Análisis de Series de Tiempo Multivariadas

Introducción

- Debido a la globalización de la economía y a las comunicaciones por Internet, los movimientos de los precios en un mercado pueden extenderse fácilmente e instantáneamente a otro mercado. Por esta razón, hoy los mercados financieros son más dependientes uno de otro y debemos considerarlos conjuntamente para comprender mejor la estructura dinámica de las finanzas globales.

- Un mercado puede guiar a otro bajo ciertas circunstancias, aunque la relación se puede reversar en otras circunstancias. Por tanto, en finanzas es de gran importancia saber cómo los mercados están interrelacionados.

- Para un inversionista o una institución financiera que posee varios activos, las relaciones dinámicas entre los retornos de dichos activos juegan un papel importante en la toma de decisiones.

Series de tiempo Multivariadas

- Una serie de tiempo multivariada consta de varias series de tiempo univariadas, llamadas *componentes.* Sea $\boldsymbol{r_t}=(r_{1t}, r_{2t}, \ldots, r_{kt})'$ un vector que contiene los retornos de k activos en el período t.

- Por ejemplo:

  - ✓ Un inversionista con acciones de IBM, Microsoft, Exxon Mobil, General Motors y Wal-Mart Stores, necesita analizar el vector 5-dimensional de los retornos diarios de esas compañías.

- ✓ Un inversionista que está interesado en inversiones globales, puede considerar los retornos del índice S&P500 de los EU, del índice FTSE 100 del Reino Unido y del índice Nikkei 225 de Japón. En este caso la serie es tridimensional.

Estacionaridad Débil y Matrices de Correlación cruzada

- Considere una serie de tiempo k-dimensional $\boldsymbol{r_t}$=(r_{1t}, r_{2t}, …, r_{kt})'. La serie es débilmente estacionaria si sus dos primeros momentos existen y son invariantes en el tiempo. Específicamente, el vector de medias y la matriz de covarianzas de una serie débilmente estacionaria son constantes en el tiempo. A menos que se diga explícitamente lo contrario, asumiremos que los retornos de las series financieras son débilmente estacionarias.

- Para una serie financiera débilmente estacionaria su vector de media y matriz de covarianzas se define, respectivamente, como

$$\boldsymbol{\mu} = E(r_t), \qquad \Gamma_0 = E[(r_t - \boldsymbol{\mu})(r_t - \boldsymbol{\mu})']$$

donde

$$\boldsymbol{\mu} = E(r_t) = \begin{bmatrix} E(r_{1t}) \\ E(r_{2t}) \\ \vdots \\ E(r_{kt}) \end{bmatrix}_{kx1} \qquad \Gamma_0 = E[(r_t - \boldsymbol{\mu})(r_t - \boldsymbol{\mu})'] = \begin{bmatrix} \gamma_{11}(0),\ \gamma_{12}(0),\cdots,\gamma_{1k}(0) \\ \gamma_{21}(0),\ \gamma_{22}(0),\cdots,\gamma_{2k}(0) \\ \cdots \\ \gamma_{k1}(0),\ \gamma_{k2}(0),\cdots,\gamma_{kk}(0) \end{bmatrix}_{kxk}$$

donde $\gamma_{ij}(0) = \begin{cases} Cov(r_{it}, r_{jt}), & i \neq j \\ Var(r_{it}), & i = j \end{cases}$

Matrices de Correlación Cruzada

Sea D una matriz diagonal cuyos elementos en la diagonal son las desviaciones estándar de las r_{it}, es decir

D=diag$\left[\sqrt{\gamma_{11}(0)},\ \sqrt{\gamma_{22}(0)},\cdots,\sqrt{\gamma_{kk}(0)}\right]$. La matriz de *correlación contemporánea o de rezago cero*, está dada por la matriz de correlación cruzada de $\boldsymbol{r_t}$ definida como

$$\boldsymbol{\rho}_0=[\rho_{i\,j}(0)]=D^{-1}\Gamma_0 D^{-1}$$

$$\text{donde } \rho_{ij}(0)=\frac{\gamma_{ij}(0)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}$$

es decir $\rho_{ij}(0)$ es el coeficiente de correlación entre r_{it} y r_{jt}. Este coeficiente es llamado el coeficiente de *correlación contemporáneo* debido a que es la correlación de las dos series en el período t.

Propiedades:

i) $\rho_{ij}(0)=\rho_{ji}(0)$

ii) $|\rho_{ij}(0)|\leq 1$

iii) $\boldsymbol{\rho}_{ii}(0)=1$

Por tanto, $\boldsymbol{\rho}_0$ es una matriz simétrica con unos en su diagonal.

- Un tópico importante en análisis series de tiempo multivariadas es la relación *adelanto-retraso* entre las series componentes. En este caso las matrices de correlación cruzada son usadas para medir la fortaleza de la dependencia lineal entre las series. La matriz de covarianza cruzada de rezago l de $\boldsymbol{r_t}$ está definida como

$$\Gamma_l=\left[\gamma_{ij}(l)\right]=E[(r_t-\mu)(r_{t-l}-\mu)']$$

El (i,j)-ésimo elemento de esta matriz es la covarianza entre r_{it} y $r_{j,t-l}$.

- En una serie de tiempo débilmente estacionaria, la matriz $\mathbf{\Gamma}_l$ es una función de l, y no del tiempo t.

- La matriz de correlación cruzada de rezago l de $\boldsymbol{r_t}$ está definida como

$$\rho_l = \left[\rho_{ij}(l)\right] = D^{-1}\mathbf{\Gamma}_l D^{-1}$$

donde

$$\rho_{ij}(l) = \frac{\gamma_{ij}(l)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}} = \frac{Cov(r_{it}, r_{jt-l})}{std(r_{it})std(r_{jt-l})}$$

- Cuando l>0, $\rho_{ij}(l)$ mide la dependencia lineal de r_{it} sobre $r_{j,t-l}$, la cual ocurrió antes del período t. Por tanto si $\rho_{ij}(l) \neq 0$ y l>0 se dice que la serie r_{jt} *lidera* a la serie r_{it} en el rezago l.

- Similarmente $\rho_{ji}(l)$ mide dependencia lineal de r_{jt} sobre $r_{i,t-l}$, y se dice que la serie r_{it} *lidera* a la serie r_{jt} en el rezago l si $\rho_{ji}(l) \neq 0$ y l>0.

- El i-ésimo elemento de la diagonal, $\rho_{ii}(l)$ es el coeficiente de autocorrelación de orden l de r_{it}.

Propiedades de las correlaciones cruzadas cuando l>0.

i) $\rho_{ij}(l) \neq \rho_{ji}(l)$, para i≠j. Por tanto, en general, $\mathbf{\Gamma}_l$ y $\boldsymbol{\rho}_l$ son matrices no simétricas.

ii) $Cov(r_{it}, r_{j,t-l},) = Cov(r_{jt}, r_{i,t+l})$. Por tanto $\mathbf{\Gamma}_l = \mathbf{\Gamma}_{-l}^{'}$ y $\boldsymbol{\rho}_l = \boldsymbol{\rho}_{-l}^{'}$. Entonces, a diferencia de una serie univariada, en la cual $\rho_l = \rho_{-l}$, para una serie de tiempo multivariada $\boldsymbol{\rho}_l \neq \boldsymbol{\rho}_{-l}$. Debido a que $\boldsymbol{\rho}_l = \boldsymbol{\rho}_{-l}^{'}$ es suficiente considerar las matrices de correlación $\boldsymbol{\rho}_l$ cuando l>0.

Dependencia lineal

- Consideradas conjuntamente, las matrices de correlación cruzada {$\boldsymbol{\rho}_l$, | *l=0,1,*,2…} de un vector de series de tiempo débilmente estacionario, contienen la siguiente información:

  ✓ Los elementos de la diagonal {$\rho_{ii}(l)$ | l=0,1,2…} son las funciones de autocorrelación de las $r_{it.}$

  ✓ Los elementos fuera de la diagonal $\rho_{ij}(0)$ miden la relación lineal contemporánea entre r_{it} y r_{jt}.

  ✓ Para l>0 los elementos $\rho_{ij}(l)$ fuera de la diagonal miden la dependencia lineal de r_{it} sobre el valor pasado $r_{j,t-l}$.

- Si $\rho_{ij}(l)$ =0 para todo l>0 entonces r_{it} no depende linealmente de cualquier valor pasado $r_{j,t-l}$ de la serie r_{jt}.

- En general, la relación lineal entre dos series de tiempo {r_{it}} y {r_{jt}} puede ser resumida de la siguiente manera:

1. r_{it} y r_{jt} no tienen relación lineal si $\rho_{ij}(l)=\rho_{ji}(l)$ =0, para todo $l\geq 0$.
2. r_{it} y r_{jt} están correlacionadas contemporáneamente si $\rho_{ij}(0)\neq 0$.
3. r_{it} y r_{jt} no tienen relación lineal adelanto-retraso si $\rho_{ij}(l)$ =0 y $\rho_{ji}(l)$ =0, para todo l>0.
4. Existe una relación unidireccional de r_{it} hacia r_{jt} si $\rho_{ij}(l)$ =0 para todo l>0, pero $\rho_{ji}(v)\neq 0$ para algún v>0. En este caso r_{it} no depende del pasado de r_{jt}, pero r_{jt} depende de algunos valores pasados de r_{it}.
5. Existe una relación de *retroalimentación* entre r_{it} y r_{jt} si $\rho_{ij}(l)\neq 0$ para

algunos l>0, y $\rho_{ji}(v) \neq 0$ para algún v>0.

Matrices de Correlación Muestral

- Dados los datos {$\boldsymbol{r_t}$ | t=1,2,…,T} la matriz de covarianza cruzada $\boldsymbol{\Gamma}_l$ puede ser estimada por

$$\hat{\Gamma}_l = \frac{1}{T}\sum_{t=l+1}^{T}(r_t - \bar{r})(r_{t-l} - \bar{r})', \qquad l \geq 0$$

donde $\bar{\boldsymbol{r}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t$ es el vector de las medias muestrales.

- La matriz de correlación cruzada $\boldsymbol{\rho}_l$ se estima usando

$$\hat{\boldsymbol{\rho}}_l = \hat{D}^{-1}\hat{\Gamma}_l\hat{D}^{-1}, \qquad l \geq 0$$

donde $\hat{D}$ es una matriz diagonal de k x k cuyos elementos en la diagonal son las desviaciones estándar muestrales de las r_{it}.

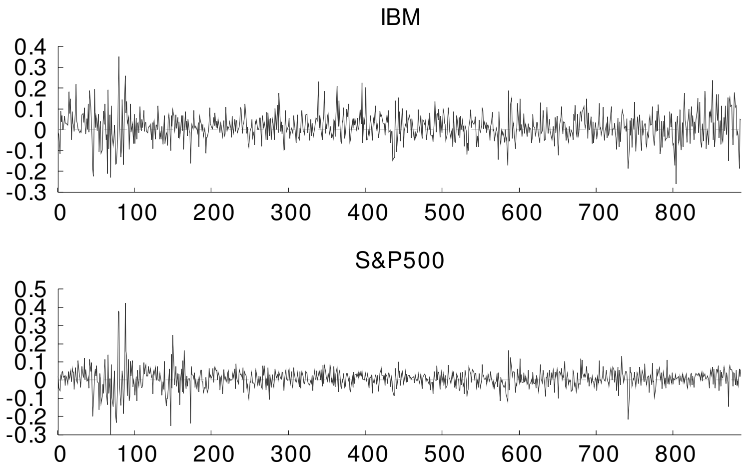
Propiedades de los estimadores

i) Son sesgados en muestras finitas
ii) Son consistentes.
iii) En series compuestas por los retornos de activos, la distribución muestral de $\hat{\boldsymbol{\rho}}_l$ es más complicada debido en parte a la presencia de heterocedasticidad condicional y alta curtosis.
iv) Si se requiere su distribución en muestras finitas se recomienda obtenerla por medio de remuestreo usando Bootstrap.

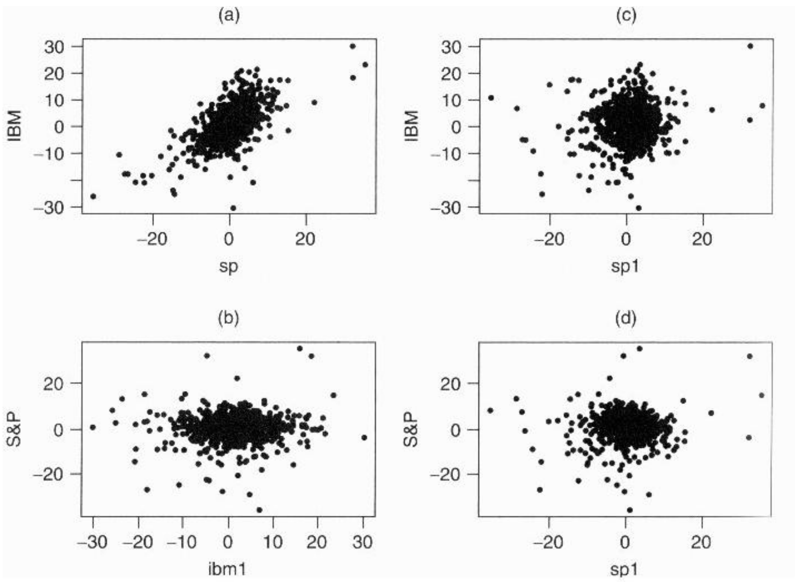
Ejemplo

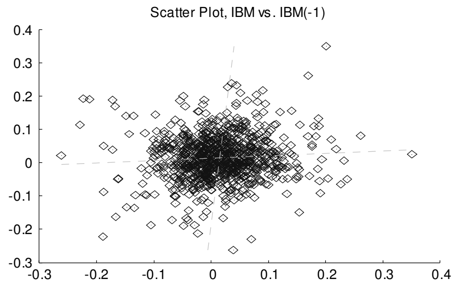
Considere los retornos mensuales de las acciones de IBM y del índice S&P500 desde Enero de 1926 hasta diciembre de 1999, para un total de 888 observaciones.

Sean r_{1t} y r_{2t} los retornos mensuales de las acciones de IBM y del índice S&P500, respectivamente. Estos dos retornos forman la serie bivariada $\boldsymbol{r}_t$=(r_{1t}, r_{2t})'.



Los siguientes gráficos muestran algunos diagramas de dispersión entre las dos series y sus primeros rezagos.





Los gráficos muestran que;

- ✓ Las series están contemporáneamente correlacionadas (la correlación muestral es de 0.64, estadísticamente significativa a un 5%).

- ✓ Las correlaciones cruzadas de rezago 1 son débiles.

La siguiente tabla presenta un resumen estadístico de las dos series junto con sus matrices de correlación cruzada muestral

Resumen estadístico

Retorno	media	std	Asimetría	Exceso-Curtosis	Mínimo	Máximo.
IBM	1.240	6.729	-0.237	1.917	-30.37	30.10
S&P	0.537	5.645	-0.521	8.117	-35.58	35.22

Matrices de correlación cruzada

Retraso 1		Retraso 2		Retraso 3		Retraso 4		Retraso 5	
0.08	0.10	0.02	-0.06	-0.02	-0.07	-0.02	-0.03	0.00	0.07
0.04	0.08	0.02	-0.02	-0.07	-0.11	0.04	0.02	0.00	0.08

Notación simplificada de las Matrices de correlación cruzada

$$\begin{bmatrix} + & + \\ \bullet & + \end{bmatrix} \quad \begin{bmatrix} \bullet & \bullet \\ \bullet & \bullet \end{bmatrix} \quad \begin{bmatrix} \bullet & - \\ - & - \end{bmatrix} \quad \begin{bmatrix} \bullet & \bullet \\ \bullet & \bullet \end{bmatrix} \quad \begin{bmatrix} \bullet & \bullet \\ \bullet & + \end{bmatrix}$$

La experiencia empírica ha mostrado que es muy difícil entender simultáneamente muchas matrices de correlación cruzada, especialmente cuando k>3. Para superar esta dificultad se emplea la notación simplificada introducida por Tiao y Box (1981). Se define la *matriz de correlación cruzada simplificada* como una matriz que puede contener a lo más los tres símbolos "+", "-", "•", donde

- ✓ "+" significa que el correspondiente coeficiente de correlación es mayor o igual a $2/\sqrt{T}$.
- ✓ "-" significa que el correspondiente coeficiente de correlación es menor o igual a $-2/\sqrt{T}$.
- ✓ "•" significa que el correspondiente coeficiente de correlación está entre $-2/\sqrt{T}$ y $2/\sqrt{T}$.

