

CUADERNOS DE MATEMÁTICA
DE LA ESCUELA POLITÉCNICA NACIONAL

HOLGER CAPA SANTOS

SERIES TEMPORALES

La ciencia y el arte de los pronósticos



Cuaderno de Matemática No. 10

SERIES TEMPORALES: LA CIENCIA Y EL ARTE DE LOS PRONÓSTICOS

HOLGER CAPA SANTOS

Responsable de la Edición: ?

Revisión técnica: ?

Asistentes: Andrés Merino

Portada: Andrés Merino

Registro de derecho autoral No.

ISBN:

Publicado por la Unidad de Publicaciones de la Facultad de Ciencias de la Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito, Ecuador.

Primera edición: 2015

Primera impresión: 2015

© Escuela Politécnica Nacional 2015

Índice general

1	Modelos de Heteroscedasticidad Condicional	3
1.1	Modelos Arch–Garch Simétricos	3
1.2	Modelos GARCH asimétricos	5
1.3	Metodología de la Modelación	7
1.4	Ejemplos con Heteroscedasticidad Condicional	8
2	Modelos Multivariantes de Series Temporales	11
2.1	Procesos Estacionarios	11
2.1.1	Proceso estrictamente estacionario	11
2.1.2	Proceso débilmente estacionario	11
2.2	Matrices de Correlación Cruzada (Cross-Correlation)	12
2.2.1	Dependencia Lineal	13
2.3	Modelos de Vectores Autoregresivos (VAR)	17
2.3.1	El caso bivalente	17
2.3.2	Representación general de un VAR	19
2.4	Representación VARMA de una Serie Multivariante	21
2.4.1	La representación VMA	21
2.4.2	Representación lineal de un VAR(p)	21
2.4.3	La representación VARMA	23
2.5	Formulación de un modelo VAR	24
2.5.1	Estimación	24
2.5.2	Determinación del número de retardos	25
2.5.3	Diagnóstico y validación del modelo	26
2.6	Predicción	29
2.7	La Causalidad	40
2.7.1	La causalidad según Granger	41
2.7.2	La causalidad según Sims	42
2.8	Análisis de los “choques”	44
2.9	Descomposición de la varianza	52
2.9.1	Representación de errores ortogonales	52
2.9.2	Descomposición de la Varianza	52
2.9.3	Elección del orden de descomposición	53
2.10	Ejemplo Práctico	57

3	La Cointegración	73
3.1	Propiedades del orden de Integración de una serie	73
3.2	Condiciones de cointegración	75
4	Modelo de Corrección del error (MCE)	79
4.1	Planteamiento del MCE	79
4.2	Prueba de Johansen	81
	Índice Alfabético	85

Índice de figuras

1.1	FAC y FACP estimadas residuales del Modelo 3 para la SVM	8
1.2	FAC y FACP estimadas de los residuos cuadráticos del Modelo 3 para la SVM	8
1.3	FAC y FACP estimadas residuales del Modelo 3-ARCH(1) para la SVM	9
1.4	FAC y FACP estimadas de los residuos cuadráticos del Modelo 3-ARCH(1) para la SVM	9
2.1	Correlaciones cruzadas entre X_{1t} y X_{2t}	16
2.2	Correlaciones cruzadas entre X_{1t} y X_{3t}	16
2.3	Correlaciones cruzadas entre X_{2t} y X_{3t}	17
2.4	Gráfico de las variaciones trimestrales de las series PIB, CI y DFI	31
2.5	Especificación de un modelo VAR en EViews	33
2.6	Criterio de estabilidad para el VAR(1) estimado	34
2.7	Comparación de las predicciones VAR y UNIVARIANTE para Y_{1t} (PIB)	39
2.8	Comparación de las predicciones VAR y UNIVARIANTE para Y_{2t} (CI)	39
2.9	Comparación de las predicciones VAR y UNIVARIANTE para Y_{3t} (DFI)	40
2.10	Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{1t}	47
2.11	Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{2t}	49
2.12	Correlaciones cruzadas entre X_{1t} y X_{2t}	58
2.13	Gráfico de secuencia de las series	59
2.14	Criterio de estabilidad para el VAR(1) estimado	61
2.15	Comparación de las predicciones VAR y UNIVARIANTE para X_{1t} (IPP)	66
2.16	Comparación de las predicciones VAR y UNIVARIANTE para X_{2t} (IAE)	67
2.17	Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{1t}	69
2.18	Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{2t}	70
3.1	Las variables X_{1t} y X_{2t} no están cointegradas	74
3.2	Las variables X_{1t} y X_{2t} están cointegradas	74
3.3	Series consideradas en el ejemplo	76
4.1	Series consideradas en el ejemplo	83
4.2	Prueba de cointegración de Johansen	84

