0.115822	D.T. de la vble. dep.	5.065299
Suma de cuad. residuos	14207.89	D.T. de la regresión	4.826139
R-cuadrado	0.093687	R-cuadrado corregido	0.092202
F(1, 610)	63.05686	Valor p (de F)	9.69e-15
Log-verosimilitud	-1830.705	Criterio de Akaike	3665.411
Criterio de Schwarz	3674.244	Crit. de Hannan-Quinn	3668.846
rho	0.106701	Durbin-Watson	1.778019

Modelo 4: MCO, usando las observaciones 1950:01-2000:12 (T = 612)
Variable dependiente: %ChgP (GRETL)

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	-0.650518	0.217728	-2.988	0.0029	***
FDD	0.469312	0.0586812	7.998	6.49e-015	***
FDD_1	0.143051	0.0586647	2.438	0.0150	**
FDD_2	0.0564233	0.0586356	0.9623	0.3363	
FDD_3	0.0722595	0.0586409	1.232	0.2183	
FDD_4	0.0343244	0.0586356	0.5854	0.5585	
FDD_5	0.0468222	0.0586647	0.7981	0.4251	
FDD_6	0.0481116	0.0586812	0.8199	0.4126	

Media de la vble. dep.	-0.115822	D.T. de la vble. dep.	5.065299
Suma de cuad. residuos	13980.81	D.T. de la regresión	4.811136
R-cuadrado	0.108173	R-cuadrado corregido	0.097837
F(7, 604)	10.46587	Valor p (de F)	1.93e-12
Log-verosimilitud	-1825.775	Criterio de Akaike	3667.550
Criterio de Schwarz	3702.884	Crit. de Hannan-Quinn	3681.293
rho	0.098782	Durbin-Watson	1.79331

Estimación de Efectos Causales Dinámicos con regresores estíictamente exógenos

- X es estríctamente exógena si $E(u_t|\ldots,X_{t+1}, X_t, X_{t-1}, \ldots) = 0$
- Si X es estrictamente exógena, entonces hay otras formas más eficientes de estimar los efectos causales dinámicos que la vista en base al modelo de retardos distribuidos:
  - Generalized Least Squares (GLS) (Mínimos Cuadrados Generalizados; MCG)
  - Autorregresivo con retardos distribuidos(CARD)
- Pero la condición de exogeneidad estricta es muy fuerte y difícil de contrastar.
- Por esta razón no vamos a desarrollar estos dos métodos en detalle. Ver la Sección 15.5 del libro.

Modelo CARD

Suponer un modelo simple

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t$$

Además, suponemos que

$$u_t = \phi_1 u_{t-1} + \tilde{u}_t$$

Introduciendo la primera en la segunda,

$$Y_t - \beta_0 - \beta_1 X_t - \beta_2 X_{t-1} = \phi_1 \left(Y_{t-1} - \beta_0 - \beta_1 X_{t-1} - \beta_2 X_{t-2} \right) + u_t$$

de donde se obtiene el modelo CARD

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + u_t$$

Con

$$\alpha_0 = \beta_0 \left(1 - \phi_1 \right), \delta_0 = \beta_1, \delta_1 = \beta_2 - \phi_1 \beta_1, \phi_2 = -\phi_1 \beta_2$$

¿Necesito los errores HAC cuando estimo un AR o un CARD?

A: NO.

- El problema ya no existe en estos modelos porque la u_t no está serialmente correlacionada por lo que los errores MCO son correctos.
- En los modelos AR y CARD los errores no están serialmente correlacionados si se incluyen suficientes retardos de Y
  - Si incluyes suficientes retardos de Y, entonces el error no puede ser predicho utilizando valores pasados de Y, o equivalentemente con el pasado de u – por lo que u no tiene autocorrelación.

Discusión: Cálculo de los multiplicadores acumulados y sus errores estándar

Los multiplicadores acumulados pueden ser estimados primero estimando el modelo de retardos distribuídos y, después, sumando los coeficientes. Sin embargo, es más fácil calcularlos de forma directa a partir de una versión modificada del modelo.

Calculando los multiplicadores acumulados

Un truco consiste en reescribir el modelo de forma que los coeficientes en la nueva regresión son los coeficientes de interés – aquí, los multiplicadores acumulados.

Ejemplo: Rescribir el modelo de retardos distribuidos con un retardo:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t \\ &= \beta_0 + \beta_1 X_t - \beta_1 X_{t-1} + \beta_1 X_{t-1} + \beta_2 X_{t-1} + u_t \\ &= \beta_0 + \beta_1 (X_t - X_{t-1}) + (\beta_1 + \beta_2) X_{t-1} + u_t \end{aligned}$$

o

$$Y_t = \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2) X_{t-1} + u_t$$

Calculando los multiplicadores acumulados

Sea $W_{1t} = \Delta X_t$ y $W_{2t} = X_{t-1}$ y estimar la regresión,

$$Y_t = \beta_0 + \delta_1 W_{1t} + \delta_2 W_{2t} + u_i$$

Entonces

$\delta_1 = \beta_1$ = efecto impacto

$\delta_2 = \beta_1 + \beta_2$ = el primer multiplicador acumulado

y los errores estándar HAC de δ_1 y δ_2 son los errores estándar para los dos multiplicadores acumulados.

Calculando los multiplicadores acumulados

En general, el modelo ARD puede ser reescrito como,

$$Y_t = \delta_0 + \delta_1 \Delta X_t + \delta_2 \Delta X_{t-1} + \ldots + \delta_{q-1} \Delta X_{t-q+1} + \delta_q \Delta X_{t-q} + u_t$$