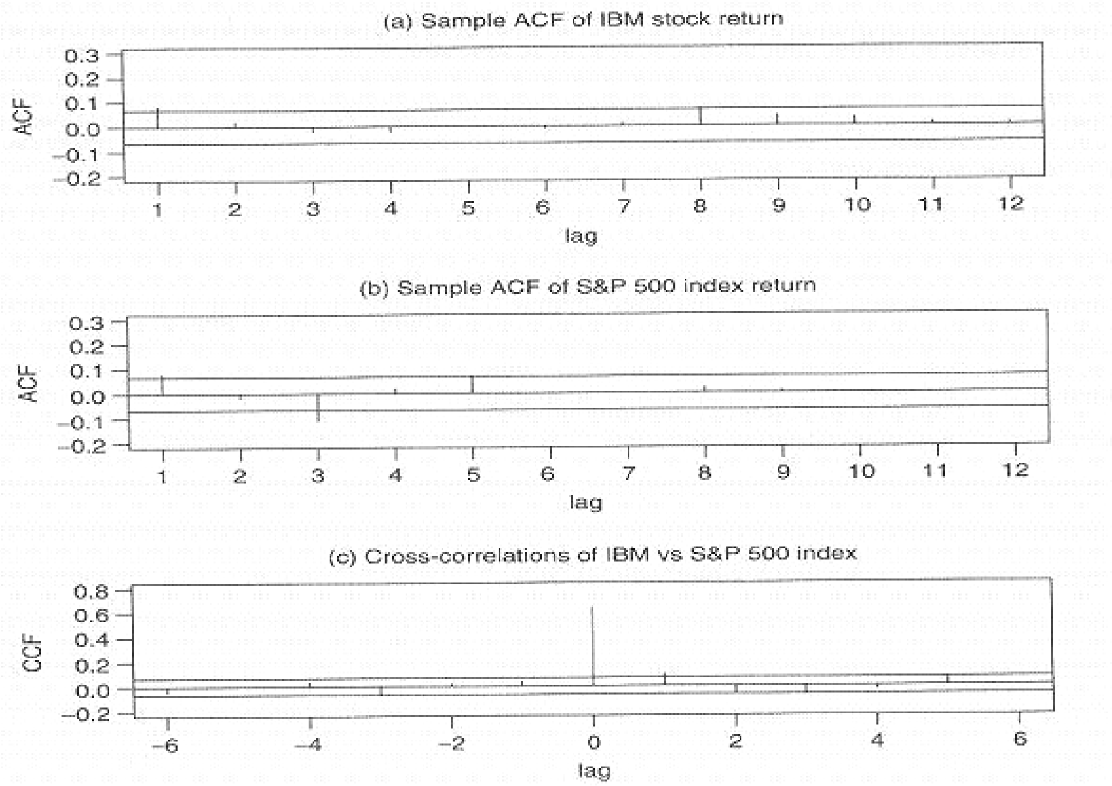
donde $2/\sqrt{T}$ es el valor crítico asintótico al 5% de la correlación muestral bajo el supuesto que la serie $\boldsymbol{r}_t$ es ruido blanco.

La notación simplificada de la CCM muestra que:

- ✓ Autocorrelaciones marginales significativas a un nivel aproximado del 5% aparecen en los rezagos 1 y 3.

- ✓ Los retornos de las acciones de IBM dependen de retornos anteriores del índice S&P 500.

✓ Los siguientes gráficos presentan las autocorrelaciones muestrales y la correlación cruzada de las dos series.



⇒ El gráfico c) asociado con valores positivos en los rezagos, muestra la dependencia de la serie de retornos de IBM del pasado de los retornos del índice S&P500.

⇒ El gráfico c) asociado con valores negativos en los rezagos, muestra la dependencia de la serie de los retornos del índice S&P500 del pasado de la serie de los retornos de IBM.

⇒ Las líneas horizontales en los gráficos indican los límites asintóticos a dos veces el error estándar de los coeficientes de autocorrelación y de correlación cruzada muestral.

⇒ De los gráficos anteriores se observa que la relación dinámica entre las dos series es débil, pero su correlación contemporánea es significativa.

Comandos para el SCA

input var ibm sp500. file 'c:\datos\m_ibm_sp500.txt'.
describe var ibm sp500.
ccm var ibm sp500. maxlag 6. output level(detailed).

Resultados:

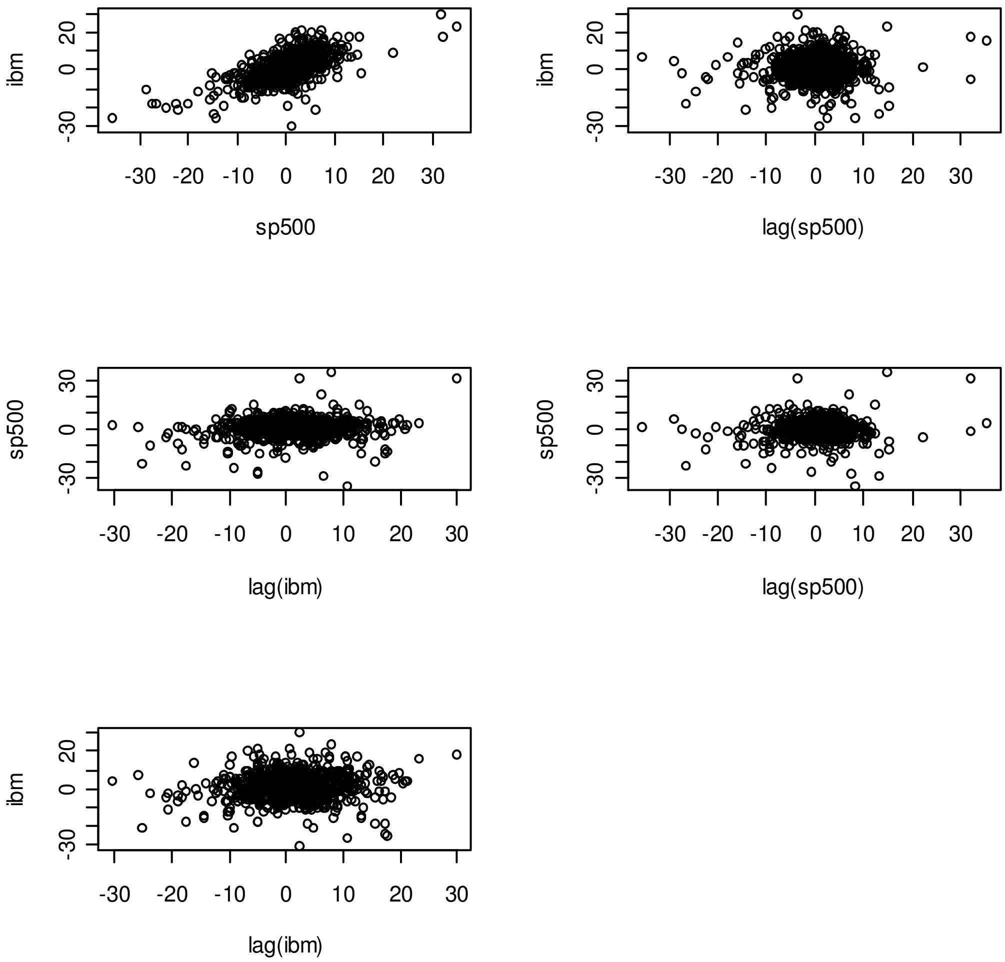
```
VARIABLE    NAME    IS       IBM
NUMBER OF OBSERVATIONS       888
NUMBER OF MISSING VALUES       0

                   STATISTIC      STD. ERROR     STATISTIC/S.E.

MEAN                 0.0148          0.0023          6.4706
VARIANCE             0.0046
STD DEVIATION        0.0680
C.  V.               4.6053
SKEWNESS             0.1452          0.0821
KURTOSIS             1.7682          0.1639

                   QUARTILE

MINIMUM             -0.2619
1ST QUARTILE        -0.0248
MEDIAN               0.0123
3RD QUARTILE         0.0526
MAXIMUM              0.3512

                      RANGE

MAX - MIN            0.6131
Q3  - Q1             0.0775


VARIABLE    NAME    IS     SP500
NUMBER OF OBSERVATIONS       888
NUMBER OF MISSING VALUES       0

                   STATISTIC      STD. ERROR     STATISTIC/S.E.

MEAN                 0.0070          0.0019          3.6802
VARIANCE             0.0032
STD DEVIATION        0.0565
C.  V.               8.0973
SKEWNESS             0.3545          0.0821
KURTOSIS             9.6264          0.1639

                   QUARTILE

MINIMUM             -0.2994
1ST QUARTILE        -0.0199
MEDIAN               0.0094
3RD QUARTILE         0.0376
MAXIMUM              0.4222

                      RANGE
```

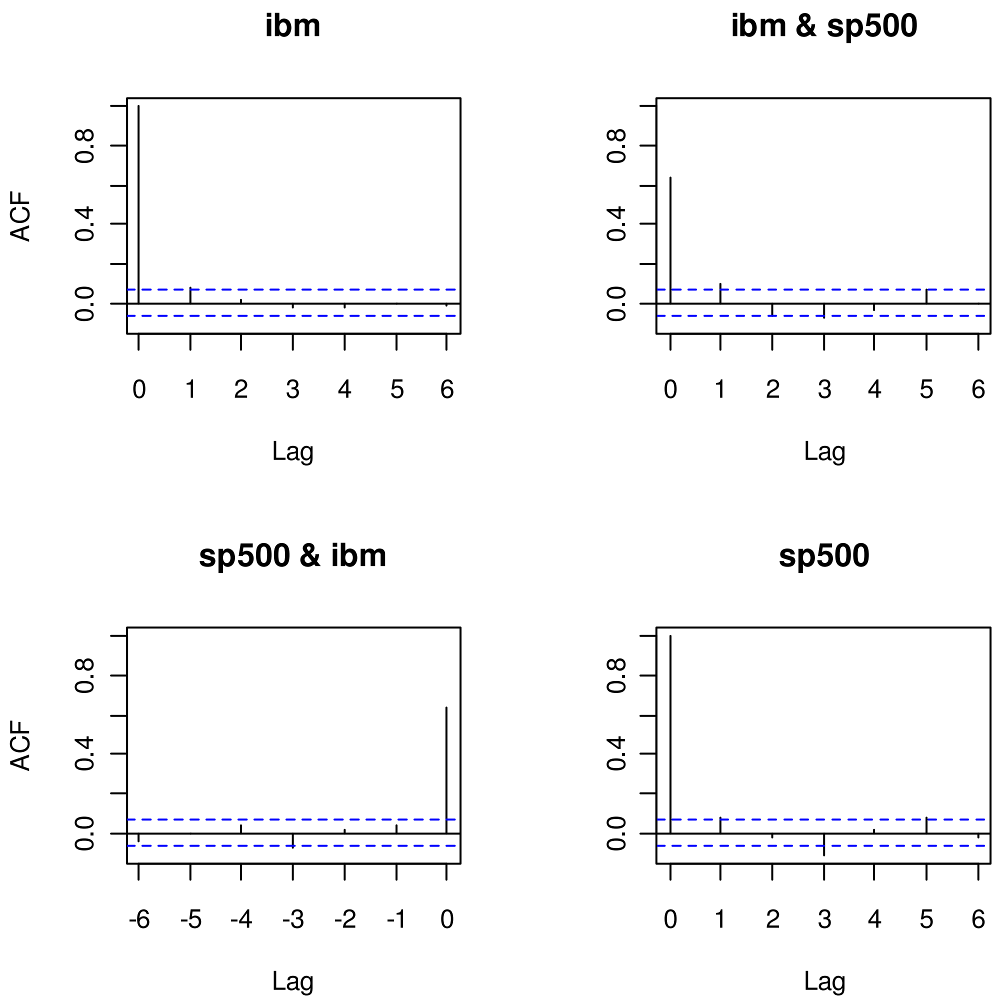
```
MAX - MIN            0.7216
Q3  - Q1             0.0575
--


TIME PERIOD ANALYZED . . . . . . . . . . . .  1  TO   888
EFFECTIVE NUMBER OF OBSERVATIONS (NOBE). . .          888

SERIES   NAME           MEAN       STD. ERROR
   1    IBM            0.0148      0.0679
   2    SP500          0.0070      0.0564


NOTE: THE APPROX. STD. ERROR FOR THE ESTIMATED CORRELATIONS BELOW
      IS (1/NOBE**.5) =    0.03356

SAMPLE CORRELATION MATRIX OF THE SERIES

  1.00
  0.63  1.00

SAMPLE CROSS CORRELATION MATRICES FOR THE ORIGINAL SERIES.
THE (I,J) ELEMENT OF THE LAG L MATRIX IS THE ESTIMATE OF
THE LAG L CROSS CORRELATION WHEN SERIES J LEADS SERIES I

LAG = 1

  0.07  0.10
  0.05  0.08

LAG = 2

  0.01 -0.06
  0.02 -0.02

LAG = 3

 -0.02 -0.07
 -0.07 -0.12


LAG = 4

 -0.03 -0.03
  0.03  0.02

LAG = 5

  0.01  0.06
  0.00  0.07

LAG = 6

 -0.01 -0.01
 -0.05 -0.04

SUMMARIES OF CROSS CORRELATION MATRICES USING +,-,., WHERE
     + DENOTES A VALUE GREATER THAN 2/SQRT(NOBE)
     - DENOTES A VALUE LESS THAN -2/SQRT(NOBE)
     . DENOTES A NON-SIGNIFICANT VALUE BASED ON THE ABOVE CRITERION

BEHAVIOR OF VALUES IN (I,J)TH POSITION OF CROSS CORRELATION MATRIX OVER
ALL OUTPUTTED LAGS WHEN SERIES J LEADS SERIES I

         1              2

 1   +.....         +.-...


 2   ..-...         +.-.+.


CROSS CORRELATION MATRICES IN TERMS OF +,-,.
LAGS  1 THROUGH  6
```

```
      + +     . .     . -     . .     . .     . .
      . +     . .     - -     . .     . +     . .
--
```

Comandos para R

```
(ibm_sp500=read.table("g:/datos/m_ibm_sp500.txt"))
ibm=ts(ibm_sp500[,1])
sp500=ts(ibm_sp500[,2])
# gráficos
ibm_sp500_m=cbind(ibm, sp500)
plot.ts(ibm_sp500_m)
par(mfrow=c(3,2))
plot(sp500,ibm)
plot(lag(sp500), ibm)
plot(lag(ibm), sp500)
plot(lag(sp500), sp500)
plot(lag(ibm), ibm)
# autocorrelaciones y correlaciones cruzadas
acf(ibm_sp500_m)
acf(ibm_sp500_m, plot=FALSE, lag.max=5)
```


ibm
sp500
lag(sp500)
sp500
lag(ibm)
sp500
lag(sp500)
ibm
lag(ibm)
-30 -20 -10 0 10 20 30




```
Autocorrelations of series ‘ibm_sp500_m’, by lag

, , ibm

 ibm          sp500
  1.000 ( 0)  0.636 ( 0)
  0.076 ( 1)  0.044 (-1)
  0.016 ( 2)  0.021 (-2)
 -0.019 ( 3) -0.073 (-3)
 -0.023 ( 4)  0.039 (-4)
  0.004 ( 5)  0.002 (-5)
 -0.008 ( 6) -0.040 (-6)

, , sp500

 ibm          sp500
  0.636 ( 0)  1.000 ( 0)
  0.101 ( 1)  0.076 ( 1)
 -0.060 ( 2) -0.016 ( 2)
```

```
-0.071 ( 3) -0.110 ( 3)
-0.033 ( 4)  0.024 ( 4)
 0.067 ( 5)  0.084 ( 5)
-0.003 ( 6) -0.021 ( 6)
```

Prueba de Portmateau Multivariada

- El estadístico univariado Q(m) de Ljung y Box fue generalizado al caso multivariado por Hosking (1980, 1981) y Li y McLeod (1981). Para series multivariadas, la hipótesis nula es H_0: $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_2 = \cdots = \boldsymbol{\rho}_m = \mathbf{0}$, y la hipótesis alternativa es H_1: $\boldsymbol{\rho}_i \neq \mathbf{0}$, para algún $i \in \{1,2,...,m\}$. El estadístico generalizado es

$$Q_k(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} tr(\hat{\boldsymbol{\Gamma}}_l' \hat{\boldsymbol{\Gamma}}_0^{-1} \hat{\boldsymbol{\Gamma}}_l \hat{\boldsymbol{\Gamma}}_0^{-1})$$

  Donde T es el número de observaciones, k es la dimensión de $\boldsymbol{r}_t$, y tr(A) es la traza de la matriz A. Bajo H_0 y algunas condiciones de regularidad $Q_k(m)$ sigue asintóticamente una distribución χ^2 con k^2m grados de libertad.

- El estadístico $Q_k(m)$ puede ser escrito en términos de correlación cruzadas muestrales $\hat{\boldsymbol{\rho}}_l$. La expresión está en términos del producto Kronecker y de la vectorización de matrices:

$$Q_k(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} b_l'(\hat{\rho}_0^{-1} \otimes \hat{\rho}_0^{-1}) b_l$$

  donde b_l =vec($\hat{\boldsymbol{\rho}}_l'$).

- El estadístico propuesto por Li y McLeod (1981) es

$$Q_k^*(m) = T\sum_{l=1}^{m}\frac{1}{T-l}b_l'(\hat{\rho}_0^{-1}\otimes\hat{\rho}_0^{-1})b_l + \frac{k^2m(m+1)}{2T}$$

el cual es asintóticamente equivalente a $Q_k(m)$.

Ejemplo. Aplicando $Q_k(m)$ a la serie bivariada de los retornos de las acciones de IBM y del índice S&P, se obtiene $Q_2(1)$ =9.81, $Q_2(5)$ =47.06 y $Q_2(10)$ =71.65. Usando la distribución asintótica para 4, 20 y 40 grados de libertad, todos los valores p de las pruebas están cercanos a cero. Por tanto, se confirma la existencia de dependencia serial entre las dos series.

- El estadístico $Q_k(m)$ es una prueba conjunta para chequear las primeras m matrices de correlación cruzada de $\boldsymbol{r_t}$. Si se rechaza la hipótesis nula, debemos construir un modelo multivariado para las series de forma que permita estudiar la relaciones adelanto-rezago entre las componentes de $\boldsymbol{r_t}$.

Modelos Vectoriales Autorregresivos

- Una serie de tiempo multivariada k dimensional $\boldsymbol{r_t}$ es un proceso VAR de orden 1 o VAR(1), si sigue el modelo

$$\boldsymbol{r_t} = \boldsymbol{\phi}_0 + \boldsymbol{\Phi r}_{t-1} + \boldsymbol{a}_t$$

donde $\boldsymbol{\phi}_0$ es un vector de constantes k dimensional, $\boldsymbol{\Phi}$ es una matriz de constantes de k x k, y $\{\boldsymbol{a}_t\}$ es una sucesión de vectores aleatorios de kx1 no correlacionados con vector de media cero y matriz de covarianza $\boldsymbol{\Sigma}$, definida positiva. $\{\boldsymbol{a}_t\}$ es llamado un proceso de ruido blanco multivariado y frecuentemente se supone que $\boldsymbol{a}_t$ sigue una distribución normal multivariada.

- En términos del operador de retraso B el VAR(1) se puede escribir como

$$(I - \Phi B)r_t = \phi_0 + a_t$$

donde I es la matriz identidad de kxk.