Índice de tablas

1.1	Información sobre los coeficientes del Modelo 3 para la SVM	8
1.2	Información estadística para el Modelo 3-ARCH(1) para la SVM	9
2.1	Resumen de estadísticas y matrices de correlación cruzada para X_{1t} , X_{2t} y X_{3t}	15
2.2	Prueba DFA para Y_{1t}	31
2.3	Prueba DFA para Y_{2t}	31
2.4	Prueba DFA para Y_{3t}	32
2.5	Criterios para escoger el retardo del VAR	32
2.6	Prueba de autocorrelación Pormanteau	35
2.7	Prueba LM	36
2.8	Prueba de Normalidad de los residuos	36
2.9	Error Medio Cuadrático estimado para los modelos VAR y univariante	40
2.10	Prueba de causalidad de Granger	44
2.11	Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{1t}	46
2.12	Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{2t}	48
2.13	Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{3t}	50
2.14	Descomposición hipotética de choques	54
2.15	Otra descomposición hipotética de choques	55
2.16	Descomposición de la Varianza (Orden de Cholesky Y_3 Y_2 Y_1)	56
2.17	Resumen de estadísticas y matrices de correlación cruzada para X_{1t} y X_{2t} . . .	58
2.18	Prueba DFA para X_{1t}	59
2.19	Prueba DFA para X_{2t}	59
2.20	Criterios para escoger el retardo del VAR	60
2.21	Prueba de autocorrelación Pormanteau	62
2.22	Prueba LM	63
2.23	Error Medio Cuadrático estimado para los modelos VAR y UNIVARIANTE .	67
2.24	Prueba de causalidad de Granger	68
2.25	Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{1t}	68
2.26	Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{2t}	69
2.27	Descomposición de la Varianza (Orden de Cholesky X_1 X_2)	71
3.1	Prueba de raíces unitarias para las X_{1t}	77
3.2	Prueba de raíces unitarias para las X_{2t}	77
3.3	Prueba de raíces unitarias para las X_{3t}	77
3.4	Prueba de raíces unitarias para X_{1t} en primera diferencia	77
3.5	Prueba de raíces unitarias para X_{2t} en primera diferencia	77

3.6	Prueba de raíces unitarias para X_{3t} en primera diferencia	78
4.1	Elección del retardo del VAR	83

Notaciones

- $E(.)$: Esperanza matemática de una variable aleatoria.
- $V(.)$: Varianza de una variable aleatoria.
- $\text{Cov}(.,.)$: Covarianza de dos variables aleatorias.
- $N(\mu, \sigma^2)$: Distribución normal con media μ y varianza σ^2 .
- $\hat{\cdot}$: Denota un estimador del parámetro estadístico que se encuentra bajo el símbolo.
- c : De manera general, representa a una constante numérica.
- i.i.d : Representa a la frase: independientes e idénticamente distribuidas.
- i.i.d (0,1) : Variables aleatorias independientes e idénticamente distribuidas con esperanza matemática 0 y varianza 1.
- $\gamma(l)$: Función de autocovarianza de orden l .
- $\rho(l)$: Función de autocorrelación de orden l .
- $r(l)$: Función de autocorrelación parcial de orden l .
- 1_A : Función indicatriz sobre el conjunto A .
- $L_c^2(\Omega)$: Espacio de variables aleatorias de Ω en C , de cuadrado integrable.
- $L_c^2(E)$: Espacio de funciones medibles de E en C , de cuadrado integrable.
- Σ : Matriz de varianzas - covarianzas
- Σ_u : Matriz de varianzas – covarianzas residuales
- \therefore : Significa: por tanto

Capítulo 1

Modelos de Heteroscedasticidad Condicional

1.1 Modelos Arch–Garch Simétricos

En esta sección, se presentará un breve resumen de la teoría desarrollada alrededor de los modelos ARCH – GARCH, los cuales determinan un patrón de comportamiento estadístico para la varianza condicional, denominados modelos Auto-Regresivos con Heteroscedasticidad Condicional.