donde

$$\delta_1 = \beta_1$$

$$\delta_2 = \beta_1 + \beta_2$$

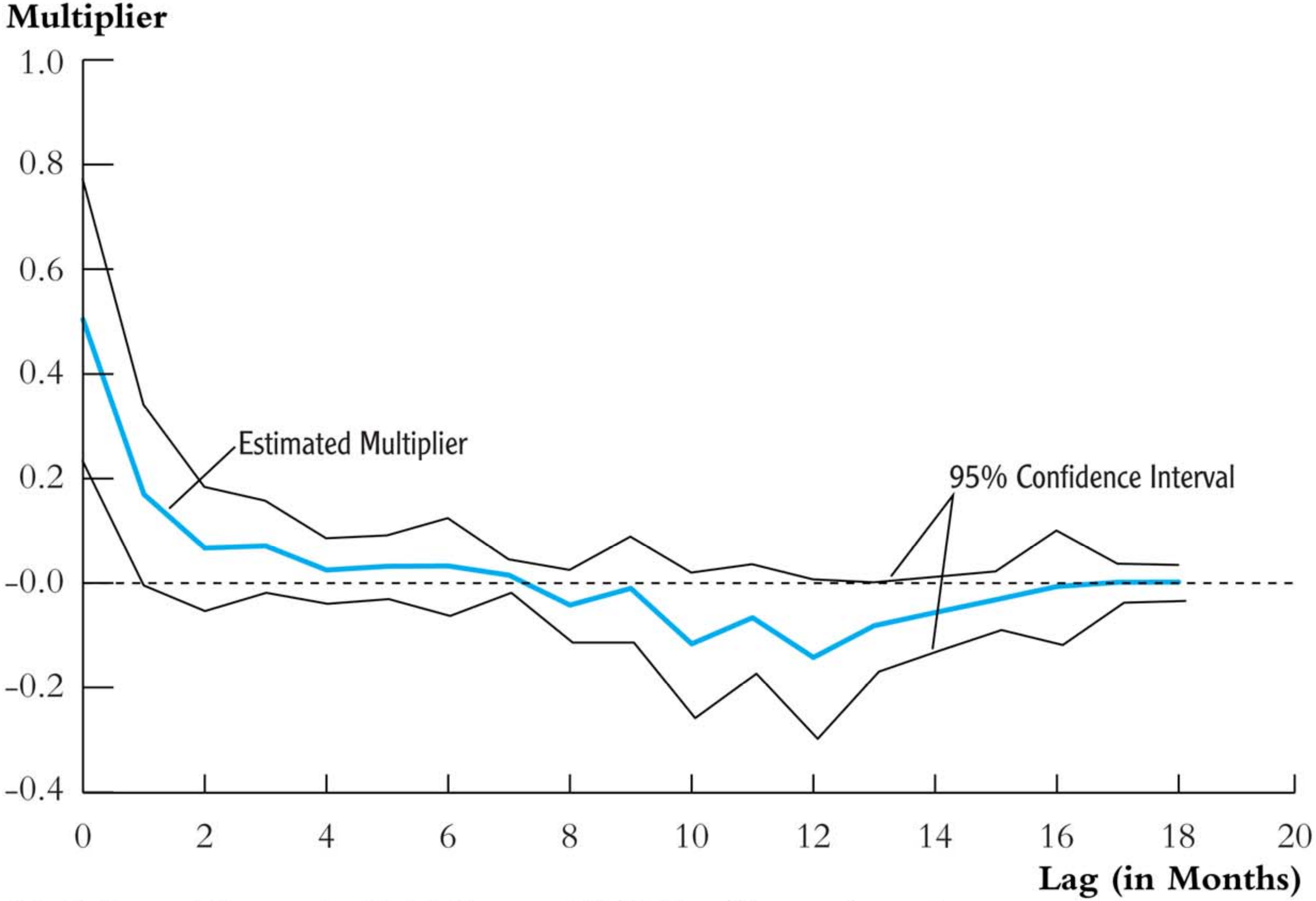
$$\delta_3 = \beta_1 + \beta_2 + \beta_3$$

$$\ldots$$

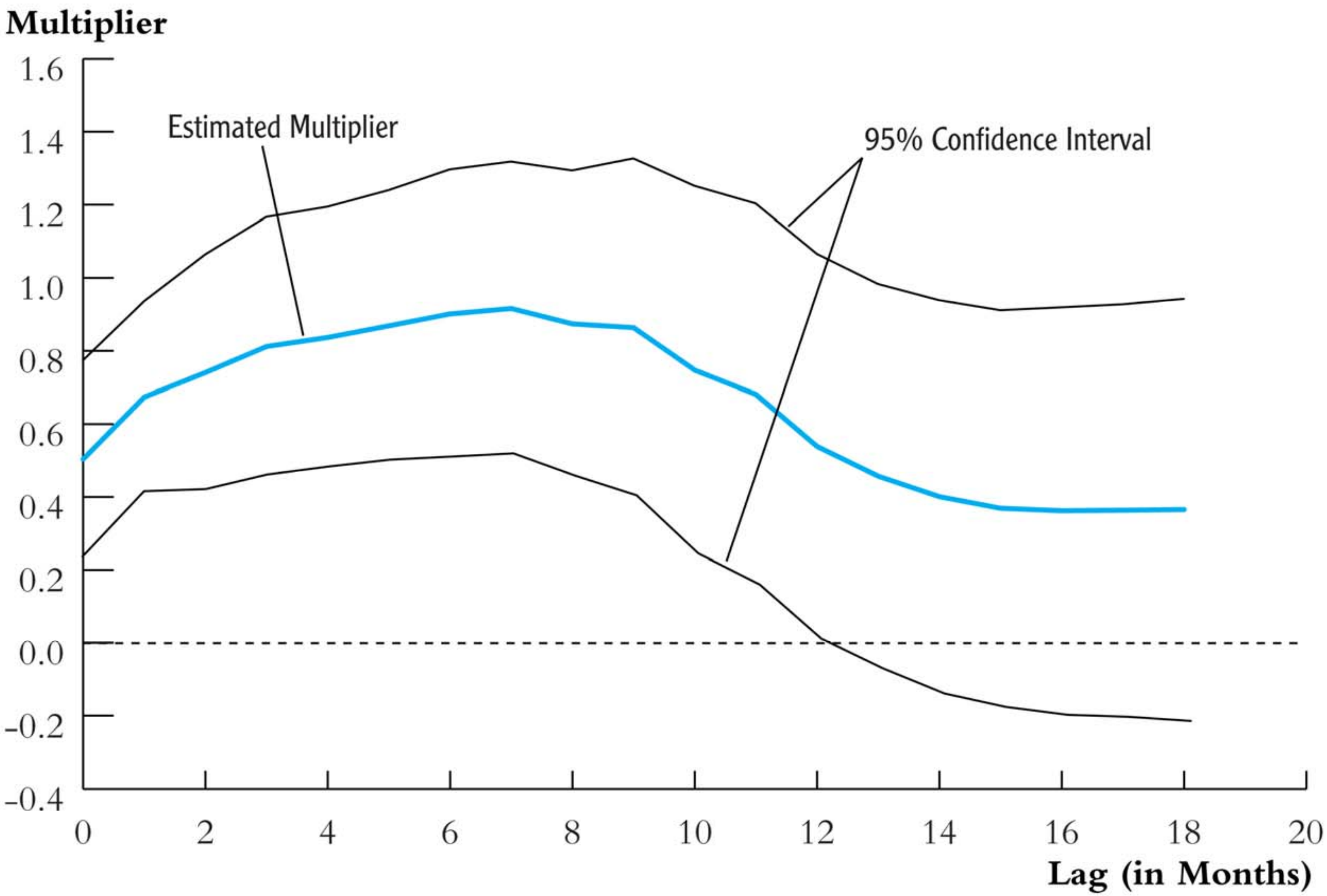
$$\delta_q = \beta_1 + \beta_2 + \ldots + \beta_q$$

Los multiplicadores acumulados y sus EE HAC pueden ser calculados directamente utilizando esta regresión transformada.

FIGURE 15.2 The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice



(a) Estimated Dynamic Multipliers and 95% Confidence Interval



(b) Estimated Cumulative Dynamic Multipliers and 95% Confidence Interval

Son los efectos dinámicos de OJ estables?