- Sea $\Phi(B) = (I - \Phi B)$. Entonces el VAR(1) también se puede escribir como

$$\Phi(B)r_t = \phi_0 + a_t$$

- Considere el caso Bivariado: k=2, $\mathbf{r}_t$=(r_{1t}, r_{2t}) y $\boldsymbol{a}_t$=(a_{1t}, a_{2t}). El VAR(1) es

$$r_{1t} = \boldsymbol{\phi}_{10} + \boldsymbol{\Phi}_{11} r_{1,t-1} + \boldsymbol{\Phi}_{12} r_{2,t-1} + a_{1t}$$
$$r_{2t} = \boldsymbol{\phi}_{20} + \boldsymbol{\Phi}_{21} r_{1,t-1} + \boldsymbol{\Phi}_{22} r_{2,t-1} + a_{2t}$$

donde $\boldsymbol{\Phi}_{ij}$ es el (i,j)-ésimo elemento de $\boldsymbol{\Phi}$, y $\boldsymbol{\phi}_{io}$ es el i-ésimo elemento de $\boldsymbol{\phi}_0$.

  - ✓ Dada la primera ecuación, $\boldsymbol{\Phi}_{12}$ denota la dependencia lineal de r_{1t} sobre $r_{2,t-1}$ en presencia de $r_{1,t-1}$. Por tanto $\boldsymbol{\Phi}_{12}$ es el efecto condicional de $r_{2,t-1}$ sobre r_{1t} dado $r_{1,t-1}$.

  - ✓ Si $\boldsymbol{\Phi}_{12}$=0, entonces r_{1t} solamente depende de su propio pasado y no de $r_{2,t-1}$.

  - ✓ La segunda ecuación muestra que si $\boldsymbol{\Phi}_{21}$=0, r_{2t} no depende de $r_{1,t-1}$ cuando $r_{2,t-1}$ está presente.

  - ✓ Considere las dos ecuaciones conjuntamente:

$\Rightarrow$ Si Φ_{12}=0 y $\Phi_{21} \neq 0$, entonces existe una relación lineal unidireccional de r_{1t} a r_{2t}.

$\Rightarrow$ Si $\Phi_{12} \neq 0$ y Φ_{21}=0, entonces existe una relación lineal unidireccional de r_{2t} a r_{1t}.

$\Rightarrow$ Si Φ_{12}=0 y Φ_{21}=0, entonces no existe una relación lineal entre r_{2t} y r_{1t}.

$\Rightarrow$ Si $\Phi_{12} \neq 0$ y $\Phi_{21} \neq 0$, entonces existe una relación lineal de retroalimentación entre r_{2t} y r_{1t}.

Formas Reducida y estructural

- En general, la matriz coeficiente Φ mide la **dependencia dinámica** de $\boldsymbol{r_t}$. El elemento σ_{12} fuera de la diagonal de la matriz Σ, muestra la relación **contemporánea** entre r_{1t}, y r_{2t}. Si σ_{12} =0, no hay relación lineal contemporánea entre r_{1t} y r_{2t}.

- En la literatura econométrica, el modelo VAR(1) es llamado un modelo en *forma reducida* debido a que no muestra explícitamente la dependencia **contemporánea** entre las series. Si es necesario, se puede obtener una expresión explícita que muestre la relación de dependencia contemporánea usando una transformación lineal simple.

  ✓ Como Σ es definida positiva, existe una matriz triangular inferior L con unos en la diagonal, y una matriz diagonal G tal que

  $$\Sigma = LGL' \quad \text{(Descomposición de Cholesky)}$$

  ✓ Multiplicando por L^{-1} el modelo VAR(1),

$$L^{-1}\boldsymbol{r}_t = L^{-1}\boldsymbol{\phi}_0 + L^{-1}\boldsymbol{\Phi}\boldsymbol{r}_{t-1} + L^{-1}\boldsymbol{a}_t$$
$$= \boldsymbol{\phi}_0^* + \boldsymbol{\Phi}^*\boldsymbol{r}_{t-1} + \boldsymbol{b}_t$$

donde $\mathbf{b}_t=(b_{1t}, b_{2t}, \ldots, b_{kt})' = L^{-1}\boldsymbol{a}_t$. Entonces

$$E(\mathbf{b}_t) = L^{-1}E(\boldsymbol{a}_t) = \mathbf{0}, \text{ y } Cov(\mathbf{b}_t) = L^{-1}\boldsymbol{\Sigma}(L^{-1})' = G$$

Puesto que G es diagonal, los elementos de $\mathbf{b}_t$ no están correlacionados.

Además $\boldsymbol{\phi}_0^* = L^{-1}\boldsymbol{\phi}_0$ es un vector k dimensional y $\boldsymbol{\Phi}^* = L^{-1}\boldsymbol{\Phi}$ es una matriz de k x k. Debido a la estructura de L, la k-ésima fila de L^{-1} es de la forma ($w_{k1}, w_{k2}, \ldots, w_{k,k-1}$, 1). Por tanto la k-ésima ecuación del modelo transformado anterior es

$$r_{kt} + \sum_{i=1}^{k-1} w_{ki} r_{it} = \boldsymbol{\phi}_{k,0}^* + \sum_{i=1}^{k} \boldsymbol{\Phi}_{ki}^* r_{i,t-1} + b_{kt}$$

- ✓ Debido a que b_{kt} no está correlacionada con b_{it} para $1 \le i < k$, la ecuación anterior muestra explícitamente la dependencia lineal contemporánea de r_{kt} sobre r_{it}, donde $1 \le i < k$. Esa ecuación es llamada una *ecuación estructural* en la literatura econométrica.

- ✓ Para cualquier otra componente r_{it} de $\boldsymbol{r}_t$, se puede re-arreglar el modelo VAR(1) de manera que r_{it} sea la última componente de $\boldsymbol{r}_t$. Por tanto, el modelo reducido dado por el VAR(1) originalmente definido, es equivalente a la forma estructural usada en la literatura econométrica.

- ✓ En el análisis de series de tiempo, se acostumbra usar la forma reducida por dos razones: 1. Fácil estimación y, 2. La principal, es que las correlaciones contemporáneas no pueden ser usadas en los pronósticos.

Ejemplo.

Para ilustrar la transformación desde un modelo de forma reducida a las ecuaciones estructurales, considere el siguiente modelo bivariante AR(1):

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

donde $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$.

- ✓ Usando la descomposición de Cholesky (es decir, $L^{-1}\Sigma(L')^{-1}$ es una matriz diagonal), la matriz triangular inferior L^{-1} es :

$$L^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix}$$

Premultiplicando el AR(1) por L^{-1} se obtiene

$$\begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix} \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.7 & 0.95 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix}$$

donde $G = \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$ es la matriz de covarianzas de $b = \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix}$.

- ✓ La segunda ecuación del modelo transformada da

$$r_{2t} = 0.3 + 0.5 r_{1t} - 0.7 r_{1,t-1} + 0.95 r_{2,t-1} + b_{2t}$$

La cual muestra explícitamente la dependencia de r_{2t} sobre r_{1t}.

- ✓ Re-arreglando el orden de los elementos en $\boldsymbol{r}_t$, el modelo bivariante se puede escribir como

$$\begin{bmatrix} r_{2t} \\ r_{1t} \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 1.1 & -0.6 \\ 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} r_{2,t-1} \\ r_{1,t-1} \end{bmatrix} + \begin{bmatrix} a_{2t} \\ a_{1t} \end{bmatrix}$$

donde $\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$.

- ✓ La matriz triangular inferior necesitada en la descomposición de Cholesky de Σ en el modelo anterior es

$$L^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -1.0 & 1.0 \end{bmatrix}$$

Premultiplicando por L^{-1} el modelo VAR(1) re-arreglado, se obtiene

$$\begin{bmatrix} 1.0 & 0.0 \\ -1.0 & 1.0 \end{bmatrix} \begin{bmatrix} r_{2t} \\ r_{1t} \end{bmatrix} = \begin{bmatrix} 0.4 \\ -0.2 \end{bmatrix} + \begin{bmatrix} 1.1 & -0.6 \\ -0.8 & 0.8 \end{bmatrix} \begin{bmatrix} r_{2,t-1} \\ r_{1,t-1} \end{bmatrix} + \begin{bmatrix} c_{2t} \\ c_{1t} \end{bmatrix}$$

donde $G = \operatorname{cov}(c) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

- ✓ Ahora la segunda ecuación da

$$r_{1t} = -0.2 + 1.0 r_{2t} - 0.8 r_{2,t-1} + 0.8 r_{1,t-1} + c_{1t}$$

Esta ecuación muestra explícitamente la relación de dependencia de r_{1t} sobre r_{2t}.

Condición de Estacionaridad y Momentos de un VAR(1)

- Suponga que el VAR(1) es débilmente estacionario. Tomando esperanza, y puesto E($\boldsymbol{a}_t$)=**0,**

$$E(\boldsymbol{r}_t)=\phi_0 + \Phi\ E(\boldsymbol{r}_{t-1})$$

Como E($\boldsymbol{r}_t$) es invariante en el tiempo,

$$\mu = E(\boldsymbol{r}_t)=(I - \Phi)^{-1}\ \phi_0$$

dado que la matriz (I - Φ) sea no singular, siendo I la matriz identidad de k x k.

- Usando ϕ_0=(I - Φ) μ, el modelo VAR(1) se puede escribir como

$$\boldsymbol{r}_t - \mu = \Phi(\boldsymbol{r}_{t-1} - \mu) + \boldsymbol{a}_t$$

Sea $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \mu$, la serie corregida por la media. Entonces el Var(1) se puede escribir como:

$$\tilde{\boldsymbol{r}}_t = \Phi\tilde{\boldsymbol{r}}_{t-1} + \boldsymbol{a}_t$$

- Usando el modelo anterior, por sustituciones repetidas,

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{a}_t + \Phi\boldsymbol{a}_{t-1} + \Phi^2\boldsymbol{a}_{t-2} + \Phi^3\boldsymbol{a}_{t-3} + \ldots$$

De esta expresión podemos observar varias características del VAR(1):

  - ✓ Puesto que $\boldsymbol{a}_t$ es serialmente no correlacionado, entonces la Cov($\boldsymbol{a}_t$, $\boldsymbol{r}_{t-l}$)=**0**, para l>0. Por esta razón $\boldsymbol{a}_t$ es llamada *shock* o *innovación* de la serie en el período t.

- ✓ Postmultiplicando la expresión anterior por $\boldsymbol{a}_t^{'}$, tomando esperanza, y usando el hecho de que no hay correlación serial en $\boldsymbol{a}_t$, obtenemos que Cov($\boldsymbol{r}_t$, $\boldsymbol{a}_t$)=$\boldsymbol{\Sigma}$.

- ✓ El proceso VAR(1) depende de la innovaciones pasadas $\boldsymbol{a}_{t-j}$ a través de los coeficientes $\boldsymbol{\Phi}^{j}$. Para que la dependencia tenga sentido, $\boldsymbol{\Phi}^{j}$ debe converger a una matriz nula a medida que $j \to \infty$. Para que esto se cumpla, **los k valores propios de Φ deben tener módulo menor que 1** (Ver, por ejemplo, Wei (2006), pág, 389). Esta es la condición necesaria y suficiente para la estacionaridad débil de $\boldsymbol{r}_t$, dado que la matriz de covarianzas de $\boldsymbol{a}_t$ exista.

- ✓ Esta condición de estacionalidad se reduce a que $|\phi|<1$ en un modelo univariado.

- ✓ Puesto que $|\lambda I - \Phi| = \lambda^k \left| I - \Phi \frac{1}{\lambda} \right|$, los valores propios de Φ son los inversos de los ceros del determinante $|I - \Phi B|$. Por tanto, una condición suficiente y necesaria para la estacionalidad de $\boldsymbol{r}_t$ equivalente es que los ceros del determinante $|\Phi(B)| = |I - \Phi(B)|$ tengan módulo mayor que 1.

- ✓ Usando la expresión $\tilde{\boldsymbol{r}}_t = \boldsymbol{a}_t + \boldsymbol{\Phi a}_{t-1} + \boldsymbol{\Phi}^2 \boldsymbol{a}_{t-2} + \boldsymbol{\Phi}^3 \boldsymbol{a}_{t-3} + \ldots$ obtenemos que

$$\text{Cov}(\boldsymbol{r}_t) = \Gamma_0 = \Sigma + \Phi\Sigma\Phi' + \Phi^2\Sigma(\Phi^2)' + \ldots = \sum_{i=0}^{\infty} \Phi^i \Sigma (\Phi^i)'$$

donde $\boldsymbol{\Phi}^0 = \mathrm{I}$.

- ✓ Postmultiplicando por $\tilde{\boldsymbol{r}}_{t-l}^{'}$ a $\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}\tilde{\boldsymbol{r}}_{t-1} + \boldsymbol{a}_t$, tomando esperanza y usando el resultado Cov($\boldsymbol{r}_{t-l}, \boldsymbol{a}_t$)=**0**, para l>0 obtenemos

$$\text{E}(\tilde{\boldsymbol{r}}_t\, \tilde{\boldsymbol{r}}_{t-l}^{'}) = \boldsymbol{\Phi}\, \text{E}(\tilde{\boldsymbol{r}}_{t-1}\, \tilde{\boldsymbol{r}}_{t-l}^{'}), \quad l>0.$$

Por tanto,

$$\boldsymbol{\Gamma}_l = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{l-1}, \quad l>0$$

$$\boldsymbol{\Gamma}_l = \boldsymbol{\Phi}^l\boldsymbol{\Gamma}_0, \quad l>0.$$

Esto es una generalización de lo que ocurre en el modelo univariado AR(1).

- ✓ Pre y post multiplicando a $\boldsymbol{\Gamma}_l = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{l-1}$ por la inversa de la matriz D=diag$\left[\sqrt{\gamma_{11}(0)}, \sqrt{\gamma_{22}(0)}, \cdots, \sqrt{\gamma_{kk}(0)}\right]$, se obtiene

$$\rho_l = D^{-1}\Gamma_l D^{-1} = D^{-1}\Phi\Gamma_{l-1}D^{-1} = D^{-1}\Phi D^{-1}D^{-1}\Gamma_{l-1}D^{-1} = \gamma\rho_{l-1}$$

donde $\gamma = D^{-1}\Phi D^{-1}$. Por tanto, el VAR(1) satisface

$$\boldsymbol{\rho}_l = \boldsymbol{\gamma}^l\boldsymbol{\rho}_0 \text{ para } l>0.$$

Modelos Vectoriales AR(p)

- La generalización del VAR(1) al VAR(p) es directa. La serie de tiempo $\boldsymbol{r}_t$ sigue un proceso VAR(p) si satisface

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \ldots + \boldsymbol{\Phi}_p\boldsymbol{r}_{t-p} + \boldsymbol{a}_t, \quad \text{p>0}$$

donde $\boldsymbol{\phi}_0$ y $\boldsymbol{a}_t$ están definidas como antes y $\boldsymbol{\Phi}_j$ son matrices de k x k.

Usando el operador de rezagos B, el VAR(p) se puede escribir

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 \boldsymbol{B} - \ldots - \boldsymbol{\Phi}_p \boldsymbol{B}^p)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t ,$$

En forma compacta,

$$\boldsymbol{\Phi}(\boldsymbol{B})\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t$$

donde $\boldsymbol{\Phi}(\boldsymbol{B}) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 \boldsymbol{B} - \ldots - \boldsymbol{\Phi}_p \boldsymbol{B}^p)$, es la matriz polinomial.

- $\boldsymbol{r}_t$ es débilmente estacionaria si los ceros de $|\boldsymbol{I} - \boldsymbol{\Phi}_1 \boldsymbol{B} - \ldots - \boldsymbol{\Phi}_p \boldsymbol{B}^p|$ caen fuera del círculo unidad.

- Si $\boldsymbol{r}_t$ es débilmente estacionaria,

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 - \ldots - \boldsymbol{\Phi}_p)^{-1} \boldsymbol{\phi}_0 = [\boldsymbol{\Phi}(1)]^{-1} \boldsymbol{\phi}_0$$

Sea $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\mu}$, la serie corregida por la media. Entonces el Var(p) se puede escribir como:

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}_1 \tilde{\boldsymbol{r}}_{t-1} + \ldots + \boldsymbol{\Phi}_p \tilde{\boldsymbol{r}}_{t-p} + \boldsymbol{a}_t$$

- Usando esta expresión y las mismas técnicas que para el VAR(1), obtenemos

  ✓ Cov($\boldsymbol{r}_t$, $\boldsymbol{a}_t$)=$\boldsymbol{\Sigma}$

  ✓ Cov($\boldsymbol{r}_{t-l}$, $\boldsymbol{a}_t$)=**0**, para l>0

  ✓ $\boldsymbol{\Gamma}_l = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{l-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{\Gamma}_{l-p}$, l>0

Para l>0, la última propiedad presenta las denominadas ecuaciones de momento del modelo VAR(p). Es una versión multivariada de las ecuaciones de Yule-Walker del modelo univariado AR(p).

- Las propiedades del modelo VAR(p) pueden ser entendidas haciendo uso del modelo VAR(1). Esto se puede lograr transformando el VAR(p) de $\boldsymbol{r}_t$ en un modelo VAR(1) kp dimensional.

  - ✓ Específicamente, sean $\boldsymbol{x}_t=(\tilde{\boldsymbol{r}}_{t-p+1}',\tilde{\boldsymbol{r}}_{t-p+2}',\cdots,\tilde{\boldsymbol{r}}_t')'$ y $\mathbf{b}_t=(0,\ \ldots,\ 0,\boldsymbol{a}_t')'$, dos procesos kp dimensionales. La media de $\mathbf{b}_t$ es un vector de ceros y su matriz de covarianzas es una matriz de kp x kp con ceros en todas partes excepto en la esquina derecha inferior, la cual es $\boldsymbol{\Sigma}$.