El análisis de series temporales económicas, tradicionalmente se ha centrado en el estudio de modelos para la media condicional en los que se asume que la varianza condicional es constante; en este caso se dice que existe homoscedasticidad. Si este no fuera el caso, se estaría enfrentado un problema de no estacionariedad de la serie.

En 1982, Robert Engle revolucionó los modelos de volatilidad introduciendo el estudio de las estructuras cuadráticas, ampliando así la visión de la metodología Box–Jenkins, en la cual los modelos lineales de tipo ARIMA admiten que las innovaciones son un ruido blanco, con media cero y varianza constante.

Los modelos de volatilidad condicional son importantes por el papel que juega el riesgo y el concepto de incertidumbre en el desarrollo de las teorías modernas de modelos financieros, que relacionan de forma directa el riesgo con la volatilidad. Estos modelos permiten relacionar el valor de la varianza condicional (no constante) en función del conjunto de información disponible en periodos anteriores; están específicamente diseñados para modelar y pronosticar varianzas condicionales.

Definición 1.1. Un Modelo ARCH(r) se define por:

$$\begin{aligned} Z_t &= \sqrt{h_t} u_t \\ h_t &= \alpha_0 + \alpha_1 Z_{t-1}^2 + \dots + \alpha_r Z_{t-r}^2 \end{aligned}$$

donde los (u_t) son independientes e idénticamente distribuidos con media cero y varianza uno (i.i.d. $(0,1)$), $\alpha_0 > 0$ y $\alpha_i \geq 0$ para $i > 0$.

En la práctica se supone que $u_t \sim N(0,1)$ o que siguen una distribución t-student. Los coeficientes α_i deben satisfacer ciertas condiciones (en general no negatividad: $\alpha_0 > 0$ y

$\alpha_i \geq 0$ para $i > 0$ y $\sum_{i=1}^r \alpha_i < 1$), dependiendo del tipo de restricciones que se coloquen sobre el proceso Z_t .

Las restricciones de signo de los coeficientes de la ecuación de varianza garantizan que la varianza condicional sea positiva en todos los períodos. Lo que se necesita es que, una vez que se haya estimado el modelo, genere una serie de varianzas positiva, lo que puede suceder aún si algunos de los coeficientes α_i fueran negativos. Esto se puede considerar como un contraste de validez del modelo.

Por la propia definición, a valores grandes de Z_t les siguen otros valores grandes de la serie. h_t es la varianza condicional de Z_t dado Z_s , $s < t$.

Observación 1.1. Se puede demostrar que un proceso $ARCH(r)$ implica que se puede representar X_t^2 como un proceso $AR(r)$, con residuos que no son gaussianos. Además, el coeficiente de apuntamiento (curtosis) es mayor que 3, por lo cual las colas de la distribución serán más pesadas que en la distribución normal.

Definición 1.2. Un modelo $GARCH(r, s)$, está definido por:

$$Z_t = \sqrt{h_t} u_t$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i Z_{t-i}^2 + \sum_{j=1}^s \beta_j h_t$$

donde los u_t son i.i.d. $(0, 1)$ con $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ para $i > 0$ y

$$\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1$$

Como en el caso de los modelos $ARCH$, usualmente se supone que los u_t son normales o siguen una distribución t -student.

Las restricciones de no negatividad impuestas sobre los coeficientes en la definición son para asegurarse que la varianza no llegue a tomar un valor negativo. Además, en algunos casos se pueden encontrar condiciones menos restrictivas para estos coeficientes, que aseguren que la varianza sea positiva.

Observación 1.2. Se puede demostrar que un proceso $GARCH(rs)$ implica que se puede representar Z_t^2 como un proceso $ARMA(rs)$, con residuos que no son gaussianos (más precisamente con residuos que representan una diferencia martingala). También, en este caso, el coeficiente de apuntamiento (curtosis) es mayor que 3, por lo cual las colas de la distribución serán más pesadas que en la distribución normal.