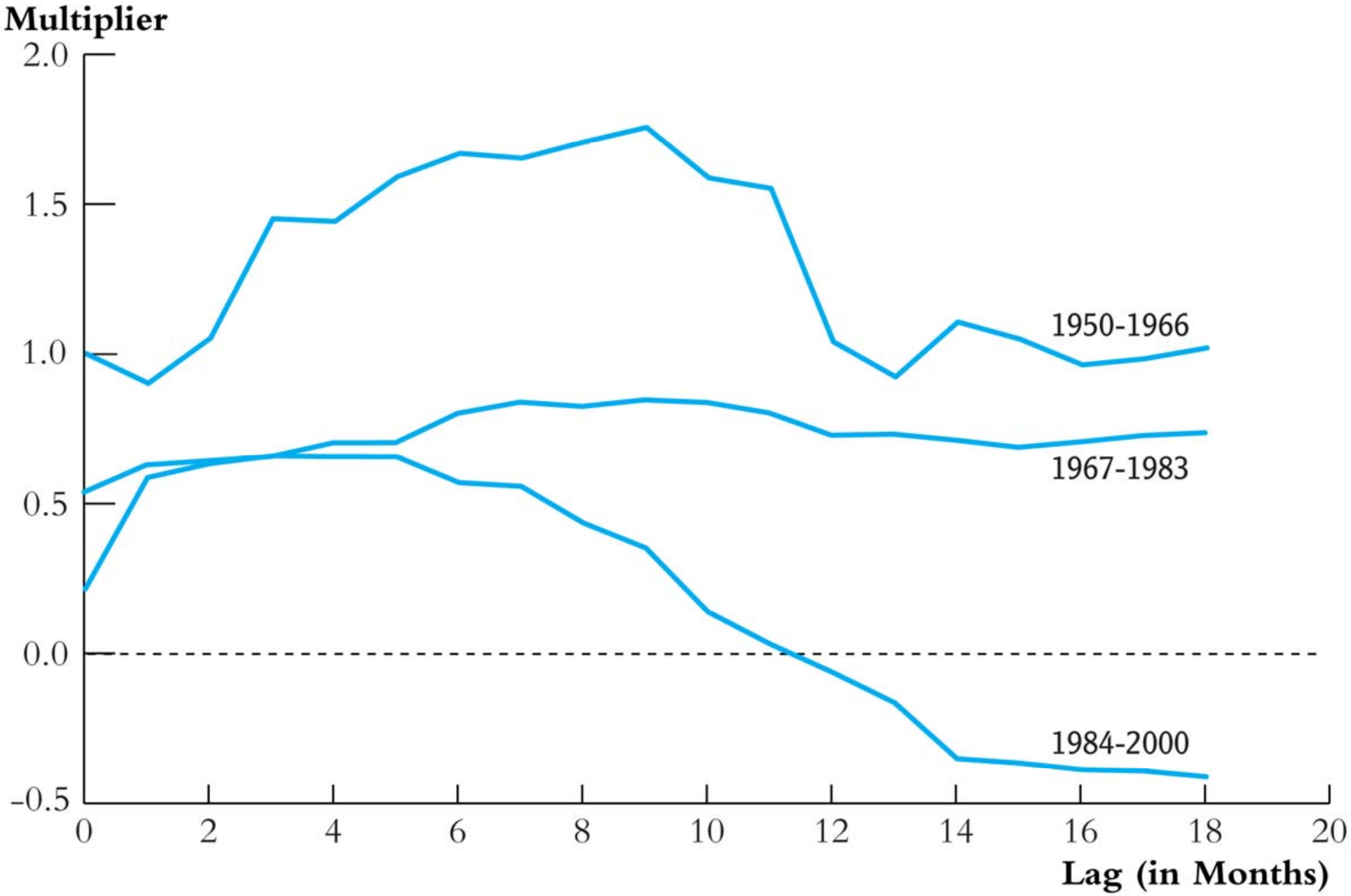
Recordar que la estabilidad se puede contrastar utilizando el estadístico QLR . Podemos calcular QLR para la regresión en (1) en la Tabla 15.1:

- *¿Son necesarios los EE HAC?¿Por qué o por qué no?*
- *¿Cómo calcularías el estadístico de Chow?*
- *¿Cómo calcularías el estadístico QLR ?*
- *¿Cuáles son los grados de libertad de los estadísticos de Chow y QLR?*
- Resultado: ***QLR* = 21.19**.
  - ¿Es significativo? (Ver Tabla 14.6)
  - ¿A que nivel de significación?
- ¿Cómo interpretar el resultado? Estimar los multiplicadores dinámicos en submuestras y ver cómo cambian en el tiempo.

TABLE 14.6 Critical Values of the QLR Statistic with 15% Trimming

Number of Restrictions (q)	10%	5%	1%
1	7.12	8.68	12.16
2	5.00	5.86	7.78
3	4.09	4.71	6.02
4	3.59	4.09	5.12
5	3.26	3.66	4.53
6	3.02	3.37	4.12
7	2.84	3.15	3.82
8	2.69	2.98	3.57
9	2.58	2.84	3.38
10	2.48	2.71	3.23
11	2.40	2.62	3.09
12	2.33	2.54	2.97
13	2.27	2.46	2.87
14	2.21	2.40	2.78
15	2.16	2.34	2.71
16	2.12	2.29	2.64
17	2.08	2.25	2.58
18	2.05	2.20	2.53
19	2.01	2.17	2.48
20	1.99	2.13	2.43

OJ: Do the breaks matter substantively?



¿El efecto acumulado del *FDD* baja con el tiempo? ¿Por qué?