  - ✓ El modelo VAR(p) puede ser escrito

$$\boldsymbol{x}_t = \boldsymbol{\Phi}^* \boldsymbol{x}_{t-1} + \boldsymbol{b}_t$$

donde $\boldsymbol{\Phi}^*$ es una matriz de kp x kp dada por

$$\boldsymbol{\Phi}^* = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \\ \boldsymbol{\Phi}_\mathbf{p} & \boldsymbol{\Phi}_\mathbf{p-1} & \boldsymbol{\Phi}_\mathbf{p-2} & \boldsymbol{\Phi}_\mathbf{p-3} & \cdots & \boldsymbol{\Phi}_\mathbf{1} \end{bmatrix}$$

donde **0** e $\mathbf{I}$ son las matrices nulas e identidad de k x k, respectivamente. En la literatura, la matriz $\boldsymbol{\Phi}^*$ es llamada la matriz *compañera* de la matriz polinomial $\boldsymbol{\Phi}(\mathrm{B})$.

  - ✓ La ecuación $\mathbf{x}_\mathbf{t} = \boldsymbol{\Phi}^*\mathbf{x}_\mathbf{t-1} + \mathbf{b}_\mathbf{t}$ es un modelo VAR(1) para $\mathbf{x}_t$, la cual contiene a $\boldsymbol{r}_t$ como su última componente. Las propiedades del modelo VAR(1) vistas anteriormente, pueden ser usadas para derivar las propiedades de modelo VAR(p). Por ejemplo, de la definición, $\mathbf{x}_t$ es débilmente estacionario si $\boldsymbol{r}_t$ lo es. Por tanto la condición de estacionaridad débil del VAR(p) es que todos los valores propios de $\boldsymbol{\Phi}^*$ tengan módulo menor que 1.

- En el análisis de series financieras, es de mucha importancia la estructura de la matriz de coeficientes $\boldsymbol{\Phi}_l$ de un modelo VAR(p). Por ejemplo si $\boldsymbol{\Phi}_{ij}(\boldsymbol{l})$ =0 para todo l, entonces r_{it} no depende de los valores pasados de r_{jt}. Por tanto, la estructura de las matrices $\boldsymbol{\Phi}_l$ proporciona información sobre las relaciones de adelanto-rezago entre as componentes de $\boldsymbol{r}_t$.

Construcción de un Modelo VAR(p)

- Para construir un modelo VAR(p), se usa de nuevo el procedimiento de identificación, estimación y chequeo del modelo.

- El concepto de función de autocorrelación parcial puede ser generalizado para determinar el orden del VAR. Considere la sucesión de modelos VAR

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{a}_t$$

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{r}_{t-2} + \boldsymbol{a}_t$$

$$\vdots$$

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \ldots + \boldsymbol{\Phi}_i\boldsymbol{r}_{t-i} + \boldsymbol{a}_t$$

$$\vdots$$

Los parámetros de estos modelos pueden ser estimados usando el método de mínimos cuadrados ordinarios (OLS). En el análisis estadístico multivariado este procedimiento se conoce como la estimación de una regresión lineal multivariada (Ver, por ejemplo, Johnson y Wichern, 1998).

Para la i-ésima ecuación (el VAR(i)), sean $\hat{\Phi}_j^{(i)}$ y $\hat{\phi}_0^{(i)}$ las estimaciones OLS de $\mathbf{\Phi}_j$ y ϕ_0. Entonces el residual es

$$\hat{\boldsymbol{a}}_t^{(i)} = r_t - \hat{\phi}_0^{(i)} - \hat{\Phi}_1^{(i)} r_{t-1} - \cdots - \hat{\Phi}_i^{(i)} r_{t-i}$$

La matriz de covarianza residual está definida como

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{\boldsymbol{T} - 2\boldsymbol{i} - 1} \sum_{t=i+1}^{T} \hat{\boldsymbol{a}}_t^{(i)} (\hat{\boldsymbol{a}}_t^{(i)})', \quad \boldsymbol{i} \geq 0$$

- Para especificar el orden p del VAR, se puede probar la hipótesis nula H0:$\mathbf{\Phi}_l = \mathbf{0}$, contra la hipótesis alternativa H1:$\mathbf{\Phi}_l \neq \mathbf{0}$, secuencialmente para l=1, 2, 3, … Por ejemplo, usando la ecuación de un VAR(1), podemos probar H0:$\mathbf{\Phi}_1 = 0$, contra la hipótesis alternativa H1:$\mathbf{\Phi}_1 \neq 0$. El estadístico de la prueba es

$$M(1) = -(T - k - 1 - \frac{3}{2}) \ln\left(\frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_0|} \right)$$

donde $|\hat{\Sigma}_i|$ indica el determinante de la matriz $\hat{\Sigma}_i$ definida antes, i=0,1. Bajo H_0 (y algunas condiciones de regularidad) $M(1) \sim \chi^2_{k^2}$ cuando $n \rightarrow \infty$. Si se rechaza H_0, procedemos a probar en un VAR(2) la hipótesis H0:$\mathbf{\Phi}_2 = 0$, contra la hipótesis alternativa H1:$\mathbf{\Phi}_2 \neq 0$, usando el estadístico

$$M(2) = -(T - k - 2 - \frac{3}{2}) \ln\left(\frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} \right)$$

Bajo H_0 (y algunas condiciones de regularidad) $M(2) \sim \chi^2_{k^2}$ cuando $n \rightarrow \infty$.

Si se rechaza el nuevo H_0, se procedería en forma similar usando un VAR(3), y así se continua hasta que no se rechace el nuevo H_0.

En general, si se quiere probar que el orden del VAR es i, se usan los modelos para el VAR(i-1) y el VAR(i) y se prueba la hipótesis H_0: $\boldsymbol{\Phi}_i = 0$, contra la hipótesis alternativa H_1: $\boldsymbol{\Phi}_i \neq 0$, usando el estadístico

$$M(i) = -(T - k - i - \frac{3}{2}) \ln\left(\frac{|\hat{\boldsymbol{\Sigma}}_i|}{|\hat{\boldsymbol{\Sigma}}_{i-1}|} \right)$$

Bajo H_0 (y algunas condiciones de regularidad) $M(i) \sim \chi^2_{k^2}$ cuando $n \rightarrow \infty$. Si no se rechaza H_0, el orden del VAR es i-1.

- Alternativamente se puede usar el criterio de información de Akaike (AIC) o sus variantes para seleccionar el orden p.

- Si suponemos que $\boldsymbol{a}_t$ sigue una distribución normal multivariada y que el modelo es un VAR(i), podemos estimar el modelo usando máxima verosimilitud (ML). Para modelos AR, los estimadores OLS y ML (condicional) de $\boldsymbol{\Phi}_j$ y $\boldsymbol{\phi}_0$ son equivalentes. Sin embargo hay diferencia entre los estimadores OLS y ML para $\boldsymbol{\Sigma}$. El estimador ML para $\boldsymbol{\Sigma}$ es

$$\tilde{\boldsymbol{\Sigma}}_i = \frac{1}{\boldsymbol{T}} \sum_{t=i+1}^{T} \hat{\boldsymbol{a}}_t^{(i)} (\hat{\boldsymbol{a}}_t^{(i)})', \quad \boldsymbol{i} \geq 0$$

El AIC de un modelo VAR(i) bajo normalidad está definido como

$$AIC(i) = \ln\left(|\, \tilde{\Sigma}_i \,|\right) + \frac{2k^2 i}{T}$$

Selecciona el modelo VAR(p) si AIC(p)=min{AIC(i) | i=1,2,,,,,p_0}, donde p_0 es un número preespecificado conocido tal que p_0>p.

- Otros criterios de información disponibles para moldeos VAR(i) son

$$BIC(i) = \ln\left(|\tilde{\Sigma}_i|\right) + \frac{k^2 i \ln(T)}{T}$$

$$HQ(i) = \ln\left(|\tilde{\Sigma}_i|\right) + \frac{2k^2 i \ln(\ln(T))}{T}$$

El criterio HQ fue propuesto por Hannan y Quinn (1979).

Ejemplo.
Suponiendo que la serie bivariada de los retornos mensuales de las acciones de IBM y del índice S&P 500 siguen un modelo VAR, la siguiente tabla muestra el cálculo de los estadísticos M(i) y del criterio AIC para varios rezagos.

Orden	1	2	3	4	5	6
M(i)	9.81	8.93	12.57	6.08	9.56	2.80
AIC(i)	6.757	6.756	6.750	6.753	6.751	6.756

Los dos estadísticos indican que el modelo VAR(3) puede ser adecuado para los datos. El estadístico M(i) es marginalmente significante al nivel del 5% en los rezagos 1, 3 y 5, puesto que el valor crítico es 9.488. Sin embargo, no es significante al 1%, pues el valor crítico es 13.277, confirmando la observación anterior de que la dependencia lineal dinámica entre las dos series es débil.

Estimación y Chequeo del Modelo

- Para la estimación de un VAR(p) se puede emplear la estimación OLS o ML. Los dos métodos son asintóticamente equivalentes. Bajo algunas condiciones de regularidad, los estimadores son asintóticamente normales.

- El modelo estimado debe ser cuidadosamente chequeado. Para probar que no hay correlación serial en los residuales, se puede usar el estadístico $Q_k(m)$ el cual tiene una distribución asintótica chi-cuadrado con k^2m-g grados de libertad, donde g es el número de parámetros estimados en las matrices de coeficientes AR.

Ejemplo.

En el ejemplo anterior, se identificó un Var(3) para la serie bivariada de los retornos de las acciones de IBM y del índice S&P 500. La especificación del modelo es

$$r_t = \phi_0 + \Phi_1 r_{t-1} + \Phi_3 r_{t-3} + a_t$$

Donde la primera componente se refiere a los retornos de las acciones de IBM. En este ejemplo se usan solamente las matrices de coeficientes en los rezagos 1 y 3, debido a la débil dependencia serial de los datos. En general, cuando se identifica un VAR(3), se usan las matrices de coeficientes en los rezagos 1, 2 y 3. La siguiente tabla presenta los resultados de la estimación del modelo completo.

Parámetros	ϕ_0	Φ_1	Φ_3	Σ
(a) Modelo Completo				
Estimación	1.20	0.011 0.108	0.039 -0.112	44.44 23.51
	0.58	-0.013 0.084	-0.007 -0.105	23.51 31.29
Error estánd.	0.23	0.043 0.051	0.044 0.052	
	0.19	0.036 0.043	0.037 0.044	
(b) Modelo Simplificado				
Estimación	1.24	0 0.117	0 -0.083	44.48 23.51
	0.57	0 0.073	0 -0.109	23.51 31.29
Error estánd.	0.23	- 0.040	- 0.040	
	0.19	- 0.033	- 0.033	

- ✓ La tabla (b) presenta los resultados de la estimación después de eliminar algunos parámetros no significativos.
- ✓ Algunos estadísticos $Q_k(m)$ de la serie de residuales para el modelo de la tabla (b) están dados por $Q_2(4)=18.17$ y $Q_2(8)=41.26$. Los valores p, obtenidos de una distribución asintótica con 12 y 28 grados de libertad, son 0.111 y 0.051, y por tanto el modelo es adecuado a un nivel del 5%.
- ✓ Como vimos antes, las series individuales poseen heterocedasticidad condicional. Más adelante se verán modelos para volatilidad multivariada.

- ✓ Para el modelo de la tabla (b):

  - ⇒ La correlación contemporánea entre las dos series de innovaciones es $23.51/(44.48*31.29)^{0.5}=0.63$, la cual está muy cerca del coeficiente de correlación muestral entre r_{1t} y r_{1t}.
  - ⇒ Las dos series de retornos tienen medias positivas y significativas, lo cual implica que los logaritmos de los precios de las dos series tienen una tendencia hacia arriba sobre el período muestral.
  - ⇒ Del modelo estimado obtenemos

$$IBM_t = 1.24 + 0.117SP500_{t-1} - 0.083SP500_{t-3} + \hat{a}_{1t}$$
$$SP500_t = 0.57 + 0.0773SP500_{t-1} - 0.109SP500_{t-3} + \hat{a}_{2t}$$

Por tanto, a un nivel de significancia del 5%, existe una relación dinámica unidireccional de los retornos del índice SP 500 a los retornos de IBM. Esto implica que si el índice S&P 500 representa el mercado de acciones de EU, entonces el retorno de IBM está afectado por movimientos pasados del mercado. Sin embargo, movimientos pasados de los retornos de IBM no afectan significativamente el mercado de EU.

$\Rightarrow$ El modelo bivariado ajustado puede ser escrito como

$$\begin{bmatrix} IBM_t \\ SP500_t \end{bmatrix} = \begin{bmatrix} 1.24 \\ 0.57 \end{bmatrix} + \begin{bmatrix} 0.117 \\ 0.073 \end{bmatrix} SP500_{t-1} + \begin{bmatrix} 0.083 \\ 0.109 \end{bmatrix} SP500_{t-3} + \begin{bmatrix} a_{1t} \\ a_{1t} \end{bmatrix}$$

lo cual que indica que S&P 500 es el factor que conduce la serie bivariada.

Pronósticos

- Como en el caso univariado, podemos obtener los pronósticos y los errores estándar de los errores de pronóstico asociados.

- Para un modelo VAR(p), el pronóstico 1-paso adelante en el período de origen h es

$$r_h(1) = \phi_0 + \sum_{i=1}^{p} \Phi_i r_{h+1-i}$$

y el error de pronóstico asociado es $e_h(1) = a_{h+1}$, con matriz de covarianza Σ.

- Para un modelo VAR(p), el pronóstico 2-pasos adelante en el período de origen h es

$$r_h(2) = \phi_0 + \Phi_1 r_h(1) + \sum_{i=2}^{p} \Phi_i r_{h+2-i}$$

y el error de pronóstico asociado es

$e_h(2) = a_{h+2} + \Phi_1 a_{h+1}$, con matriz de covarianza $\Sigma + \Phi_1 \Sigma \Phi_1'$.

- Si $\boldsymbol{r}_t$ es débilmente estacionario entonces,
  - ✓ El pronóstico l-pasos adelante, $r_h(l)$ converge a su vector de medias μ, a medida que el horizonte del pronóstico se incrementa.

  - ✓ La matriz de covarianzas del error del pronóstico converge a la matriz de covarianzas de $\boldsymbol{r}_t$.

Ejemplo

Para el modelo refinado en el ejemplo anterior, la siguiente tabla proporciona los pronósticos hasta 6 períodos adelante en el origen h=888.

Horizonte	1	2	3	4	5	6
Pronóstico IBM	1.40	1.12	0.82	1.21	1.27	1.31
Error Estándar	6.67	6.70	6.70	6.72	6.72	6.72
Pronóstico S&P500	0.32	0.38	-0.02	0.53	0.56	0.61
Error Estándar	5.59	5.61	5.61	5.64	5.64	5.64

Comandos del SCA: miden, mtsm, mestim. mforecast

```
input var ibm sp500. file 'c:\datos\m_ibm_sp500.txt'.

c  Selección del orden
miden var ibm sp500. arfits 1 to 6.

c  Especificación de un Var(3)  con rezagos 1 y 3 solamente.
mtsm modvar3. series ibm sp500.  @
model (i-p1*B-p3*B**3)series=c+noise.

c Estimación del modelo completo
mestim modvar3.

c  Asignación de ceros a los parámetros no significativos
p1(1,1)=0
p1(2,1)=0
```

```
p3(1,1)=0
p3(2,1)=0


c  Asignación de restricciones a los parámetros fijos
cp1(1,1)=1
cp1(2,1)=1
cp3(1,1)=1
cp3(2,1)=1

c estimación del modelo simplificado
mestim modvar3. hold residuals(res1, res2).

c pronósticos
mforecast modvar3. hold forecasts(pron_ibm pron_sp)  @
             std(std_ibm std_sp). nofs 6.
```

Función Impulso Respuesta

- Como en el caso univariado, un modelo VAR(p) puede ser escrito como una función de las innovaciones pasadas

$$r_t = \mu + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + ...$$

donde $\mu = [\Phi(1)]^{-1}\phi_0$, dado que la inversa exista. Las matrices coeficientes Ψ_i se pueden obtener igualando los coeficientes de B^i en la ecuación

$$(I - \Phi_1 B - \Phi_2 B^2 - ... - \Phi_p B^p)(I + \Psi_1 B + \Psi_2 B^2 + ...) = I$$

- Esta es una representación de medias–móviles de $\boldsymbol{r}_t$, donde las matrices de coeficientes Ψ_i indican el impacto de la innovación pasada a_{t-i} sobre $\boldsymbol{r}_t$. Equivalentemente Ψ_i es el efecto de a_t sobre $\boldsymbol{r}_{t+i}$.

- Ψ_i es denominada la función impulso respuesta de $\boldsymbol{r}_t$. Sin embargo, como las componentes de a_t generalmente están correlacionadas, la interpretación de Ψ_i no es directa. Para ayudar a la interpretación se puede usar la descomposición de Cholesky mencionada anteriormente, que permite transformar las innovaciones de manera que queden incorrelacionadas.