Modelos IGARCH

Este modelo fue descrito originalmente por Engle y Bollerslev (1986). Si el modelo polinomial AR del modelo $GARCH$ tiene una raíz unitaria, se tiene un modelo $IGARCH$. Los modelos $IGARCH$ son modelos $GARCH$ con una raíz unitaria. Estos modelos tienen la característica que los impactos de los choques al cuadrado sobre Z_t^2 son persistentes; esto se puede expresar por:

$$\eta_{t-i} = Z_{t-i}^2 - h_{t-i}, \quad \text{para } i > 0$$

Definición 1.3. Un modelo $IGARCH(r, s)$, se define por:

$$Z_t = \sqrt{h_t} u_t$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i Z_{t-i}^2 + \sum_{j=1}^s (1 - \beta_j) h_{t-j}$$

tal que,

$$\sum_{j=1}^s \beta_j + \sum_{i=1}^r \alpha_i = 1$$

Como en el caso de los modelos ARCH, usualmente se supone que los u_t son normales o siguen una distribución t -student.

Los modelos IGARCH son un caso específico dentro de la familia de los “modelos con varianza persistente” en los que la información actual (en el instante “ t ”) es importante para realizar predicciones óptimas con cualquier horizonte temporal.

1.2 Modelos GARCH asimétricos

Una característica de los modelos GARCH, dado que la varianza depende fundamentalmente de valores cuadráticos pasados, es que la volatilidad que generan frente a cambios positivos o negativos inesperados de la variable (piénsese en retornos de inversiones), dan una respuesta simétrica a estos. Sin embargo, se ha probado empíricamente que la reacción que tiene la volatilidad o varianza condicional de muchas variables financieras a este tipo de cambios es asimétrica; es decir, existe diferencia en la respuesta de la volatilidad de la variable, dependiendo de si el cambio es positivo o negativo.

Con el fin de modelar esta respuesta asimétrica se han desarrollado una variedad de modelos asimétricos; los más representativos son los modelos EGARCH, TARCH, PARCH.

Modelos EGARCH

Las restricciones de no negatividad para asegurar la positividad de la varianza en los modelos GARCH en muchas ocasiones son difíciles de lograr. Nelson (1991) propuso los modelos EGARCH (*Exponential GARCH*), como solución a este problema; éstos, además, incorporan efectos asimétricos.

Definición 1.4. Un modelo $EGARCH(r, s)$, se define por:

$$Z_t = \sqrt{h_t} u_t$$

$$\ln(h_t) = \alpha_0 + \sum_{j=1}^s \beta_j \ln(h_{t-j}) + \sum_{i=1}^r (\alpha_i |u_t| + \gamma_i u_t)$$

Nótese que el lado izquierdo de la ecuación es el logaritmo de la varianza condicional; esto implica que su efecto es exponencial y garantiza que las predicciones de ésta serán no negativas. Cuando $\gamma_i \neq 0$, el efecto asimétrico deber incorporarse al Modelo GARCH.

Obsérvese también que se puede escribir $\frac{Z_t}{\sqrt{h_t}}$ en lugar de u_t , en las fórmulas anteriores.

Como en el caso de los modelos ARCH, usualmente se supone que los u_t son normales o siguen una distribución t -student.

Modelos TARCh

Los modelos TARCh (*Threshold ARCH*) fueron introducidos independientemente por Zakoian (1990) y Glosten, Jaganathan y Runklen (1993), por lo que también se conocen como GJR–GARCH. Estos modelos incluyen una variable adicional d_t , que determina el carácter asimétrico del modelo.

Definición 1.5. Un modelo TARCh(r, s), se define por:

$$Z_t = \sqrt{h_t} u_t$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i Z_{t-i}^2 + \gamma Z_{t-1}^2 d_{t-1} + \sum_{j=1}^s \beta_j h_{t-j}$$

donde los u_t son i.i.d. $(0, 1)$ con $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ para $i, j > 0$ y $\gamma \neq 0$

$$d_t = \begin{cases} 1, & \text{si } Z_t < 0 \\ 0, & \text{si } Z_t \geq 0 \end{cases}$$

Si $\gamma = 0$ se pierde el efecto asimétrico del modelo.

En este modelo, las malas noticias ($Z_t < 0$) y las buenas noticias ($Z_t \geq 0$) (piénsese otra vez en retornos), tienen efectos diferentes sobre la varianza condicional.

Modelos PARCH

Los modelos PARCH (*Power ARCH*) desarrollados independientemente por Taylor (1986) y Schwert (1989), introducen la desviación estándar a los modelos ARCH; donde se modela la desviación estándar en lugar de la varianza. Este modelo fue generalizado por Ding y otros (1993).

En el modelo PARCH, el parámetro de potencia δ de la desviación estándar puede ser estimado antes que impuesto y los parámetros opcionales γ se agregan para capturar la asimetría dentro de los datos.