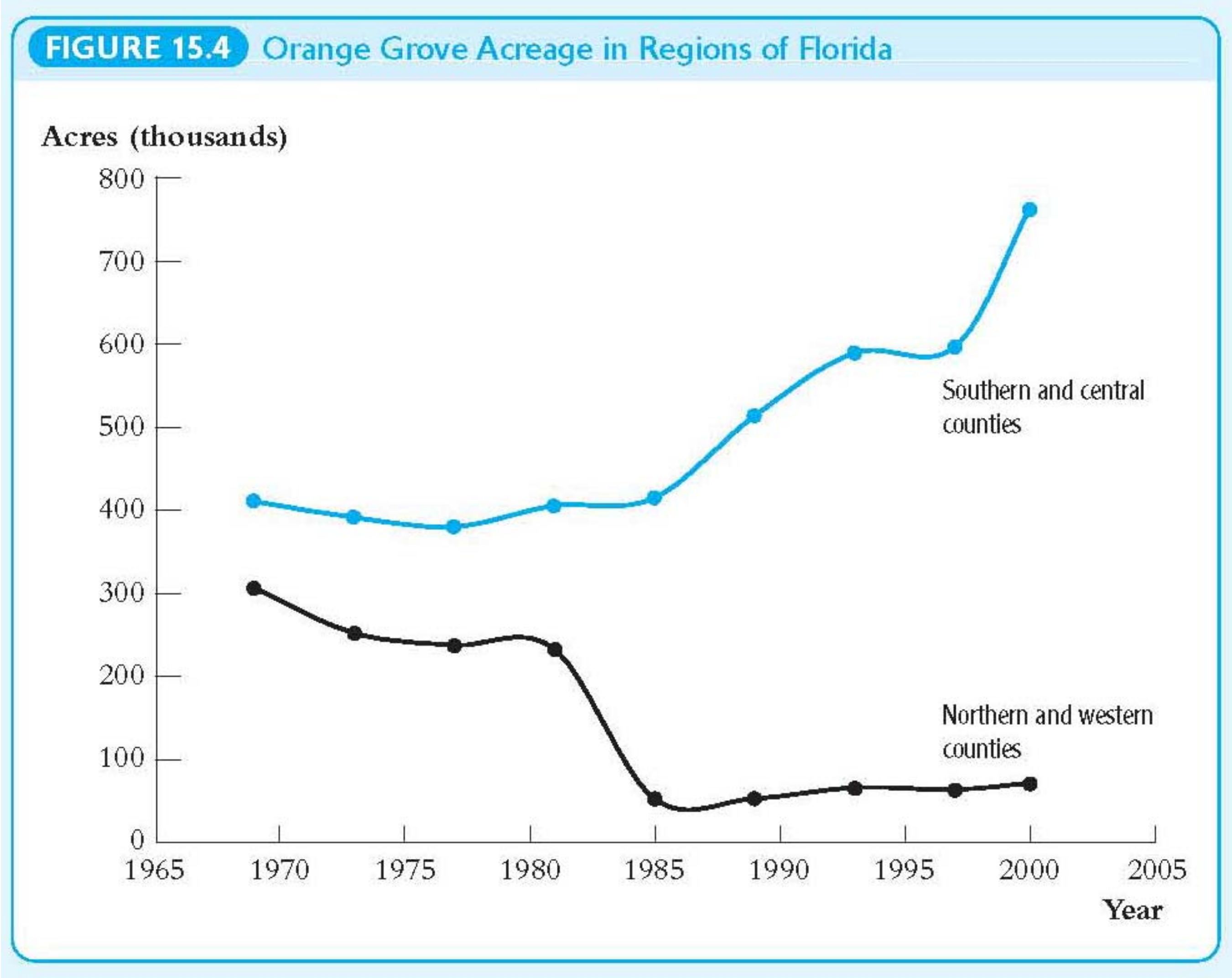


FIGURE 15.4 Orange Grove Acreage in Regions of Florida
Acres (thousands)
800
700
600
500
400
300
200
100
0
1965
1970
1975
1980
1985
1990
1995
2000
2005
Year
Southern and central counties
Northern and western counties

Hecho: Después de perder muchos árboles por las heladas en el norte, los agricultores de Florida incrementaron las plantaciones en el sur. ¿Cómo se relaciona esto con el cambio en las respuestas acumuladas?

¿Es la exogeneidad plausible? Algunos Ejemplos

Si X is exógena (y las hipótesis #2-4se cumplen), entonces un modelo con retardos distribuidos proporciona estimadores consistentes de los efectos causales dinámicos.

Como ocurría con los datos de sección cruzada tenemos que ser críticos acerca de si X es exógena en cualquier aplicación:

- ¿Es X exógena, i.e. $E(u_t|X_t, X_{t-1}, \ldots) = 0$?
- ¿es X estríctamente exógena, i.e. $E(u_t|\ldots, X_{t-1}, X_t, X_{t+1}, \ldots) = 0$?

En los siguientes ejemplos, ¿es (a) exógena y/o (b) estrictamente exógena plausible? ¿Qué piensas?

1. *Y* = OJ precios, *X* = *FDD* en Orlando
2. *Y* = exportaciones australianas, *X* = US GDP (efecto de la renta en USA sobre la demanda de exportaciones de Australia)
3. *Y* = UE exportaciones, *X* = US GDP (efecto de la renta en USA sobre las exportaciones de la UE)
4. *Y* = US tasa de inflación, *X* = cambio porcentual en los precios del petróleo (fijados por la OPEC) (efecto de un cambio en el precio del petróleo sobre la inflación)
5. *Y* = GDP crecimiento, *X* =tipo de la Reserva Federal(el efecto de la política monetaria sobre el crecimiento del output)

6. Y = cambio en la tasa de inflación , X = desempleo (la curva de Phillips)

Exogeneidad

- Hay que evaluar la exogeneidad y la exogeneidad estricta caso por caso.
- La exogeneidad a menudo no es plausible en datos de series temporales debido a la causalidad simultánea.
- La causalidad estricta es muy poco plausible con datos de series temporales debido a las interacciones.

Conclusiones

- Cuando X es exógena, puedes estimar consistentemente los efectos causales dinámicos utilizando un modelo con retardos distribuidos.
- Si u tiene correlación serial, los EE MCO convencionales son incorrectos; tienes que usar los EE HAC.
- Para decidir si X es exógena, pensar en profundidad lo específico del problema!

En **Econometría es muy difícil justificar la hipótesis de exogeneidad (estricta o no) en el modelo CRD.**

Se han propuesto contrastes de exogeneidad-el más conocido es el de HAUSMAN- pero ninguno de ellos está libre de serios problemas.

Eso explica que la solución que parece más razonable sea la utilización del modelo VAR.

CC 16.1 Vectores Autorregresivos. Modelo VAR

Un vector autorregresivo (VAR) es un conjunto de k regresiones de series temporales en las que los k regresores son los valores retardados de las k series. Un VAR extiende el modelo autorregresivo univariante a una lista, o vector, de variables de series temporales. Cuando el número de retardos en cada una de las ecuaciones es el mismo y es igual a p, el sistema de ecuaciones se denomina VAR(p).

En el caso de dos variables de series temporales, $Y_t \; y \; X_t$, el VAR(p) consta de dos ecuaciones

$$Y_t = \beta_{10} + \beta_{11} Y_{t-1} + \ldots + \beta_{1p} Y_{t-p} + \gamma_{11} X_{t-1} + .. + \gamma_{1p} X_{t-p} + u_{1t}$$

$$X_t = \beta_{20} + \beta_{21} Y_{t-1} + \ldots + \beta_{2p} Y_{t-p} + \gamma_{21} X_{t-1} + .. + \gamma_{2p} X_{t-p} + u_{2t}$$

Donde las $\beta \; y \; las \; \gamma$ son coeficientes desconocidos y $u_{1t} y \, u_{2t}$ son los términos de error.

Los supuestos del VAR son los supuestos de regresión de series temporales de **CC 14.6,** aplicados a cada una de las ecuaciones. Los coeficientes de un VAR se calculan mediante la estimación de cada una de las ecuaciones por MCO.

Modelo VAR

Para utilizar el modelo VAR hay que prestar atención a cinco aspectos:

Estimación

a. Determinación del número de retardos
b. Análisis de permanencia estructural
c. Predicción
d. Causalidad

a. Estimación

Todas las relaciones del modelo VAR tienen los mismos regresores. Por esa razón, las estimaciones MCO son las mismas utilizando todo el sistema o utilizando cada relación por separado.

Tomando una sola relación, si se cumplen los supuestos recogidos en el CC 14.6, que hemos comentado al comienzo del tema, entonces los estimadores MCO son consistentes.

Como recordatorio:

CC 14.6 Regresión de series temporales con varios predictores.

El modelo general de regresión de series temporales permite k predictores adicionales, en el que se incluyen q_1 retardos del primer predictor, q_2 retardos del segundo predictor, y así sucesivamente:

$$
\begin{aligned}
Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \\
\delta_{11} X_{1t-1} + \ldots\ldots + \delta_{1q_1} X_{1t-q_1} + \\
+ \ldots. + \delta_{k1} X_{kt-1} + \ldots\ldots + \delta_{kq_k} X_{kt-q_k} + u_t
\end{aligned}
$$

Donde los supuestos son

1.

$$E\left(u_t \,/\, Y_{t-1}, Y_{t-2}, \ldots X_{1t-1}, X_{1t-2} \ldots.. X_{kt-1}, X_{kt-2}, \ldots\right)$$

2. (a) Las variables aleatorias $\left(Y_t, X_{1t}, \ldots. X_{kt}\right)$ son conjuntamente estacionarias.