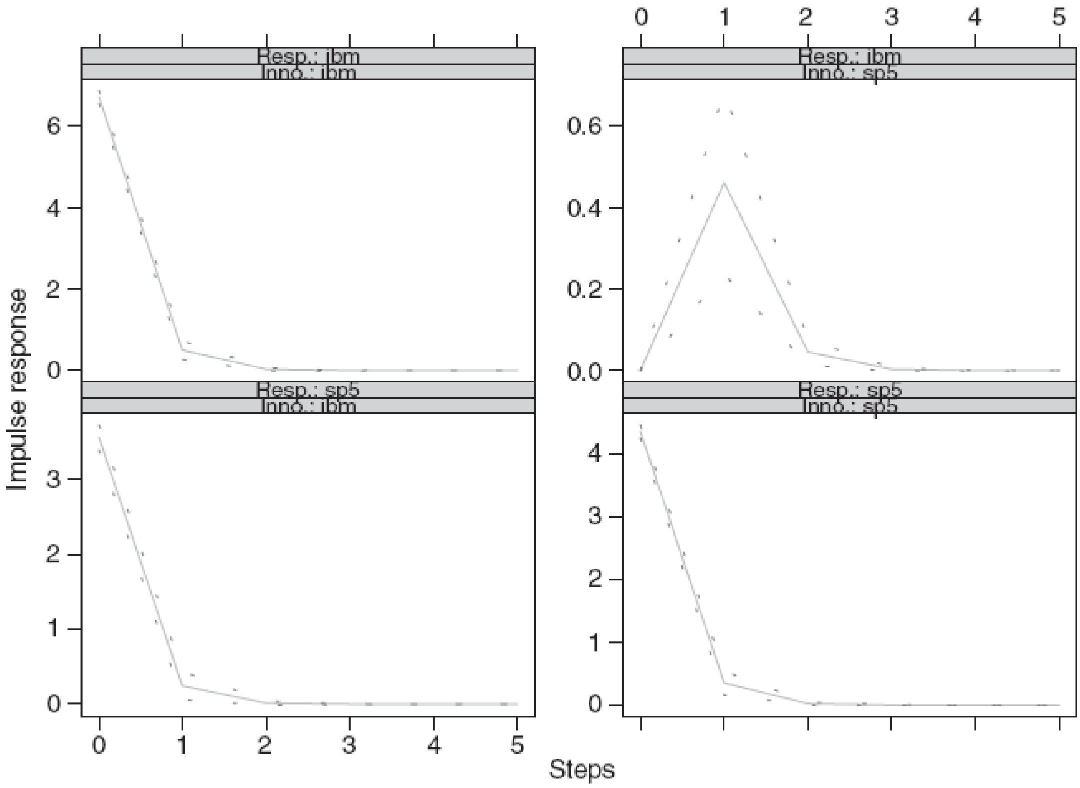
- ✓ Específicamente, existe una matriz triangular inferior con unos en su diagonal *L* tal que $\Sigma = LGL'$, donde *G* es una matriz diagonal.
- ✓ Sea b=$L^{-1} a_t$. Entonces Cov(b)=*G*, de manera que las componentes de b están no correlacionadas.
- ✓ La ecuación $r_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \ldots$ puede ser escrita como

$$\begin{aligned} r_t &= \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \ldots \\ &= \mu + LL^{-1} a_t + \Psi_1 LL^{-1} a_{t-1} + \Psi_2 LL^{-1} a_{t-2} + \ldots \\ &= \mu + \Psi_0^* b_t + \Psi_1^* b_{t-1} + \Psi_2^* b_{t-2} + \ldots \end{aligned}$$

  donde $\Psi_0^* = L$ y $\Psi_i^* = \Psi_i L$. Las matrices Ψ_i^* son llamadas la *función impulso respuesta* de $\boldsymbol{r}_t$.

- ✓ Específicamente, $\Psi_{ij}^*(l)$, el (i,j)-ésimo elemento de Ψ_l^*, es el impacto de $b_{j,t}$ sobre la observación futura $r_{i,t+l}$.
- ✓ En la práctica, se puede normalizar la innovación ortogonal b_t de forma que tenga varianza 1.
- ✓ Una debilidad de la ortogonalización anterior es que el resultado depende del ordenamiento de las componentes $\boldsymbol{r}_t$. Diferentes ordenamientos de las componentes de $\boldsymbol{r}_t$, pueden conducir a diferentes funciones impulso respuesta.

Ejemplo. El siguiente gráfico contiene la función impulso respuesta para un modelo VAR(1) ajustado a las series de los retornos de las acciones IBM y S&P500, cuando los retornos de IBM es la primera serie del vector. Observe que debido a la dependencia débil de los retornos, las funciones presentan patrones simples y decaen rápidamente.



Modelos de Medias Móviles Vectoriales

- Un modelo de medias móviles vectorial de orden q, o VMA(q), es de la forma:

$$\boldsymbol{r_t} = \theta_0 + a_t + \Theta_1 a_{t-1} + \ldots + \Theta_q a_{t-q}, \quad \text{o} \quad \boldsymbol{r_t} = \theta_0 + \Theta(B) a_t$$

donde θ_0 y $\boldsymbol{a}_t$ es un vector k-dimensional de ruido blanco y $\boldsymbol{\Theta}_j$ son matrices de k x k y $\boldsymbol{\Theta}(\boldsymbol{B}) = \boldsymbol{I} - \boldsymbol{\Theta}_1 \boldsymbol{B} - \ldots - \boldsymbol{\Theta}_q \boldsymbol{B}^q$ es la matriz polinomial MA en el operador de rezagos *B*.

- Como en el caso univariado, el proceso VMA(q) es débilmente estacionario dado que la matriz de covarianza $\boldsymbol{\Sigma}$ de $\boldsymbol{a}_t$ exista.

- Tomando esperanza, $\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{r}_t) = \theta_0$.

- Sea $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0$. Entonces se cumple que:

  - ✓ Cov($\boldsymbol{r}_t, \boldsymbol{a}_t$)=$\boldsymbol{\Sigma}$
  - ✓ $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1\boldsymbol{\Sigma}\boldsymbol{\Theta}_1^{'} + ... + \boldsymbol{\Theta}_q\boldsymbol{\Sigma}\boldsymbol{\Theta}_q^{'}$
  - ✓ $\boldsymbol{\Gamma}_l$=**0** si l>q
  - ✓ $\boldsymbol{\Gamma}_l = \sum_{j=l}^{q} \Theta_j \Sigma \Theta_{j-l}^{'}$ si $1 \le \boldsymbol{l} \le \boldsymbol{q}$, donde $\boldsymbol{\Theta}_0 = -\boldsymbol{I}$

- Puesto que $\boldsymbol{\Gamma}_l$=**0** si l>q, las matrices de correlación cruzadas (CCM) de un proceso VMA(q) satisfacen, como en el caso univariado,

$$\boldsymbol{\rho}_l = 0, \quad si \;\; l > q$$

- Considere el modelo MA(1) bivariado o VMA(1),

$$\boldsymbol{r_t} = \theta_0 + a_t - \Theta_1 a_{t-1}$$

explícitamente,

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Theta}_{11} & \boldsymbol{\Theta}_{12} \\ \boldsymbol{\Theta}_{21} & \boldsymbol{\Theta}_{22} \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

El valor contemporáneo de $\boldsymbol{r_t}$ solamente depende del valor contemporáneo y pasado del shock. Por tanto, el modelo es de memoria finita.

- Considere la ecuación para r_{1t}.

  - ✓ El parámetro $\boldsymbol{\Theta}_{12}$ indica la dependencia lineal de r_{1t} sobre $a_{2,t-1}$ en presencia de $a_{1,t-1}$. Si $\boldsymbol{\Theta}_{12}$=0, entonces r_{1t} no depende de valores rezagados de a_{2t} y por tanto, de valores rezagados de r_{2t}.

- ✓ Similarmente, si Si $\mathbf{\Theta}_{21}$=0, entonces r_{2t} no depende de valores rezagados de r_{1t}.
- ✓ Por tanto los elementos fuera de la diagonal de $\mathbf{\Theta}_1$ muestran la dependencia dinámica entre las componentes de la serie.

- Para el modelo VMA(1), las relaciones entre r_{1t} y r_{2t} se pueden clasificar de la siguiente manera:

  - ✓ Son series no relacionadas dinámicamente si $\mathbf{\Theta}_{12}=\mathbf{\Theta}_{21}$=0.
  - ✓ Existe una relación dinámica unidireccional de r_{1t} a r_{2t} si $\mathbf{\Theta}_{12}$=0 y $\mathbf{\Theta}_{21} \neq 0$.
  - ✓ Existe una relación dinámica unidireccional de r_{2t} a r_{1t} si $\mathbf{\Theta}_{21}$=0 y $\mathbf{\Theta}_{12} \neq 0$.
  - ✓ Existe una relación de retroalimentación entre r_{1t} y r_{2t} si $\mathbf{\Theta}_{12} \neq 0$ y $\mathbf{\Theta}_{21} \neq 0$.
  - ✓ La correlación contemporánea entre r_{1t} y r_{2t} es la misma que entre a_{1t} y a_{2t}.

Estimación del VMA

- Como en el caso univariado, a diferencia del modelo VAR(p), la estimación del modelo VMA(q) es mucho más compleja. Vea, por ejemplo, Hillmer y Tiao (1979), Lutkepöhl (1991). Para la estimación máximo verosímil hay dos procedimientos: estimación máximo verosímil condicional y estimación máximo verosímil exacta.

✓ **Estimación máximo verosímil condicional del VMA(1)**

Supone que $\boldsymbol{a}_0$=**0**. Bajo este supuesto, podemos escribir $\boldsymbol{a}_t = \boldsymbol{r}_t - \theta_0 + \Theta_1 \boldsymbol{a}_{t-1}$, y podemos calcular recursivamente el shock como

$$\boldsymbol{a}1 = \boldsymbol{r}_1 - \theta_0, \quad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \theta_0 + \Theta_1 \boldsymbol{a}_1, \quad \ldots$$

Por tanto, la función de verosimilitud de los datos es

$$f(r_1,\ldots,r_T \mid \theta_0, \Theta_1, \Sigma) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{T/2} \mid \boldsymbol{\Sigma} \mid^{1/2}} \exp\left[-\frac{1}{2} \boldsymbol{a}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_t \right]$$

La cual puede ser evaluada para obtener los parámetros.

✓ **Estimación máximo verosímil exacta del VMA(1)**

Para el método de verosimilitud exacta, $\boldsymbol{a}_0$ es un vector desconocido y debe ser estimado usando los datos. Para simplificar, suponga que $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \theta_0$ es la serie corregida por la media. Entonces

$$\boldsymbol{a}_t = \tilde{\boldsymbol{r}}_t + \Theta \boldsymbol{a}_{t-1}$$

Por sustituciones repetidas obtenemos que $\boldsymbol{a}_0$ está relacionado con todos los $\tilde{\boldsymbol{r}}_t$ como

$$\begin{aligned}
\boldsymbol{a}_1 &= \tilde{\boldsymbol{r}}_1 + \Theta \boldsymbol{a}_0 \\
\boldsymbol{a}_2 &= \tilde{\boldsymbol{r}}_2 + \Theta \boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_2 + \Theta \tilde{\boldsymbol{r}}_1 + \Theta^2 \boldsymbol{a}_0 \\
\vdots &= \vdots \\
\boldsymbol{a}_T &= \tilde{\boldsymbol{r}}_T + \Theta \tilde{\boldsymbol{r}}_{T-1} + \ldots + \Theta^{T-1} \tilde{\boldsymbol{r}}_1 + \Theta^T \boldsymbol{a}_0
\end{aligned}$$

Por tanto, $\boldsymbol{a}_0$ es una función lineal de los datos si θ_0 y Θ son dados.

Este resultado permite estimar a $\boldsymbol{a}_0$ usando los estimadores iniciales para θ_0 y $\boldsymbol{\Theta}$. Concretamente, si definimos

$$\boldsymbol{r}_t^* = \tilde{\boldsymbol{r}}_t + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{t-1} + \ldots + \boldsymbol{\Theta}^{T-1}\tilde{\boldsymbol{r}}_1, \text{ para t=1,2,…,T}$$

Podemos reescribir las ecuaciones anteriores como

$$\begin{aligned}
\boldsymbol{r}_1^* &= -\boldsymbol{\Theta}\boldsymbol{a}_0 + \boldsymbol{a}_1 \\
\boldsymbol{r}_2^* &= -\boldsymbol{\Theta}^2\boldsymbol{a}_0 + \boldsymbol{a}_2 \\
\vdots &= \vdots \\
\boldsymbol{r}_T^* &= -\boldsymbol{\Theta}^T\boldsymbol{a}_0 + \boldsymbol{a}_T
\end{aligned}$$

Esto constituye un modelo de regresión lineal múltiple con vector de parámetros $\boldsymbol{a}_0$, aunque la matriz de covarianza $\boldsymbol{\Sigma}$ no es diagonal.

Si hay disponible un estimador inicial para $\boldsymbol{\Sigma}$, podemos pre-multiplicar cada una de las ecuaciones del sistema anterior por $\boldsymbol{\Sigma}^{-1/2}$. El sistema resultante es una regresión lineal múltiple y podemos usar el método de los mínimos cuadrados ordinarios para obtener una estimación de $\boldsymbol{a}_0$.

Sea $\hat{\boldsymbol{a}}_0$ dicho estimador. Los shocks $\boldsymbol{a}_t$ pueden ser calculados recursivamente como,

$$\boldsymbol{a}1 = \boldsymbol{r}_1 - \theta_0 + \boldsymbol{\Theta}\hat{\boldsymbol{a}}_0, \quad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \theta_0 + \boldsymbol{\Theta}_1\boldsymbol{a}_1, \quad \ldots$$

Además, podemos derivar la función de verosimilitud de los datos usando la distribución conjunta de {$\boldsymbol{a}_t$ | t=0,1,…,T}. De esta forma la función de verosimilitud resultante puede ser evaluada para obtener los estimadores máximo verosímiles exactos.

Resumiendo, la estimación máximo verosímil exacta trabaja de la siguiente forma:

✓ Dadas estimaciones iniciales para θ_0, $\boldsymbol{\Theta}$ y $\boldsymbol{\Sigma}$ se usan las ecuaciones

$$\begin{aligned} \boldsymbol{a}_1 &= \tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}\boldsymbol{a}_0 \\ \boldsymbol{a}_2 &= \tilde{\boldsymbol{r}}_2 + \boldsymbol{\Theta}\boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_2 + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}^2\boldsymbol{a}_0 \\ \vdots &= \vdots \\ \boldsymbol{a}_T &= \tilde{\boldsymbol{r}}_T + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{T-1} + \boldsymbol{\Theta}^{T-1}\tilde{\boldsymbol{r}}_1 + \ldots + \boldsymbol{\Theta}^T\boldsymbol{a}_0 \end{aligned}$$

para obtener una estimación de $\boldsymbol{a}_0$.

✓ Esta estimación es usada para calcular recursivamente a las $\boldsymbol{a}_t$ usando a ecuación

$$\boldsymbol{a}_t = \tilde{\boldsymbol{r}}_t + \boldsymbol{\Theta}\boldsymbol{a}_{t-1}$$

comenzando con $\boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}\hat{\boldsymbol{a}}_0$

✓ Las resultantes $\{\boldsymbol{a}_t\}$ t=1,…,T se emplean para evaluar la función de verosimilitud exacta de los datos para actualizar las estimaciones de θ_0, $\boldsymbol{\Theta}$ y $\boldsymbol{\Sigma}$.

✓ El proceso completo se repite hasta que las estimaciones convergen. Este proceso también es válido en el caso general de un VMA(q).

- La estimación máximo verosímil exacta requiere computación más intensiva que la máxima verosimilitud condicional, pero proporciona estimadores más precisos, especialmente cuando algunos valores propios de $\boldsymbol{\Theta}$ tienen módulos cercanos a 1.

- Si se sospecha de sobrediferenciación, la estimación máximo verosímil exacta es muy importante.

- Cuando el tamaño muestral no es muy grande, es preferible usar la estimación máximo verosímil exacta.

- En conclusión, la construcción de un modelo VMA contempla tres pasos:

  - ✓ Use las matrices de correlación cruzada muestrales para especificar el orden q. Para un VMA(q), $\boldsymbol{\rho}_l = 0, \quad si \;\; l > q$.
  - ✓ Estime el modelo usando estimación máximo verosímil condicional o estimación máximo verosímil exacta.
  - ✓ Realice chequeos sobre el modelo ajustado.

- Los pronósticos de un modelo VMA se obtienen de la misma forma que los de un modelo univariado MA.

Ejemplo

Considere de nuevo la serie bivariada de los retornos de las acciones de IBM y del índice S&P 500. Debido a que se presentaban correlaciones significantes en los rezagos 1 y 3 podemos usar un modelo VMA(3) de la forma

$$r_t = \theta_0 + a_t - \Theta_1 a_{t-1} - \Theta_3 a_{t-3}$$

La siguiente tabla muestra los resultados de la estimación.

Parámetros	θ_0	Θ_1	Θ_3	Σ

(a) Modelo Completo con Verosimilitud Condicional

Estimación	1.24 0.54	-0.013 -0.121 0.020 -0.101	-0.038 0.118 0.014 0.105	44.48 23.52 23.52 31.20
Error estánd.	0.24 0.18	0.043 0.051 0.036 0.043	0.044 0.052 0.036 0.043	

Parámetros	θ_0	Θ_1	Θ_3	Σ

(b) Modelo Completo con Verosimilitud Exacta

Estimación	1.24 0.54	-0.013 -0.121 0.020 -0.101	-0.038 0.118 0.013 0.105	44.48 23.52 23.52 31.20
Error estánd.	0.24 0.18	0.043 0.051 0.036 0.043	0.044 0.052 0.036 0.043	

(c) Modelo Simplificado con Verosimilitud Exacta

Estimación	1.24 0.54	0.000 -0.126 0.000 -0.084	0.000 0.082 0.000 0.114	44.54 23.51 23.51 31.21
Error estánd.	0.23 0.18	- 0.040 - 0.033	- 0.040 - 0.033	

- ✓ Los estadísticos $Q_k(m)$ para el modelo simplificado son $Q_2(4)=17.25$ y $Q_2(8)=39.30$. Comparado con la chi-cuadrado con 12 y 28 grados de libertad, los valores p son 0.1404 y 0.0762, respectivamente. El modelo es adecuado a un nivel del 5%.

- ✓ Para este ejemplo, la diferencia entre la verosimilitud condicional y exacta es pequeña. Esto se debe a que el tamaño muestral no es pequeño y lo más importante es que la estructura dinámica de los datos es débil.

- ✓ El modelo VMA(3) proporciona esencialmente la misma estructura dinámica que la del modelo VAR(3) vista anteriormente:
  - ⇒ Los retornos mensuales de IBM dependen de los retornos anteriores del índice S&P 500.
  - ⇒ Los retornos del mercado no dependen de los retornos pasados de IBM.
  - ⇒ La correlación contemporánea entre los dos retornos es fuerte.