(b) $\left(Y_t, X_{1t}, \ldots. X_{kt}\right)$ y $\left(Y_{t-j}, X_{1t-j}, \ldots. X_{kt-j}\right)$ se hacen independientes cuando j se hace grande.

3. Los valores extremos elevados son poco probables

4. No existe multicolinealidad perfecta.

b. Determinación del número de retardos.

Utilizaremos una doble vía:

- Criterios de Información
- Análisis de Esfericidad

Criterios de Información

Anteriormente, hemos comentado la justificación y la forma que adoptaban, para un modelo ARD, algunos criterios de información como el AIC y el BIC. Estos criterios pueden extenderse a un marco multiecuacional. La forma que adoptan es:

$$AIC(p) = \ln\left|\hat{\Sigma}_u\right| + k(kp+1)\frac{2}{T}$$

$$BIC(p) = \ln\left|\hat{\Sigma}_u\right| + k(kp+1)\frac{\ln(T)}{T}$$

En donde, p es el número de retardos y k es el número de variables. $\hat{\Sigma}_u$ es la estimación de la matriz de varianzas y covarianzas del vector de errores, siendo su elemento i,j-ésimo $\frac{1}{T}\sum_1^T \hat{u}_{it}\hat{u}_{jt}$ en donde $\hat{u}_{it}$ es el residuo MCO de la ecuación i-ésima y $\hat{u}_{jt}$ es el residuo de la ecuación j-ésima.

Esfericidad

Se trata de utilizar los procedimientos adecuados para estar seguros a la luz de la evidencia disponible que el modelo no tiene ni autocorrelación, ni herocedasticidad ni asimetrías o apuntamientos poco ajustados a una distribución normal.

c. Análisis de Permanencia Estructural

Seguiremos el mismo esquema comentado en el Tema 6, distinguiendo si el periodo de posible ruptura es conocido o no.

Periodo de ruptura conocido: Contraste de Chow

Periodo de ruptura desconocido: contraste QLR

d.Predicción

Se trata de utilizar la evidencia disponible para informar sobre acontecimientos futuros. Se dispone de T observaciones muestrales y se va a predecir el valor de la variable dependiente en el periodo T+h siendo h el horizonte temporal de predicción. Distinguiremos dos tipos de predicción:

- Predicciones multiperiodo iteradas (GRETL las llama "Dinámicas")
- Predicciones multiperiodo no iteradas (GRETL las llama "Estáticas")

Predicciones Dinámicas

Suponemos un modelo VAR(4) con solo dos variables. Primero, estimamos los parámetros del modelo utilizando las T observaciones de las dos variables. Las predicciones para sucesivos horizontes de predicción son:

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\beta}_2 Y_{T-1} + \hat{\beta}_3 Y_{T-2} + \hat{\beta}_4 Y_{T-3} + \hat{\gamma}_1 X_T + \hat{\gamma}_2 X_{T-1} + \hat{\gamma}_3 X_{T-2} + \hat{\gamma}_4 X_{T-3}$$

$$\hat{Y}_{T+2|T} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{T+1|T} + \hat{\beta}_2 Y_T + \hat{\beta}_3 Y_{T-1} + \hat{\beta}_4 Y_{T-2} + \hat{\gamma}_1 \hat{X}_{T+1|T} + \hat{\gamma}_2 X_T + \hat{\gamma}_3 X_{T-1} + \hat{\gamma}_4 X_{T-2}$$

$$\hat{Y}_{T+3|T} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{T'+2|T} + \hat{\beta}_2 \hat{Y}_{T+1|T} + \hat{\beta}_3 Y_T + \hat{\beta}_4 Y_{T-1} + \hat{\gamma}_1 \hat{X}_{T+2|T} + \hat{\gamma}_2 \hat{X}_{T+1|T} + \hat{\gamma}_3 X_T + \hat{\gamma}_4 X_{T-1}$$

Predicciones Estáticas

Suponemos el mismo modelo. La diferencia con las dinámicas consiste en utilizar en la parte de la derecha no las predicciones generadas en etapas anteriores sino los valores reales de las variables. Pero eso no se puede hacer si se trata de una verdadera predicción porque los valores futuros no son conocidos. Por eso se habla de una pseudo-predicción. Del total de las T observaciones, se eliminan las P últimas, y con el resto S=T-P se estima el modelo. A continuación, se utiliza el modelo estimado para llevar a cabo la predicción de los periodos excluidos. La forma que adoptarían las predicciones sería la siguiente

Predicciones Estáticas

$$\hat{Y}_{S+1|S} = \hat{\beta}_0 + \hat{\beta}_1 Y_S + \hat{\beta}_2 Y_{S-1} + \hat{\beta}_3 Y_{S-2} + \hat{\beta}_4 Y_{S-3} + \hat{\gamma}_1 X_S + \hat{\gamma}_2 X_{S-1} + \hat{\gamma}_3 X_{S-2} + \hat{\gamma}_4 X_{S-3}$$

$$\hat{Y}_{S+2|S} = \hat{\beta}_0 + \hat{\beta}_1 Y_{S+1} + \hat{\beta}_2 Y_S + \hat{\beta}_3 Y_{S-1} + \hat{\beta}_4 Y_{S-2} + \hat{\gamma}_1 X_{S+1} + \hat{\gamma}_2 X_S + \hat{\gamma}_3 X_{S-1} + \hat{\gamma}_4 X_{S-2}$$

$$\hat{Y}_{S+3|S} = \hat{\beta}_0 + \hat{\beta}_1 Y_{S+2} + \hat{\beta}_2 Y_{S+1} + \hat{\beta}_3 Y_S + \hat{\beta}_4 Y_{S-1} + \hat{\gamma}_1 X_{S+2} + \hat{\gamma}_2 X_{S+1} + \hat{\gamma}_3 X_S + \hat{\gamma}_4 X_{S-1}$$

y así sucesivamente.

e. Causalidad

Lo que sigue es una parte de “Usos de los modelos econométricos”

Decimos que la variable *x* causa a la variable *y*, si controlando la variable *x* podemos controlar la variable *y*. Y esto no solo cualitativamente sino también cuantitativamente; es decir si podemos determinar la magnitud del cambio inducido en *y* como consecuencia de un cambio unitario en *x*.

Podemos decir que en un grupo de variables relacionadas con un fenómeno seguramente que existen relaciones de causalidad entre la mayor parte de la parejas de variables que se pueden formar, unas en sentido más fuerte y otras de tipo más débil. Pero el problema es aflorar, sacar a la superficie esas relaciones causales porque, en principio, todo determina a todo y la causalidad bivariante queda camuflada ya que somos observadores pasivos.

La forma en que se ha resuelto este problema en otras ciencias como la medicina y la agricultura ha sido mediante experimentos aleatorizados. De forma aleatoria se distribuyen las unidades objeto de estudio en dos grupos de forma que la composición de ambos es muy similar. A continuación se aplica un estímulo a las unidades de uno de los grupos y la causalidad podrá determinarse comparando los resultados de ambos grupos. El ejemplo tipo son dos parcelas de cultivo que son vecinas y, por lo tanto, tienen las mismas características de suelo y condiciones meteorológicas. A una se aplica fertilizante y a la otra no. La causalidad vendrá dada por los diferentes rendimientos obtenidos en las dos parcelas.

Pero estos ejercicios son difíciles de llevar a cabo en Economía, porque somos observadores pasivos y los datos nos vienen dados.

Para resolver estos problemas la Econometría ha desarrollado dos grandes estrategias, la primera que va de lo específico a lo general y la segunda que va de lo general a lo específico. En la primera estrategia, se parte del modelo más simple que pone en relación las dos variables cuya causalidad se estudia. Según sean los resultados que se van obteniendo se van incorporando nuevas variables como regresores hasta que se llega a una especificación en la que la estimación del parámetro de interés se le puede asignar una interpretación causal. Los conceptos de exogeneidad estricta y variable instrumental juegan un papel clave para poder hacer esa interpretación causal. Una descripción completa de esta estrategia puede encontrarse en el libro de Stock y Watson (2012).

En lo que respecta a la segunda estrategia, la que propugna ir de lo general a lo específico, el punto de partida es una modelización estocástica sin restricciones de un conjunto de variables. En esta primera etapa el objetivo es que los datos hablen por si mismos sin ningún tipo de limitaciones. En las etapas sucesivas se trata de ver si restricciones con contenido económico sugeridas por la Teoría Económica no distorsionan de forma relevante y pueden ser incorporadas. En esta estrategia el modelo VAR y la causalidad en el sentido de Granger son dos referencias básicas. Ver Juselius (2006).

Veamos ahora ambas estrategias con un poco más de detalle.

Como hemos dicho, en la primera estrategia se parte del modelo simple entre las dos variables cuya causalidad se quiere estudiar

$$y_t = \beta x_t + u_t \qquad \text{t=1,2…T} \qquad (1)$$

Si la variable x_t es independiente de la perturbación aleatoria, es estrictamente exógena, entonces a la estimación MCO de β se le puede asignar una interpretación causal. Si la influencia es pequeña entonces estadísticamente no será significativa y no podremos detectar la causalidad. Si es una influencia causal grande entonces podría encontrarse una relación causal significativa pero entonces el problema es como saber que la variable x es estrictamente exógena. Como hemos visto hay muchas razones que pueden poner en duda la exogeneidad del regresor en (1). Los cinco casos en los que el regresor no es estrictamente exógeno son: variable omitida, forma funcional errónea, sesgo de selección muestral relaciones simultáneas y errores en variables. La solución parece girar en torno al uso de variables instrumentales. Pero el uso de estas variables plantea similares problemas porque tienen que relacionarse con la variable de interés y, al mimo tiempo, ser independientes de la perturbación. ¿De donde sacamos la información sobre la independencia? Solo queda el recurso de la información a priori.

Si consideramos el segundo enfoque, **de lo general a lo específico,** el punto de partida es un modelo VAR. Se supone en principio una forma estructural que toma la forma siguiente,

$$y_t = \theta x_t + \beta_{11} y_{t-1} + \beta_{12} x_{t-1} + \varepsilon_{1t} \qquad (2)$$

$$x_t = \gamma y_t + \beta_{21} y_{t-1} + \beta_{22} x_{t-1} + \varepsilon_{2t} \qquad (3)$$

A continuación, obtenemos la forma reducida

$$y_t = \pi_{11} y_{t-1} + \pi_{12} x_{t-1} + v_{1t} \qquad (4)$$

$$x_t = \pi_{21} y_{t-1} + \pi_{22} x_{t-1} + v_{2t} \qquad (5)$$

$$\pi_{11} = \frac{\beta_{11} + \theta\beta_{21}}{1 - \theta\gamma} \qquad \pi_{12} = \frac{\beta_{12} + \theta\beta_{22}}{1 - \theta\gamma}$$

$$\pi_{21} = \frac{\gamma\beta_{11} + \beta_{21}}{1 - \theta\gamma} \qquad \pi_{22} = \frac{\gamma\beta_{12} + \beta_{22}}{1 - \theta\gamma}$$

$$v_{1t} = \frac{\varepsilon_{1t} + \theta\varepsilon_{2t}}{1-\theta\gamma} \qquad v_{2t} = \frac{\gamma\varepsilon_{1t} + \varepsilon_{2t}}{1-\theta\gamma}$$

$$Cov(v_{1t}, v_{2t}) = \frac{\gamma\sigma_1^2 + \theta\sigma_2^2}{(1-\theta\gamma)^2}$$

Suponemos que $\varepsilon'_t = (\varepsilon_{1t}, \varepsilon_{2t})$ es una secuencia de ruidos blancos con media cero y matriz de varianzas y covarianzas

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Dada la definición de $v_t^{'} = (v_t, v_{2t})$ su media es cero y su matriz de varianzas y covarianzas puede escribirse como

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix} = \frac{1}{(1-\theta\gamma)^2}\begin{bmatrix} \sigma_1^2 + \theta^2\sigma_2^2 & \gamma\sigma_1^2 + \theta\sigma_2^2 \\ - & \gamma^2\sigma_1^2 + \sigma_2^2 \end{bmatrix}$$

Hay que destacar que la forma reducida del VAR tiene 7 parámetros y que todos ellos están identificados. Por otra parte, la forma estructural tiene 8 parámetros por lo que se requiere una restricción a priori para poder identificar sus parámetros.

En el sentido de Granger se dice que la variable x causa a la variable y cuando la predicción de y, utilizando el error cuadrático medio de predicción, es mejor cuando se tiene en cuenta el pasado de x al lado de toda la información que se considere relevante. Podemos decir que x ayuda a predecir y.

Para formalizar y ver las implicaciones de la propuesta de Granger consideremos algunos aspectos de notación. Sean $\{A_i; i = 0, \pm 1, \pm 2, ...\}$ toda la información relevante; $\overline{A}_i = \{A_j, j < i\}$ la información pasada y $\overline{\overline{A}}_i = \{A_j, j = i\}$ la

información contemporánea. Similármente, $\overline{X}_i, \overline{Y}_i, \overline{\overline{X}}_i, \overline{\overline{Y}}_i$.
Sea $\sigma^2(y/B)$ el error cuadrático medio de predicción definido como

$$\sigma^2(y/B) = E\left[y_{T+h} - \hat{y}_T(h)/B\right]^2$$

siendo B la base informativa en cada caso.

Definición 1. La variable x no causa en el sentido de Granger a la variable y, cuando

$$\sigma^2(y/\overline{A}) = \sigma^2(y/\overline{A} - \overline{X})$$

Utilizando el sistema (4)-(5) esto es equivalente a no rechazar la hipótesis nula.

$$H_0 : \pi_{12} = 0$$

Definición 2. x no causa a y instantaneamente si

$$\sigma^2(y/\overline{A}, \overline{\overline{X}}) = \sigma^2(y/\overline{A})$$

En el sistema (4)-(5) esto es equivalente a o bien no rechazar $\omega_{12} = 0$ o bien si este valor se rechaza, la hipótesis nula $\theta = 0$ no se rechaza.