Modelos ARMA Vectoriales

- Los modelos ARMA univariados pueden ser extendidos a vectores de series de tiempo. Los modelos resultantes son llamados modelos VARMA. Sin embargo, la generalización presenta algunos problemas que no aparecían en el caso univariado. Uno de ellos es el problema de la *identificabilidad*.

- A diferencia de los modelos ARMA univariados, los modelos VARMA pueden no estar definidos en forma única. Por ejemplo, el modelo VMA(1)

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{1t-1} \\ \boldsymbol{a}_{2t-1} \end{bmatrix}$$

Es *idéntico* al modelo VAR(1)

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1t-1} \\ r_{2t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

Su equivalencia puede ser observada examinando sus modelos componentes. Para el modelo VMA(1),

$$r_{1t} = a_{1t} - 2a_{2t-1}, \qquad r_{2t} = a_{2t}$$

Para el modelo VAR(1),

$$r_{1t} + 2r_{2t-1} = a_{1t}, \qquad r_{2t} = a_{2t}$$

Reemplazando r_{2t-1} por a_{2t-1}, vemos que los dos modelos son idénticos. Este problema de identificación no es dañino puesto que cualquiera de los modelos puede ser usado en una aplicación real.

- Hay otro tipo de problema de identificabilidad más problemático. Considere el modelo VARMA(1,1),

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1t-1} \\ r_{2t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

Este modelo es idéntico al modelo VARMA(1,1),

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2+\eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} r_{1t-1} \\ r_{2t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & \eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

para $\omega \neq 0$ y $\eta \neq 0$.

Esta clase de problema de identificabilidad es serio, sin restricciones apropiadas, la función de verosimilitud de un modelo vectorial

ARMA(1,1) no está definida en forma única, lo que produce una situación similar al problema de la multicolinealidad exacta en el análisis de regresión. Este problema puede ocurrir aún si una de las componentes es ruido blanco.

- Los ejemplos anteriores muestran los nuevos problemas que aparecen en la generalización a los modelos VARMA. Existen algunas propuestas para superar el problema de la identificabilidad, las cuales son denominadas métodos de *especificación estructural* (Tiao y Tsay, 1989). Sin embargo, en las aplicaciones financieras los modelos VAR y VMA son generalmente suficientes y no las discutiremos. Cuando los modelos VARMA son usados, generalmente modelos de bajos órdenes son empleados (por ejemplo, VARMA(1,1) o VARMA(2,1)) especialmente si las series no son estacionales.

- Un modelo VARMA(p,q) pude ser escrito como

$$\boldsymbol{\Phi}(\boldsymbol{B})\boldsymbol{r}_t = \phi_0 + \boldsymbol{\Theta}(\boldsymbol{B})\boldsymbol{a}_t$$

  donde $\boldsymbol{\Phi}(\boldsymbol{B}) = \boldsymbol{I} - \boldsymbol{\Phi}_1\boldsymbol{B} - ... - \boldsymbol{\Phi}_p\boldsymbol{B}^p$ y $\boldsymbol{\Theta}(\boldsymbol{B}) = \boldsymbol{I} - \boldsymbol{\Theta}_1\boldsymbol{B} - ... - \boldsymbol{\Theta}_q\boldsymbol{B}^q$ son dos matrices polinomiales de k x k. Suponemos que ellas no tiene factores comunes. En otro caso, el modelo puede ser simplificado.

- Condición necesaria y suficiente de de estacionaridad débil para $\boldsymbol{r_t}$, es la misma que la del VAR(p) para la matriz polinomial $\boldsymbol{\Phi}(\boldsymbol{B})$.

- Para $\boldsymbol{v}$ >0, los (i,j)-ésimos elementos de las matrices $\boldsymbol{\Phi}_v$ y $\boldsymbol{\Theta}_v$ miden, respectivamente, la dependencia lineal de r_{it} sobre $r_{j,t-v}$ y $\boldsymbol{a}_{j,t-v}$.

- Si el (i,j)-ésimo elemento de todas las matrices de coeficientes AR y MA son cero, entonces r_{it} no depende de los valores rezagados de r_{jt}. Sin embargo, lo contrario no es cierto en un modelo VARMA, es decir, pueden existir matrices de coeficientes AR y MA que no son cero y sin embargo r_{it} puede no depender de los valores rezagados de r_{jt}. Considere el siguiente modelo bivariante.

$$\begin{bmatrix} \boldsymbol{\Phi}_{11}(\boldsymbol{B}) & \boldsymbol{\Phi}_{12}(\boldsymbol{B}) \\ \boldsymbol{\Phi}_{21}(\boldsymbol{B}) & \boldsymbol{\Phi}_{22}(\boldsymbol{B}) \end{bmatrix} \begin{bmatrix} \boldsymbol{r}_{1t} \\ \boldsymbol{r}_{2t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Theta}_{11}(\boldsymbol{B}) & \boldsymbol{\Theta}_{12}(\boldsymbol{B}) \\ \boldsymbol{\Theta}_{21}(\boldsymbol{B}) & \boldsymbol{\Theta}_{22}(\boldsymbol{B}) \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{1t} \\ \boldsymbol{a}_{2t} \end{bmatrix}$$

Las condiciones *necesarias y suficientes* para la existencia de una relación univariante dinámica de r_{1t} hacia r_{2t} son

$$\boldsymbol{\Phi}_{22}(\boldsymbol{B})\boldsymbol{\Theta}_{12}(\boldsymbol{B}) - \boldsymbol{\Phi}_{12}(\boldsymbol{B})\boldsymbol{\Theta}_{22}(\boldsymbol{B}) = 0, \quad \text{pero}$$
$$\boldsymbol{\Phi}_{11}(\boldsymbol{B})\boldsymbol{\Theta}_{21}(\boldsymbol{B}) - \boldsymbol{\Phi}_{21}(\boldsymbol{B})\boldsymbol{\Theta}_{11}(\boldsymbol{B}) \neq 0,$$

- Estas condiciones pueden ser obtenidas de la siguiente manera. Sea

$$\boldsymbol{\Omega}(\boldsymbol{B}) = |\boldsymbol{\Phi}(\boldsymbol{B})| = \boldsymbol{\Phi}_{11}(\boldsymbol{B})\boldsymbol{\Phi}_{22}(\boldsymbol{B}) - \boldsymbol{\Phi}_{12}(\boldsymbol{B})\boldsymbol{\Phi}_{21}(\boldsymbol{B})$$

el determinante de la matriz polinomial AR: premultiplicando el modelo por la matriz

$$\begin{bmatrix} \boldsymbol{\Phi}_{22}(\boldsymbol{B}) & -\boldsymbol{\Phi}_{12}(\boldsymbol{B}) \\ -\boldsymbol{\Phi}_{21}(\boldsymbol{B}) & \boldsymbol{\Phi}_{11}(\boldsymbol{B}) \end{bmatrix}$$

podemos reescribir el modelo como

$$\boldsymbol{\Omega}(\boldsymbol{B}) \begin{bmatrix} \boldsymbol{r}_{1t} \\ \boldsymbol{r}_{2t} \end{bmatrix} =$$

$$\begin{bmatrix} \boldsymbol{\Phi}_{22}(\boldsymbol{B})\boldsymbol{\Theta}_{11}(\boldsymbol{B}) - \boldsymbol{\Phi}_{12}(\boldsymbol{B})\boldsymbol{\Theta}_{21}(\boldsymbol{B}) & \boldsymbol{\Phi}_{22}(\boldsymbol{B})\boldsymbol{\Theta}_{12}(\boldsymbol{B}) - \boldsymbol{\Phi}_{12}(\boldsymbol{B})\boldsymbol{\Theta}_{22}(\boldsymbol{B}) \\ \boldsymbol{\Phi}_{11}(\boldsymbol{B})\boldsymbol{\Theta}_{21}(\boldsymbol{B}) - \boldsymbol{\Phi}_{21}(\boldsymbol{B})\boldsymbol{\Theta}_{11}(\boldsymbol{B}) & \boldsymbol{\Phi}_{11}(\boldsymbol{B})\boldsymbol{\Theta}_{22}(\boldsymbol{B}) - \boldsymbol{\Phi}_{21}(\boldsymbol{B})\boldsymbol{\Theta}_{12}(\boldsymbol{B}) \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{1t} \\ \boldsymbol{a}_{2t} \end{bmatrix}$$

Consideremos la ecuación para r_{1t} y las condiciones vistas

$$\boldsymbol{\Phi}_{22}(\boldsymbol{B})\boldsymbol{\Theta}_{12}(\boldsymbol{B}) - \boldsymbol{\Phi}_{12}(\boldsymbol{B})\boldsymbol{\Theta}_{22}(\boldsymbol{B}) = 0, \quad \text{pero}$$

$$\boldsymbol{\Phi}_{11}(\boldsymbol{B})\boldsymbol{\Theta}_{21}(\boldsymbol{B}) - \boldsymbol{\Phi}_{21}(\boldsymbol{B})\boldsymbol{\Theta}_{11}(\boldsymbol{B}) \neq 0,$$

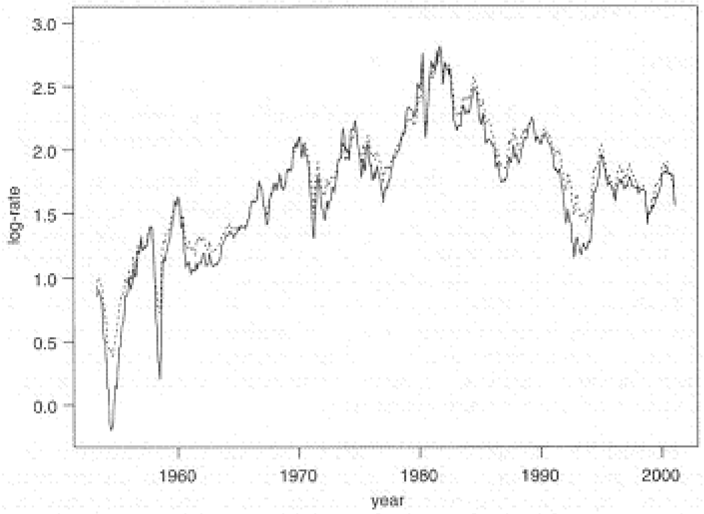
- ✓ La primera ecuación muestra que r_{1t} no depende de valores pasados de $\boldsymbol{a}_{2t}$ o de r_{2t}.
- ✓ De la ecuación para r_{2t}, la segunda condición implica que r_{2t} depende de algunos valores pasados de $\boldsymbol{a}_{1t}$.
- ✓ Basados en las ecuaciones anteriores, $\boldsymbol{\Theta}_{12}(\boldsymbol{B}) = \boldsymbol{\Phi}_{12}(\boldsymbol{B}) = \mathbf{0}$, es una condición suficiente, pero no necesaria para la existencia de una relación unidireccional de r_{1t} hacia r_{2t}.

Estimación de un modelo VARMA

- Puede ser realizada por máxima verosimilitud condicional o exacta.
- El estadístico $Q_k(m)$ continúa siendo válido, pero los grados de libertad de la chi-cuadrado deben ser cambiados por k^2m-g, donde g es el número de parámetros estimados en las dos matrices de parámetros AR y MA.

Ejemplo

Considere las siguientes series de tasas de interés mensuales de EU. La primera serie es la tasa del Tesoro con madurez a un año, y la segunda es tasa del Tesoro con madurez a tres años. Los datos van desde Abril de 1953 hasta Enero de 2001, para un total de 574 observaciones. Para asegurar que las tasas de interés de EU sean positivas, se analizan los logaritmos de las series. El siguiente gráfico presenta la evolución de las dos series.

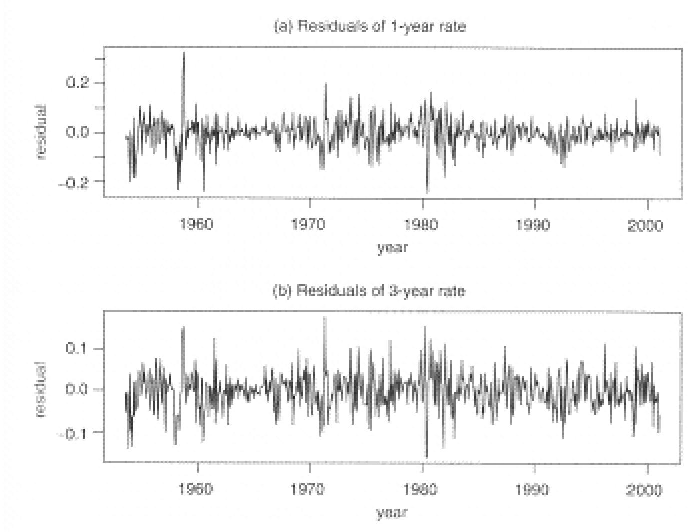


- ✓ Los criterios M(i) y AIC especifican un modelo VAR(4). Sin embargo se empleó un VARMA(2,1) debido a que los dos modelos proporcionan ajustes similares.

- ✓ La siguiente tabla presenta la estimación del modelo usando verosimilitud exacta.

Parameter	Φ_1		Φ_2		ϕ_0	Θ_1		$\Sigma \times 10^3$	
Estimate	1.82	−0.97	−0.84	0.98	0.028	0.90	−1.66	3.58	2.50
		0.99			0.025		−0.47	2.50	2.19
Standard error	0.03	0.08	0.03	0.08	0.014	0.03	0.10		
		0.01			0.011		0.04		

- ✓ Se removieron los parámetros no significativos y se re-estimó el modelo simplificado. Las series de residuales tienen correlaciones serial bajas en los rezagos 7 y 11. Las siguientes gráficas muestran los residuales y señalan la existencia de observaciones atípicas. El modelo puede ser mejorado, pero parece capturar razonablemente bien la estructura dinámica de los datos.



✓ El modelo VARMA final presenta algunas características interesantes de los datos:

⇒ Las series de las tasas de interés están altamente correlacionadas contemporáneamente. El coeficiente de correlación contemporáneo es $2.5/(3.58*2.19)^{0.5}=0.893$

⇒ Existe una relación lineal unidireccional de la tasa de interés a tres años hacia la tasa de interés a un año, debido a que los elementos (2,1) de todas las matrices AR y MA son cero, pero algunos elementos (1,2) no son cero.

⇒ Del modelo ajustado se obtiene

$$\begin{aligned} r_{3t} &= 0.025 + 0.99 r_{3,t-1} + a_{3t} + 0.47 a_{3,t-1} \\ r_{1t} &= 0.028 + 1.82 r_{1,t-1} - 0.84 r_{1,t-2} - 0.97 r_{3,t-1} + 0.98 r_{3,t-2} \\ &\quad + a_{1t} - 0.90 a_{1,t-1} + 1.66 a_{3,t-1} \end{aligned}$$

donde r_{it} es el logaritmo de la series a i-años y a_{it} es la serie de innovaciones correspondientes.

⇒ Las dos series parecen ser no estacionarias con raíces unitarias.

Modelos Marginales de las Componentes

- Dado un modelo vectorial para $\boldsymbol{r}_t$, los modelos implicados para las componentes r_{it} son los modelos *marginales*. Para un modelo k-dimensional ARMA(p,q) los modelos marginales son ARMA(kp,(k-1)p+q). Este resultado puede ser obtenido en dos pasos:

  ✓ Primero, los modelos marginales de un modelo VMA(q) son MA(q) univariados. Puesto que la matriz de correlación cruzada de $\boldsymbol{r}_t$ desaparece después del rezago q, la ACF de r_{it} se vuelve cero más allá del rezago q. Por tanto, r_{it} es un proceso MA y su modelo univariante es de la forma $r_{it} = \theta_{i,0} + \sum_{j=1}^{q} \theta_{i,j} b_{i,t-j}$, donde $\{b_{i,t}\}$ es una sucesión de variables aleatorias no correlacionadas con media cero y varianza $\sigma^2_{i,b}$. Los parámetros $\theta_{i,j}$ y $\sigma^2_{i,b}$ son funciones de los parámetros del modelo VMA para $\boldsymbol{r}_t$.

  ✓ El segundo paso se basa en la diagonalización de la matriz polinomial AR del modelo VARMA(p,q). Por ejemplo, considere el modelo bivariado AR(1)

$$\begin{bmatrix} 1-\boldsymbol{\Phi}_{11}B & -\boldsymbol{\Phi}_{12}B \\ -\boldsymbol{\Phi}_{21}B & 1-\boldsymbol{\Phi}_{22}B \end{bmatrix} \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

Premultiplicando por la matriz polinomial

$$\begin{bmatrix} 1-\boldsymbol{\Phi}_{22}B & -\boldsymbol{\Phi}_{12}B \\ -\boldsymbol{\Phi}_{21}B & 1-\boldsymbol{\Phi}_{11}B \end{bmatrix}$$

obtenemos

$$[(1-\mathbf{\Phi}_{11}B)(1-\mathbf{\Phi}_{22}B)-\mathbf{\Phi}_{12}\mathbf{\Phi}_{21}B^2]\begin{bmatrix} r_{1t} \\ r_{1t} \end{bmatrix}=\begin{bmatrix} 1-\mathbf{\Phi}_{22}B & -\mathbf{\Phi}_{12}B \\ -\mathbf{\Phi}_{21}B & 1-\mathbf{\Phi}_{11}B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

El lado izquierdo de la ecuación anterior muestra que los polinomios univariantes AR para las r_{it} son de orden 2. En contraste, el lado derecho de las ecuaciones está en forma VMA(1): por tanto, los modelos univariados para las r_{it} son ARMA(2,1).

- La técnica se generaliza fácilmente al modelo VAR(1) k-dimensional, donde los modelos marginales son ARMA(k, k-1).

- En general, para un VAR(p) k-dimensional, los modelos marginales son ARMA(kp, (k-1)p). El resultado general, para un modelo VARMA(p,q) se sigue de directamente de los de los modelos VMA y VAR.

- El orden [kp, (k-1)p+q] es el orden máximo para los modelos marginales. El orden marginal para r_{it} puede ser mucho menor.