Definición 3. x no causa a y si

$$\sigma^2(y / \bar{A}, \bar{\bar{X}}) = \sigma^2(y / \bar{A} - \bar{X})$$

Es decir, si se cumplen las dos definiciones previas.

Definición 4. x causa unidireccionalmente a y en el sentido de Granger si

- x causa a y en el sentido de Granger, es decir si

$$\sigma^2(y / \bar{A}) < \sigma^2(y / \bar{A} - \bar{X})$$

O, equivalentemente, si la hipótesis nula $\pi_{12} = 0$ es rechazada.

- y no causa a x en el sentido de Granger, es decir si

$$\sigma^2(x / \bar{A}) = \sigma^2(x / \bar{A} - \bar{Y})$$

O, equivalentemente, si la hipótesis nula $\pi_{21} = 0$ no es rechazada.

Definicion 5. x causa a y unidireccionalmente e instantaneamente si

- Existe correlación contemporanea entre x e y, esto es
$$\omega_{12} = Cov(v_{1t}, v_{2t}) \neq 0$$
- Además se cumple que $\theta \neq 0$ y $\gamma = 0$.

Definición 6. x causa unidireccionalmente a y si

- x causa a y unidireccionalmente en el sentido de Granger y, además,
- x causa a y unidireccionalmente e instantáneamente.

Notar que para establecer la causalidad unidireccional es necesario apelar a la información a priori para identificar los parámetros de la forma estructural. Volvemos a necesitar algo más que los datos para poder establecer una relación causal unidireccional.

La Función Impulso-Respuesta

La Función Impulso-Respuesta es otro enfoque utilizado para derivar relaciones de causalidad unidireccionales. En forma compacta, el modelo (2)-(3) puede escribirse como

$$\Gamma z_t = \mathrm{B} z_{t-1} + \varepsilon_t \qquad (6)$$

en donde $z_t^{'} = (y_t, x_t)$ $\quad y \quad$ $\varepsilon_t^{'} = \left(\varepsilon_{1t}, \varepsilon_{2t}\right)$.

Premultiplicando por Γ^{-1} se obtiene

$$z_t = A z_{t-1} + v_t \qquad (7)$$

en donde

$$v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = \Gamma^{-1}\varepsilon_t = \frac{1}{1-\theta\gamma}\begin{bmatrix} 1 & \theta \\ \gamma & 1 \end{bmatrix}\varepsilon_t = \begin{bmatrix} \dfrac{\varepsilon_{1t}+\theta\varepsilon_{2t}}{1-\theta\gamma} \\ \dfrac{\gamma\varepsilon_{1t}+\varepsilon_{2t}}{1-\theta\gamma} \end{bmatrix} \qquad (8)$$

Sustituyendo sucesivamente en (7) se obtiene

$$z_t = \sum_0^\infty A^i v_{t-i} = \sum_0^\infty A^i \Gamma^{-1} \varepsilon_{t-i} \qquad (9)$$

y definiendo

$$\Phi_i = A^i \Gamma^{-1} = \frac{1}{1-\theta\gamma} A^i \begin{bmatrix} 1 & -\theta \\ -\gamma & 1 \end{bmatrix}$$

(9) puede escribirse como

$$z_t = \sum_0^\infty \Phi_i \varepsilon_{t-i} = \sum_0^\infty \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-i} \\ \varepsilon_{2t-i} \end{bmatrix} \qquad (10)$$

Los cuatro conjuntos de coeficientes $\phi_{11}(i), \phi_{12}(i), \phi_{21}(i)\ y\ \phi_{22}(i)$ se llaman funcionas de impulso-respuesta. Teniendo en cuenta que los ε son independientes entre si, cualquier variación en uno de ellos tiene unos efectos en cada una de las variables que viene dado por los correspondientes ϕ. Por ejemplo, el efecto acumulado de una variación de ε_{2t}, es decir de x_t, después de pasados n periodos viene dado por

$$\sum_{i=0}^{n} \phi_{12}(i)$$

Parece resolverse el problema de la causalidad unidireccional. Pero se trata de un espejismo porque para llegar a la expresión (10) es necesario conocer la matriz Γ y ya hemos comentado que no todos elementos de esta matriz están identificados debido a que en la matriz de varianzas y covarianzas de las v hay tres términos diferentes y tenemos que derivar los valores de cuatro parámetros, los dos de la matriz de varianzas y covarianzas de las ε y los dos que resultan después de normalizar de la matriz Γ. Por lo tanto, para definir las funciones impulso-respuesta necesitamos información a priori. Supongamos que $\gamma = 0$, entonces a partir de (8) podemos escribir

$$\varepsilon_{1t} = v_{1t} - \theta v_{2t}$$

$$\varepsilon_{2t} = v_{2t}$$

Ahora los parámetros están identificados y podemos determinar la causalidad de x sobre y. Podemos modificar v_{2t} y, por lo tanto, ε_{2t} sin modificar ε_{1t} ya que : $Cov(\varepsilon_{1t},\varepsilon_{2t})=0$. Tener en cuenta que, como puede verse en el sistema (7), la modificación de v_{2t} es equivalente a la modificación de x_t.Pero todo esto es consecuencia de suponer, es decir, identificar, con $\gamma=0$. Podríamos haber identificado el sistema suponiendo que $\theta=0$ y los resultados serían diferentes.

CC16.5 Cointegración

Supongamos que X_t e Y_t son integradas de orden uno. Si, para algún coeficiente θ, $Y_t - \theta X_t$ es integrada de orden cero, entonces X_t e Y_t se dice que están cointegradas. El coeficiente θ se denomina **coeficiente de cointegración.**
Si X_t e Y_t están cointegradas, entonces tienen la misma, o común, tendencia estocástica. Calculando la diferencia $Y_t - \theta X_t$ se elimina esta tendencia estocástica común.

Contrastes de Cointegración

A partir del análisis univariante realizado en el tema anterior, podemos conocer la forma que adoptan los elementos deterministas en cada una de las variables y si tienen o no una tendencia estocástica. En esta sección se trata de determinar si las tendencias estocásticas de las dos variables están estrechamente relacionadas. En este caso concreto en que las dos variables son I(1) decimos que las dos están cointegradas si se puede encontrar una relación entre ambas de forma que el residuo de la misma es I(0), es decir, es estacionario e invertible.

Para contrastar la existencia de cointegración, existen dos grandes familias de procedimientos: uniecuacionales y multiecuacionales.