No estacionaridad de Raíces Unitarias y Co-Integración

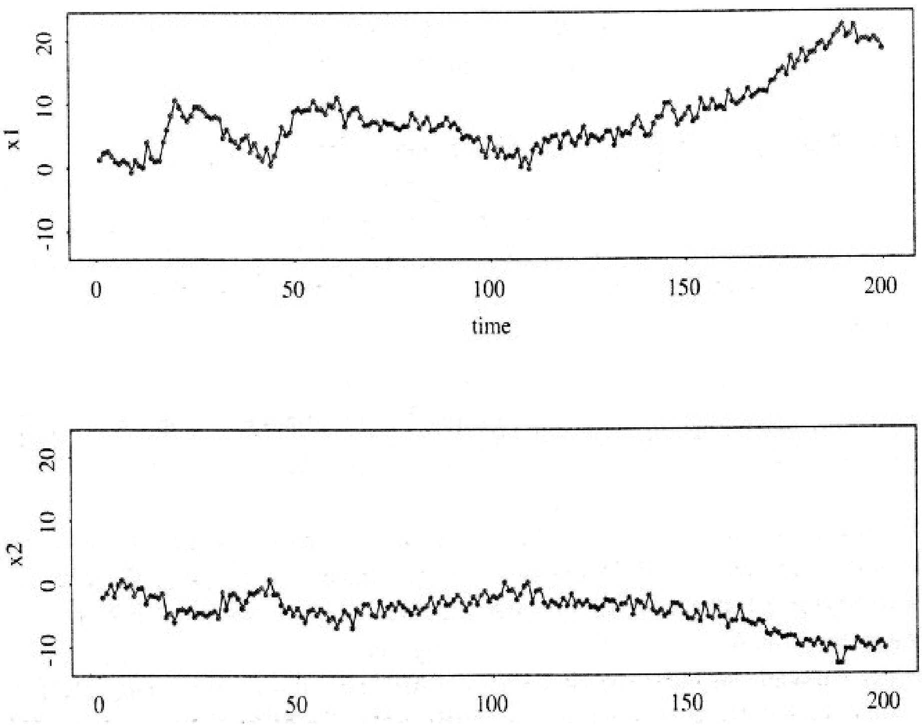
- El caso de la Co-integración puede suceder cuando se modelan varias series de tiempo conjuntamente.
- Considere el modelo bivariante

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}-\begin{bmatrix} 0.5 & -1.0 \\ -0.25 & 0.5 \end{bmatrix}\begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix}=\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}-\begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix}\begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

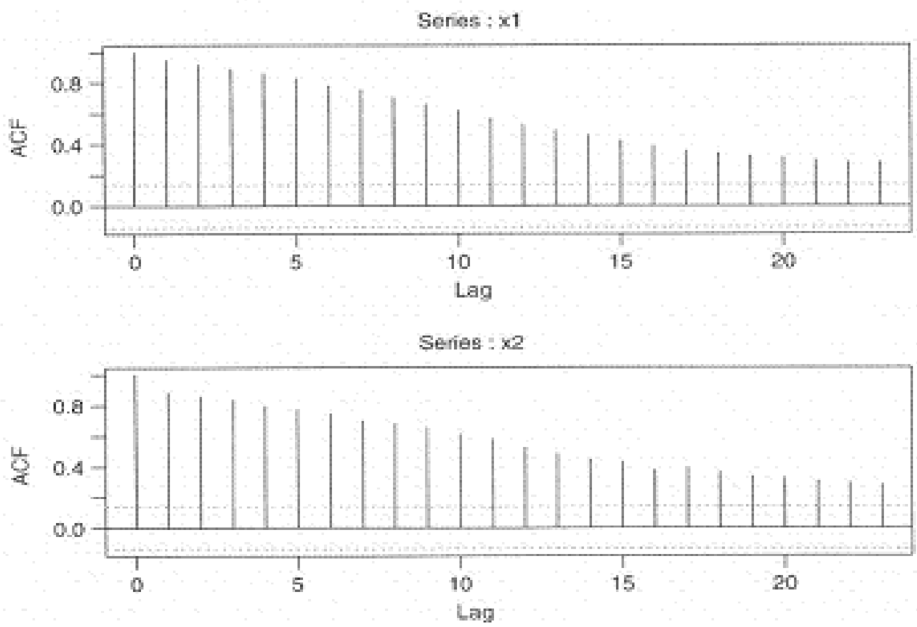
donde la matriz de covarianza Σ del shock $\boldsymbol{a}_t$ es definida positiva.

✓ Este proceso no es débilmente estacionario puesto que los valores propios de la matriz de coeficientes AR son 0 y 1.

✓ La siguiente gráfica muestra el comportamiento para series simuladas de 200 observaciones con $\Sigma = \mathbf{I}$



✓ Los correlogramas muestrales de las series son



- ✓ Las dos series están altamente autocorrelacionadas y exhiben características de no estacionaridad de raíces unitarias. El modelo se puede reescribir como

$$\begin{bmatrix} 1-0.5B & B \\ 0.25B & 1-0.5B \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.2B & -0.4B \\ 0.1B & 1-0.2B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

Premultiplicando la ecuación anterior por

$$\begin{bmatrix} 1-0.5B & -B \\ -0.25B & 1-0.5B \end{bmatrix}$$

Obtenemos

$$\begin{bmatrix} 1-B & 0 \\ 0 & 1-B \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.7B & -0.6B \\ -0.15B & 1-0.7B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

Por tanto cada componente x_{it} del modelo es no estacionaria y sigue un modelo ARIMA(0,1,1).

- ✓ Ahora consideremos la siguiente transformación lineal de $\mathbf{x}_t$:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \boldsymbol{L}\boldsymbol{x}_t$$

$$\begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} = \boldsymbol{L}\boldsymbol{a}_t$$

El modelo VARMA de la serie transformada se puede obtener de la siguiente manera.

$$\begin{aligned} \boldsymbol{L}\boldsymbol{x}_t &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{a}_{t-1} \end{aligned}$$

$$= \boldsymbol{L\Phi L}^{-1}(\boldsymbol{Lx}_{t-1}) + \boldsymbol{b}_t - \boldsymbol{L\Theta L}^{-1}\boldsymbol{b}_{t-1}$$

Por tanto el modelo para $\mathbf{y}_t$ es

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} - \begin{bmatrix} 1.0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} = \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} - \begin{bmatrix} 0.4 & -0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_{1t-1} \\ b_{2t-1} \end{bmatrix}$$

- ✓ Del modelo anterior observamos que:

  - ⇒ y_{1t} y y_{2t} son series no relacionadas dinámicamente con correlación contemporánea igual a la correlación entre los shocks b_{1t} y b_{2t}.
  - ⇒ y_{1t} sigue un proceso ARIMA(0,1,1).
  - ⇒ y_{2t} es un proceso de ruido blanco.

- ✓ En particular, el modelo anterior muestra que solamente hay una raíz unitaria en el sistema. Por tanto, las raíces unitarias de x_{1t} y x_{2t} son introducidas por la raíz unitaria de y_{1t}.

- El fenómeno para el cual x_{1t} y x_{2t} son no estacionarias de raíces unitarias, pero existe una única raíz unitaria en el vector de series, es conocido como *co-integración* en la literatura econométrica y de series de tiempo.

- Otra manera de definir co-integración descansa sobre transformaciones lineales de las series no estacionarias de raíces unitarias. En el ejemplo, la transformación realizada muestra que $y_{2t}=0.5\, x_{1t}+ x_{2t}$ no posee una raíz unitaria y es estacionaria. Por tanto, x_{1t} y x_{2t} son co-integradas si:

  - ✓ Las dos series son no estacionarias de raíces unitarias, y
  - ✓ Existe una combinación lineal de ellas que es estacionaria.

- En general, se dice que una serie de tiempo k-dimensional no estacionaria de raíz unitaria está cointegrada si hay **menos** de k raíces unitarias en el sistema. Sea h el número de raíces unitarias en la serie de tiempo k-dimensional. Entonces,

    - ✓ La cointegración existe si 0<h<k.
    - ✓ k-h es llamado el número de factores cointegrantes.
    - ✓ Alternativamente, k-h es el número de combinaciones lineales diferentes que son estacionarias de raíz unitaria.
    - ✓ Los vectores con los coeficientes de dichas combinaciones lineales son llamados *vectores cointegrantes.*

- En el ejemplo anterior, $y_{2t} = (0.5, 1.0)\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$ es estacionaria y por tanto $\begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix}$ es un vector cointegrante para el sistema. Para más detalles sobre la cointegración y las pruebas de cointegración, vea Box y Tiao (1977), Engle y Granger (1987), Stock y Watson (1988) y Johansen (1988), Wei (2006).

- El concepto de cointegración ha atraído la atención en la literatura y es muy relevante en estudios económicos y financieros. En general las series económicas y financieras son series integradas no estacionarias y pueden variar ampliamente. Sin embargo, las fuerzas económicas y de mercado tienden a mantener juntas muchas de esas series integradas formando relaciones de equilibrio. Algunos ejemplos son las tasas de interés de corto y largo plazo, el ingreso y el consumo.

Un modelo de Corrección de Error

- Debido a que hay más componentes no estacionarias con raíz unitaria, que el número de raíces unitarias en un sistema cointegrado, la diferenciación de cada una de las componentes individualmente para obtener estacionaridad, produce sobrediferenciación.

- Esta sobrediferenciación produce raíces unitarias en la matriz polinomial MA lo que puede generar dificultades en el problema de estimación. Cuando existen raíces unitarias en la matriz polinomial MA se dice que el vector de series de tiempo **no es invertible**.

  ✓ Para el modelo del ejemplo anterior, donde las series x_{1t} y x_{2t} aparecen diferenciadas, es decir,

$$\Delta x_t = \begin{bmatrix} 1-B & 0 \\ 0 & 1-B \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.7B & -0.6B \\ -0.15B & 1-0.7B \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

  los ceros del determinante de la matriz polinomial MA son 2.5 y 1, y por tanto hay una raíz unitaria y el entonces vector de series de tiempo no es invertible.

- Engle y Granger (1987) proponen una representación de corrección de error para un sistema cointegrado que resuelve la estimación de un modelo VARMA no invertible.
  ✓ Considere el sistema cointegrado anterior dado por

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \begin{bmatrix} 0.5 & -1.0 \\ -0.25 & 0.5 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

  y sea $\Delta x_t = x_t - x_{t-1}$, la serie bivariada diferenciada. Restando x_{t-1} a ambos lados de la ecuación se obtiene

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} -0.5 & -1.0 \\ -0.25 & -0.5 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} -1.0 \\ -0.5 \end{bmatrix} [0.5,\ 1.0] \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

Este es un modelo estacionario puesto que tanto Δx_t como $[0.5, 1.0]\, x_t = y_{2t}$ son estacionarios.

- ✓ En este último modelo, la matriz polinomial MA es la misma que antes y por tanto el modelo no tiene el problema de no-invertibilidad. Esta formulación es llamada **modelo de corrección de error para** Δx_t y puede ser extendida para un modelo general VARMA cointegrado.

- ✓ Para un modelo VARMA(p,q) cointegrado con m factores cointegrantes (m<k), una representación de corrección de error es

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + a_t - \sum_{j=1}^{q} \Theta_j a_{t-j}$$

donde α y β son matrices de k x m de rango completo m, y

$$\Phi_i^* = -\sum_{i=j+1}^{p} \Phi_i$$

$$\alpha\beta' = \Phi_p + \Phi_{p-1} + \cdots + \Phi_1 - I = -\Phi(1)$$

- ✓ La serie de tiempo m-variante $\beta' x_{t-1}$ es estacionaria de raíz unitaria y las columnas de β son los vectores cointegrantes de x_t.

- ✓ Observe que la única manera que Δx_t puede relacionarse con sentido con x_{t-1} es a través de la serie estacionaria $\beta' x_{t-1}$

Modelos VAR Cointegrados

- Suponga que una serie de tiempo k-dimensional x_t sigue un proceso VAR(p), con una posible tendencia temporal determinística. El modelo puede ser escrito como

$$x_t = \mu_t + \Phi_1 x_{t-1} + \cdots + \Phi_p x_{t-p} + a_t$$

donde a_t es una innovación Gaussiana y $\mu_t = \mu_0 + \mu_1 t$ es una componente determinística, donde μ_0 y μ_1 son vectores k-dimensionales de constantes.

  - ✓ El proceso VAR(p) es estacionario (es decir, es un proceso I(0)) si todos los ceros del determinante $|\Phi(B)|$ caen fuera del círculo unidad, donde $\Phi(B) = I - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p$.

  - ✓ Si $|\Phi(1)|$=0, entonces x_t es un proceso no estacionario de raíz unitaria. Suponga que x_t es a lo más un proceso integrado de orden 1, es decir, x_t es I(1). Entonces $(1-B)x_{it}$ es estacionario de raíz unitaria si x_{it} no lo es.

- Un modelo de corrección de error (ECM) para el proceso VAR(p) es

$$\Delta x_t = \mu_t + \Pi x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + a_t$$

donde las Φ_i^* fueron definidas antes y $\Pi = \alpha\beta' = -\Phi(1)$. El término Πx_{t-1} es llamado *término de corrección de error* y juega un papel clave en el análisis de cointegración.

- Observe que Φ_i puede ser recuperada desde la representación ECM a través de

$$\Phi_1 = I + \Pi + \Phi_1^*$$

$$\Phi_i = \Phi_i^* - \Phi_{i-1}^*,\ i=2,\ldots,p$$

donde $\Phi_p^* = 0$, la matriz nula.

- Si x_t tiene raíces unitarias, entonces $|\Phi(1)|=0$, y entonces $\Pi = -\Phi(1)$ es singular. Por tanto hay tres casos de interés al considerar el ECM:

1. El Rango(Π)=0. Esto implica que $\Pi=\mathbf{0}$ y que x_t no está cointegrada. El ECM se reduce a

$$\Delta x_t = \mu_t + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + a_t$$

En este caso Δx_t sigue un VAR(p-1) con tendencia determinística μ_t.

2. El Rango(Π)=k. Esto implica que $|\Phi(1)| \neq \mathbf{0}$ y que x_t no contiene raíces unitarias, es decir x_t es I(0). En este caso se estudia a x_t directamente.

3. El Rango(Π)=m<k. En este caso Π se puede escribir como

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

donde α y β son matrices de k x m de rango completo m. El ECM es

$$\Delta x_t = \mu_t + \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta x_{t-i} + a_t$$

Esto quiere decir que x_t es un proceso cointegrado con m vectores cointegrantes linealmente independientes, $w_t = \beta' x_t$ y hay k-m raíces unitarias que dan k-m tendencias aleatorias comunes en x_t.

- Si x_t es cointegrada con Rango(Π)=m, una manera simple de obtener una presentación de las k-m tendencias comunes es obtener una matriz $\alpha_\perp$, un complemento ortogonal de la matriz α; es decir $\alpha_\perp$ es una matriz de k x (k-m) tal que $\alpha_\perp^{'}\alpha$=**0** y use $y_t = \alpha_\perp^{'} x_t$ para representar las k-m tendencias aleatorias comunes .

  ✓ En efecto, si pre-multiplicamos al ECM

$$\Delta x_t = \mu_t + \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} \Phi_i^{*} \Delta x_{t-i} + a_t$$

  por $\alpha_\perp^{'}$ no habrá término de corrección de error en la ecuación resultante. Por tanto, la (k-m)-dimensional serie resultante, y_t tendrá k-m raíces unitarias.

  Por ejemplo, en el modelo anterior donde

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} -1.0 \\ -0.5 \end{bmatrix} [0.5,\ 1.0] \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

  Para este caso, $\alpha = [-1,\ -0.5]$ y $\alpha_\perp = [1,\ -2]'$. Por tanto $y_t = [1,\ -2] x_t = x_{1t} - 2x_{2t}$. Observe que esta es la serie no estacionaria de raíz unitaria y_{1t} vista anteriormente.

- La factorización $\Pi = \alpha\beta'$ no es única: para cualquier matriz de m x m ortogonal Ω, se obtiene

$$\alpha\beta' = \alpha\Omega\Omega'\beta' = \alpha\Omega(\alpha\Omega)' = \alpha_* \beta_*'$$

donde α_* y β_* son también de rango m.

- ✓ Se necesitan restricciones adicionales que permitan identificar de manera única a α y β.

- ✓ Es común requerir que $\beta' = [I_m, \beta_1']$ donde I_m es la matriz identidad de m x m, y β_1 es una matriz de (k-m) x m. En la práctica esto puede necesitar del reordenamiento del vector x_t, tal que las primeras m componentes todas tengan una raíz unitaria.

- ✓ Los elementos de α y β deben satisfacer también otras restricciones del proceso $w_t = \beta' x_t$. Por ejemplo, considere el caso de un modelo VAR(1) bivariado con un vector cointegrante. En este caso k=2 y m=1, y el EMC es

$$\Delta x_t = \mu_t + \begin{bmatrix} \alpha_1 \\ \alpha_1 \end{bmatrix} [1,\ \beta_1] x_{t-1} + a_t$$

  Pre-multiplicando la ecuación anterior por β', usando $w_{t-i} = \beta' x_{t-i}$ y moviendo w_{t-1} al lado derecho de la ecuación, se obtiene,

$$w_t = \beta' \mu_t + (1 + \alpha_1 + \alpha_2 \beta_1) w_{t-1} + b_t$$

  donde $b_t = \beta' a_t$. Esto implica que w_t es un proceso AR(1) estacionario. Por tanto α_1 y β_1 deben satisfacer la restricción de estacionaridad $|1 + \alpha_1 + \alpha_2 \beta_1| < 1$.

- El rango de la matriz Π en el ECM es el número de vectores cointegrantes. Entonces, para probar la existencia de cointegración se puede examinar el rango de Π.

Especificación de la función determinística

- Como en el caso univariante de raíces unitarias, las distribuciones límites de las pruebas de cointegración dependen de la función determinística μ_t. A continuación se discuten algunas de las especificaciones de μ_t que han sido propuestas en la literatura:

  - ✓ μ_t=**0**. En este caso todas las series componentes de x_t son I(1) sin deriva y la serie estacionaria $w_t = \beta' x_t$ tiene media cero.

  - ✓ $\mu_t = \mu_0 = \alpha c_0$, donde c_0 es un vector m–dimensional constante y no nulo. En este caso el ECM es

$$\Delta x_t = \alpha(\beta' x_{t-1} + c_0) + \Phi_1^* \Delta x_{t-1} + \Phi_2^* \Delta x_{t-2} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + a_t$$

    Todas las componentes de x_t son I(1) sin deriva, pero w_t tiene media no nula $-c_0$. Este es el caso de constante restringida.

  - ✓ $\mu_t = \mu_0$, no nulo. En este caso todas las series componentes de x_t son I(1) con deriva μ_0 y w_t puede tener media no nula.

  - ✓ $\mu_t = \mu_0 + \alpha c_1 t$ donde c_1 es un vector no nulo. El modelo ECM es

$$\Delta x_t = \mu_0 + \alpha(\beta' x_{t-1} + c_1 t) + \Phi_1^* \Delta x_{t-1} + \Phi_2^* \Delta x_{t-2} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + a_t$$

    de manera que las series componentes de x_t son I(1) con deriva μ_0 y w_t tiene una tendencia lineal. Este es el caso de tendencia restringida.

- ✓ $\mu_t = \mu_0 + \mu_1 t$ donde los μ_i son no nulos. En este caso tanto la constante como la tendencia no tienen restricciones. Las series componentes de x_t son I(1) y tienen una tendencia cuadrática, y w_t tiene una tendencia lineal.

- El último caso es poco común en trabajo aplicado. El primer caso no es común en series de tiempo económicas, pero puede servir para representar las series del log del precio de algunos activos financieros. El tercer caso también es útil para modelar las series del log del precio de algunos activos financieros.

Estimación Máximo Verosímil

- A continuación se describirá la EMV para un modelo VAR(p) cointegrado. Suponga que los datos son $\{x_t \mid T = 1,2,...,T\}$. Sin pérdida de generalidad, sea $\mu_t = \mu d_t$ donde $d_t = [1, t]'$, donde μ_t sigue alguna de las especificaciones anteriores. Si m es el rango de Π, el modelo ECM es

$$\Delta x_t = \mu d_t + \alpha\beta' x_{t-1} + \Phi_1^* \Delta x_{t-1} + \Phi_2^* \Delta x_{t-2} + \cdots + \Phi_{p-1}^* \Delta x_{t-p+1} + a_t$$

donde t=p+1,…,T.

- ✓ Un paso clave en la EMV es concentrar la función de verosimilitud con respecto a las componentes determinísticas y a los efectos estacionarios. Esto se logra considerando las siguientes dos regresiones multivariadas:

$$\Delta x_t = \gamma_0 d_t + \Omega_1 \Delta x_{t-1} + \Omega_2 \Delta x_{t-2} + \cdots + \Omega_{p-1} \Delta x_{t-p+1} + u_t$$

$$x_{t-1} = \gamma_1 d_t + \Xi_1 \Delta x_{t-1} + \Xi_2 \Delta x_{t-2} + \cdots + \Xi_{p-1} \Delta x_{t-p+1} + v_t$$

Sean $\hat{u}_t$ y $\hat{v}_t$, respectivamente, los residuales de las dos ecuaciones anteriores.

- ✓ Defina las matrices de covarianzas muestrales

$$S_{00} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t\hat{u}_t^{'}, \qquad S_{01} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t\hat{v}_t^{'} \text{ y } S_{11} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{v}_t\hat{v}_t^{'}$$

y calcule los valores propios de $S_{10}S_{00}^{-1}S_{01}$ con respecto a S_{11}. Esto implica resolver el problema de valores propios

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$$

- ✓ Denote los pares formados por el valor propio y su respectivo vector propio como $(\hat{\lambda}_i, e_i)$, donde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_k$, y donde los vectores propios e_i están normalizados de forma tal que $e' S_{11} e = I$, siendo $e = [e_1, e_2, \cdots, e_k]$ la matriz de vectores propios.

- ✓ El EMV no normalizado de β es $\hat{\beta} = [e_1, e_2, \cdots, e_m]$. Usando $\hat{\beta}$ se puede obtener un EMV para β que satisfaga las restricciones de identificación y la condición de normalización. Denote por $\hat{\beta}_c$ el estimador resultante. El EMV de los otros parámetros se obtiene de la ecuación de regresión lineal multivariada

$$\Delta x_t = \mu d_t + \alpha\hat{\beta}_c^{'} x_{t-1} + \Phi_1^{*}\Delta x_{t-1} + \Phi_2^{*}\Delta x_{t-2} + \cdots + \Phi_{p-1}^{*}\Delta x_{t-p+1} + a_t$$

- ✓ El valor de la función de verosimilitud maximizada basada en m vectores cointegrantes es

$$L_{máx}^{-2/T} \propto |S_{00}| \prod_{i=1}^{m}(1-\hat{\lambda}_i)$$

Este valor puede ser usado para una prueba del cociente de verosimilitud para contrastar la hipótesis Rango(Π)=m

✓ Estimaciones de los complementos ortogonales de α y β se pueden obtener usando

$$\hat{\alpha}_\perp = S_{00}^{-1}S_{11}[e_{m+1},\cdots,e_k], \qquad \hat{\beta}_\perp = S_{11}[e_{m+1},\cdots,e_k]$$

Una Prueba de Cointegración

- Sea H(m) la hipótesis nula de que el rango de Π es m. Por ejemplo, bajo H(0), el Rango(Π)=0, y entonces Π=**0** y no hay cointegración. Matemáticamente el rango de Π es el número de valores propios no nulos de Π. Este número se puede obtener si existe un estimador consistente para Π.

- Basados en el modelo ECM

$$\Delta x_t = \mu d_t + \Pi x_{t-1} + \Phi_1^*\Delta x_{t-1} + \Phi_2^*\Delta x_{t-2} + \cdots + \Phi_{p-1}^*\Delta x_{t-p+1} + a_t$$

la cual es una regresión lineal multivariada, se observa que Π está relacionada con la matriz de covarianzas entre x_{t-1} y Δx_t después de ajustar por los efectos de d_t y Δx_{t-i}, para i= 1,..,p-1. Claramente, las series x_{t-1} y Δx_t ajustadas son $\hat{u}_t$ y $\hat{v}_t$, respectivamente.

- La ecuación de interés para la prueba de cointegración es

$$\hat{u}_t = \Pi\hat{v}_t + a_t$$

Bajo el supuesto de normalidad, la prueba del cociente de verosimilitud para probar el rango de Π se puede realizar usando análisis de correlación canónica entre $\hat{u}_t$ y $\hat{v}_t$ (Vea, por ejemplo, Johnson y Wichern, 1998). Las correlaciones canónicas asociadas son las *correlaciones canónicas parciales* entre Δx_t y x_{t-1} debido a que se han ajustado los efectos de d_t y Δx_{t-i}. Los valores propios $\hat{\lambda}_i$ son los cuadrados de las correlaciones canónicas entre $\hat{u}_t$ y $\hat{v}_t$.

- Considere la hipótesis H$_0$: Rango(Π)=m contra H$_a$: Rango(Π)>m. Para realizar la prueba, Johansen (1988) propuso el estadístico del cociente de verosimilitud,

$$LK_{tr}(m) = -(T-p)\sum_{i=m+1}^{k}\ln(1-\hat{\lambda}_i)$$

Si el rango de Π es m, entonces $\hat{\lambda}_i$ debería ser pequeño para i>m y por tanto $LK_{tr}(m)$ sería pequeño. Esta prueba es denominada la *prueba de cointegración de la traza*.

- Debido a la presencia de raíces unitarias, la distribución asintótica de $LK_{tr}(m)$ no es chi-cuadrada, sino una función de movimientos Brownianos estándar. Por tanto, los valores críticos de $LK_{tr}(m)$ pueden ser obtenidos por medio de simulación.

- Johansen (1988) también propuso un procedimiento secuencial para determinar el número de vectores cointegrantes. Las hipótesis de interés son

H$_0$: Rango(Π)=m contra H$_a$: Rango(Π)=m+1

El estadístico del cociente *LK*, llamado el *estadístico del máximo valor propio* está definido como

$$LK_{máx}(m) = -(T-p)\ln(1-\hat{\lambda}_{m+1})$$

Los valores críticos del estadístico son no estándar y se obtienen por medio de simulación.

Pronósticos de Modelos VAR Cointegrados

- El modelo ECM ajustado puede ser usado para producir pronósticos.

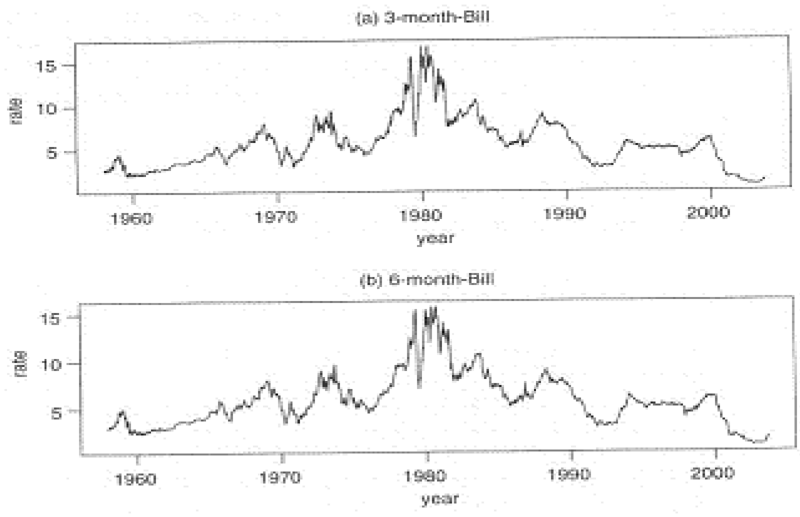
  ✓ Primero, condicionado sobre los parámetros estimados, el modelo ECM puede ser usado para producir pronósticos de la serie diferenciada Δx_t.

  ✓ Estos pronósticos pueden ser usados para obtener los pronósticos de la serie x_t.

- Una diferencia entre los pronósticos del modelo ECM y los del modelo VAR es que el modelo ECM impone las relaciones cointegrantes para obtener los pronósticos.

Un ejemplo

Considere las series de las tasas a 3 y a 6 meses de los Bonos del Tesoro (TB), desde Diciembre 12 de 1958 hasta Agosto 6 de 2004, para un total de 2383 observaciones. A continuación se presentan sus gráficos.



- ✓ La prueba de raíces unitarias de Dickey-Fuller aumentada (ADF) no rechaza la hipótesis nula de una raíz unitaria: usando un modelo AR(3) los estadísticos de la prueba son -2.34 y -2.33 para un valor p aproximado de 0.16. Se puede proceder a la modelación del VAR.

- ✓ Para la serie bivariada formada por las 2 tasas anteriores, r_t=[tb3m, tb6m]', el criterio BIC selecciona un modelo VAR(3).

- ✓ Para realizar la prueba de cointegración, elegimos una constante restringida para μ_t, debido a que a priori no hay razón para creer en la existencia de una deriva en la tasa de interés de E.U. Las dos pruebas de Johansen confirman que las dos series están cointegradas, con un vector cointegrante, cuando se usa un VAR(3). El modelo ECM a estimar es de la forma:

$$\Delta x_t = \alpha(\beta' x_{t-1} + c_0) + \Phi_1^* \Delta x_{t-1} + \Phi_2^* \Delta x_{t-2} + a_t$$

Resultados del S-Plus

```
Pruebas para el Rango de Cointegración

   Valor propio Estadíst. Traza    95%CV   99%CV
H(0)   0.0322       83.2712         19.96    24.60
H(1)   0.0023       5.4936           9.24    12.97
```

	Estadíst. Máxim	95%CV	99%CV
H(0)	77.7776	15.67	20.20
H(1)	5.4936	9.24	12.97

- ✓ Estimación máximo verosímil del modelo VAR(3) cointegrado usando la presentación del ECM.

Vectores Cointegrantes:

	Coint.1
	1.0000
tb6m	-1.0124
(err.std)	0.0086
(t.stat)	-118.2799
Intercep	0.2254
(err.std)	0.0545
(t.stat)	4.1382

Coeficientes del VECM

	tb3m	tb6m
Coint.1	-0.0949	-0.0211
(err.std)	0.0199	0.0179
(t.stat)	-4.7590	-1.1775
tbm3t.lag1	0.0466	-0.0419
(err.std)	0.0480	0.0432
(t.stat)	0.9696	-0.9699
tbm6t.lag1	0.2650	0.3164
(err.std)	0.0538	0.0484
(t.stat)	4.9263	6.5385
tbm3t.lag2	-0.2067	-0.0346
(err.std)	0.0481	0.0433
(t.stat)	-4.2984	-0.8005
tbm6t.lag2	0.2547	0.0994
(err.std)	0.0543	0.0488
(t.stat)	4.6936	2.0356

Diagnósticos :

	tbm3	tbm6
R^2	0.1081	0.0913
R^2-ajust	0.1066	0.0898
Escala resid.	0.2009	0.1807

- ✓ Los resultados muestran que:

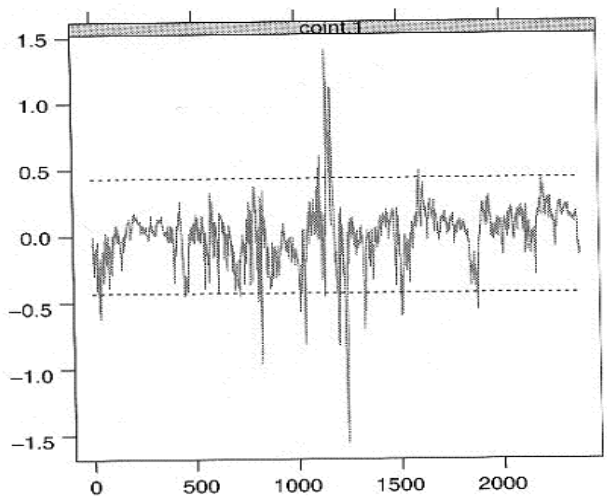
⇒ La serie estacionaria es $w_t \approx tb3m_t - tb6m_t$ con una media de -0.225 aproximadamente.

⇒ El modelo ECM ajustado es

$$\Delta x_t = \begin{bmatrix} -0.09 \\ -0.02 \end{bmatrix}(w_{t-1} + 0.23) + \begin{bmatrix} 0.05 & 0.27 \\ -0.04 & 0.32 \end{bmatrix}\Delta x_{t-1} + \begin{bmatrix} -0.21 & 0.25 \\ -0.03 & 0.10 \end{bmatrix}\Delta x_{t-2} + a_t$$

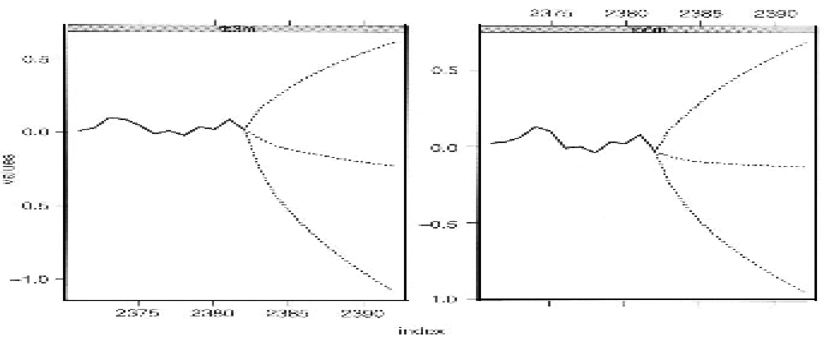
⇒ Los errores estándar de a_{1t} y a_{2t} son 0.2009 y 0.1807, respectivamente.

⇒ El ajuste del modelo ECM estimado se puede examinar usando varios gráficos, como el siguiente gráfico de residuales cointegrantes.

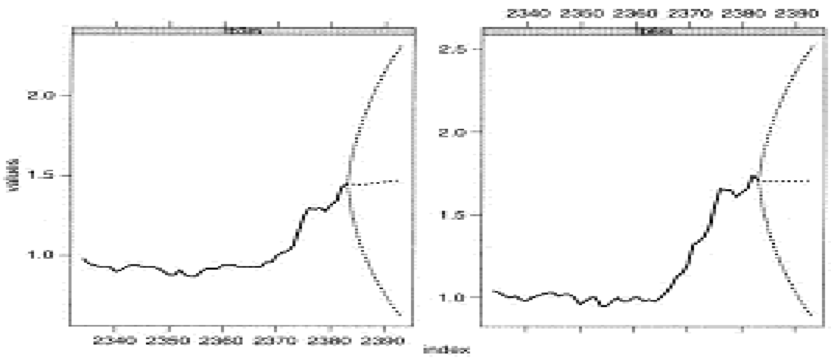


✓ El modelo ECM ajustado puede ser empleado para producir pronósticos tanto para Δx_t como para x_t. Usando como origen Agosto 6 de 2004, los siguientes gráficos presentan los pronósticos hasta 10 períodos adelante para Δx_t y x_t, junto con los intervalos de predicción al 95% de confianza.

Pronósticos para Δx_t



Pronósticos para x_t



$\Rightarrow$ Debido a la no estacionaridad de raíz unitaria, los intervalos son amplios y no son informativos.

Comandos para el S-Plus

```
# Obtención de una matriz con las series tb6m y tb6m como columnas
x=cbind(tb3m, tb6m)
# Obtención del data frame
y=data.frame(x)
# Selección del orden p del VAR
ord.choice=VAR(y, max.ar=12)
ord.choice$info

# Prueba de Cointegración de Johansen con VECM de orden p-1
cointst.rc=coint(x, trend='rc', lags=2)
cointst.rc

# Estimación Máxima Verosímil del VAR(3) usando la representación VECM
vecm.fit=VECM(cointst.rc)
summary(vecm.fit)

# Diagnósticos gráficos
plot(vecm.fit)
```

```
# Pronósticos del modelo VECM
vecm.fst=predict(vecm.fit, n.predict=10)
summary(vecm.fst)
```