2.1 Métodos Uniecuacionales

Estos métodos se basan en los residuos MCO de la relación lineal entre las dos variables. Estos residuos se definen como

$$\hat{u}_{'t} = y_t - \hat{\alpha} - \hat{\theta} x_t$$

En donde y_t y x_t son las dos variables I(1) , $\hat{\theta}$ es el estimador MCO de θ que es el coeficiente de cointegración. Par concluir que las dos variables están cointegradas el residuo MCO tiene que ser I(0). Siguiendo Engle y Granger(1987), para contrastar el orden de integración del residuo vamos a prestar atención al estadístico Durbin-Watson de la regresión y al contraste de Dickey-Fuller de raíz unitaria. Respecto al primero, seguiremos un tratamiento informal considerando que la evidencia es favorable a la cointegración cuando el estadístico de Durbin y Watson toma un valor superior a 0,50. Respecto al contraste de Dickey-Fuller, se utilizaran los puntos críticos que aparecen en Stock y Watson(2012). La tabla es la siguiente

Valores Críticos del estadístico ADF de Engle-Granger

Número de regresores	10%	5%	1%
1	-3,12	-3,41	-3,96
2	-3,52	-3,80	-4,36
3	-3,84	-4,16	-4,73
4	-4,20	-4,49	-5,07

2.2 Métodos Multiecuacionales

En estos métodos, la base no es la posible relación de cointegración. Sino el modelo multiecuacional autorregresivo con mecanismo de corrección del error (CVAR). Si el sistema tiene n variables el máximo valor de la función de verosimilitud queda en función de las raíces características de una matriz de varianzas y covarianzas de los residuos de las variables y de sus incrementos respecto a los valores pasados de esos incrementos. Si una raíz es cero, entonces solo puede haber n-1 relaciones de cointegración. Si puede aceptarse que todas las raíces son cero, entonces no hay cointegración en el sistema. Por lo tanto, el problema es saber cuantas raíces son cero. Si el rango de cointegración es r entonces el número de raíces que son cero es n-r.

El método multiecuacional más conocido y utilizado es el de Johansen. En este procedimiento se propone una secuencia de contrastes que comienza con la hipótesis nula de que no hay cointegración y, en caso de ir rechazando las sucesivas hipótesis nulas, termina con la hipótesis nula de que el rango de cointegración es n-1. Es importante destacar que los puntos críticos dependen de dos cosas; de los elementos deterministas que se incluyan en el modelo CVAR y de la etapa concreta del proceso secuencial. Una información completa sobre estos puntos críticos puede encontrarse en el libro de Juselius(2006). De todas maneras, la mayor parte de los paquetes informáticos ya tienen incorporados estos puntos críticos y los correspondientes p-valores.

Un aspecto importante en todo este proceso es la determinación de los elementos deterministas del modelo CVAR. Johansen distingue cinco casos. En el primero, las variables no tienen ningún elemento determinista o giran en torno a constantes cointegradas; en este caso, ni el modelo CVAR ni las relaciones de cointegración tienen elementos deterministas. En el segundo caso, las variables giran en torno a constantes no cointegradas; ahora, el modelo CVAR no incluye ningún elemento determinista pero en la relación de cointegración hay una constante. En el tercer caso, las variables giran en torno a tendencias lineales cointegradas, el modelo CVAR incorpora una constante pero la relación de cointegración no tiene ningún elemento determinista. El caso cuarto se refiere a cuando las variables giran en torno a tendencias lineales no cointegradas; en este caso el modelo CVAR tiene una constante y las relaciones de cointegración tienen tendencia lineal. El último caso se refiere a cuando las tendencias son cuadráticas.

Conclusiones de los contrastes de cointegración

- Variables cointegradas $\rightarrow$ modelo con Mecanismo de Corrección del error.
- Variables no cointegradas $\rightarrow$ modelo VAR en primeras diferencias.

3. Modelo con Mecanismo de Corrección de Error (CVAR)

Este modelo es la especificación correcta cuando las variables tienen un orden de integración igual o superior a uno y están cointegradas.
Para dos variables integradas de orden uno y cointegradas la forma del modelo es

$$\Delta Y_t = \beta_{10} + \beta_{11}\Delta Y_{t-1} + \ldots \beta_{1p}\Delta Y_{t-p} + \gamma_{11}\Delta X_{t-1} + \ldots\ldots$$
$$+ \gamma_{1p}\Delta X_{t-p} + \alpha_1\left(Y_{t-1} - \theta X_{t-1}\right) + u_{1t}$$
$$\Delta X_t = \beta_{20} + \beta_{21}\Delta Y_{t-1} + \ldots \beta_{2p}\Delta Y_{t-p} + \gamma_{21}\Delta X_{t-1} + \ldots\ldots$$
$$+ \gamma_{2p}\Delta X_{t-p} + \alpha_2\left(Y_{t-1} - \theta X_{t-1}\right) + u_{2t}$$

El término $Y_t - \theta X_t$ se denomina **término de corrección de error.** Notar su retardo.

FORMAS DEL MODELO

A partir de los resultados obtenidos tras la aplicación de los contrastes comentados se puede dar respuesta a la pregunta:

¿Qué modelo es el adecuado para explicar el comportamiento de las dos variables Y y X?

Según sean los resultados caben varias respuestas:

A) Las dos variables son I(1)

A.1) Las dos variables están cointegradas. En este caso el modelo a utilizar es el MCE. Las diferentes versiones dentro de esta familia dependen de la forma que adopten los elementos deterministas:

- Si las dos variables giran en torno a cero, ni las relaciones del MCE ni la relación de cointegración tienen elementos deterministas
- Si una, o las dos variables giran en torno a una constante no hay elementos deterministas en las relaciones, pero si una constante en las relaciones de cointegración.
- Si una o las dos variables giran en torno a una tendencia lineal, entonces las dos relaciones tienen una constante y la relación de cointegración tiene una tendencia lineal.

A.2) <u>Las dos variables no están cointegradas.</u> En este caso, el modelo a utilizar es el modelo VAR con las primeras diferencias de ambas variables. Las diferencias surgen por la forma que adoptan los elementos deterministas:

- Si una o las dos variables giran en torno a cero o a una constante, en las dos relaciones del modelo no aparece ningún elemento determinista.
- Si una o las dos variables giran en torno a una tendencia lineal, entonces aparece una constante en las dos relaciones del VAR.

B) Una variable es I(1) y la otra es I(0)

En este caso, el modelo apropiado es el modelo VAR en el que la variable I(1) aparece en primeras diferencias y la variable I(0) aparece en niveles. También en este caso, las diferentes versiones dependen de la forma adoptada por los elementos deterministas.

- Si la variable I(1) gira en torno a cero o a una constante, entonces las relaciones del modelo VAR tendrán los mismos elementos deterministas que la variable I(0). Si esta variable gira en torno a cero, entonces las relaciones del modelo VAR no tienen elementos deterministas, si tienen una constante, las dos relaciones del VAR tendrán una constante y lo mismo con la tendencia lineal.

- Si la variable I(1) gira en torno a una tendencia lineal, entonces las dos relaciones del VAR tendrán constante excepto en el caso en el que la variable I(0) gire en torno a una tendencia lineal, en cuyo caso las dos relaciones del VAR tendrán tendencia lineal.

C) Las dos variables son I(0)

En este caso, el modelo apropiado es el modelo VAR con las variables en niveles. Las diferentes versiones son consecuencia de la forma adoptada por los elementos deterministas.

- Si las dos variables giran en torno a cero, entonces las dos relaciones del VAR no tienen elementos deterministas.
- Si una o las dos variables giran en torno a una constante, las dos relaciones del VAR tendrán una constante.
- Si una o las dos variables giran en torno a una tendencia lineal, entonces las dos relaciones del VAR incorporan una relación lineal.

Referencias

Juselius, K.(2006): "The Cointegrated VAR Model". Oxford University Press.

Stock, J.H. y M.M.Watson (2012): "Introducción a la Econometría". Pearson.