Definición 1.6. Un modelo PARCH(r, s), se define por:

$$Z_t = \sqrt{h_t} u_t$$

$$h_t^\delta = \alpha_0 + \sum_{i=1}^r \alpha_i (|u_{t-i}| - \gamma_i u_{t-i})^\delta + \sum_{j=1}^s \beta_j h_{t-j}^\delta$$

donde $\delta > 0$, es el parámetro del término de la potencia. γ_i se dicen los parámetros de apalancamiento.

En series de valores simétricos $\gamma_i = 0$ para todo i . Nótese que si $\delta = 1$ y $\gamma_i =$ para todo i , el modelo PARCH es simplemente una especificación GARCH estándar. Si los $\gamma_i = 0$ se pierde el efecto asimétrico del modelo.

Los modelos GARCH asimétricos, se estiman por el método de máxima verosimilitud condicional, por lo cual se requiere de ciertos supuestos acerca del comportamiento de los errores. Por lo general, se suponen i.i.d con distribución normal o incluso con una distribución *t*-student.

1.3 Metodología de la Modelación

El objetivo es encontrar un modelo que represente adecuadamente a los datos históricos de una determinada variable, combinando especificaciones tanto para la media como para la varianza condicional. Los tipos de modelos que se considerarán serán los ARIMA – GARCH, de tal manera que la media condicional de la serie sea descrita por un modelo del tipo ARIMA y su varianza condicional por uno de la familia de modelos ARCH – GARCH o de sus extensiones asimétricas PARCH, TARCH y EGARCH. La modelación se realizará utilizando el paquete EViews.

El primer paso es, por tanto, modelar la serie de datos por un modelo del tipo ARIMA o incluso SARIMA, con lo que se obtiene un modelo para la media condicional de la serie.

Luego de haberse eliminado toda correlación lineal en la serie, se debe indagar si existe heteroscedasticidad condicional residual, para lo cual deben analizarse los residuos estandarizados estimados al cuadrado; el correleograma correspondiente, permite llevar a cabo un análisis gráfico de identificación, para ver si algún valor es estadísticamente diferente de cero, y por tanto, existe autocorrelación en su forma residual cuadrática.

Si se verifica la existencia de heteroscedasticidad condicional en los residuos, se rechaza el supuesto de la varianza constante; se intentará entonces obtener una especificación para la varianza condicional, a través de la modelación de los residuos estimados obtenidos por el modelo ARIMA, mediante un modelo del tipo ARCH – GARCH o sus extensiones asimétricas.

Inicialmente se mantiene la estructura para la media condicional, obtenida por el modelo ARIMA, pero esta puede modificarse con la nueva especificación. Los residuos estimados deben analizarse, tanto en su forma simple como en la cuadrática, para eliminar toda evidencia de autocorrelación lineal (deben aceptarse como un ruido blanco).

La estimación y verificación permiten encontrar uno o varios modelos que cumplan las condiciones que se impusieron en la modelación ARIMA; es decir, todos los coeficientes deben ser significativos; las raíces de los polinomios característicos, tanto de la parte autorregresiva como de la media móvil, deben estar fuera del círculo unidad, para así asegurar la estacionariedad e invertibilidad del proceso. Además, los coeficientes de la ecuación de la varianza condicional deben satisfacer las restricciones de no negatividad para la varianza (modelos ARCH – GARCH).

Para la verificación de la presencia de una estructura ARIMA en los residuos (simples o cuadráticos) pueden utilizarse la FAC y la FACP; además, también se debe realizar la prueba global (estadístico *Q*) de Box – Pierce – Ljung.

Una vez que un modelo ha sido estimado y ha superado las diversas verificaciones, se convierte en un instrumento útil para las predicciones de valores futuros. Como en la modelación ARIMA, si varios modelos son plausibles, se elige entre estos al mejor, mediante los criterios ya citados previamente.

1.4 Ejemplos con Heteroscedasticidad Condicional

Aunque los datos de las ventas que se vienen utilizando no corresponden al ámbito financiero, sirven muy bien para ilustrar la modelación para la varianza condicional. En esta ocasión se adoptará el Modelo 3 con el cual se modeló la media condicional (SARIMA). La Figura 4.2 no permite aceptar la hipótesis de que la serie tenga una varianza constante.

Una posibilidad para amortiguar los efectos de varianza no constante es utilizar la transformación logaritmo o, en general, la transformación de Box y Cox; sin embargo, en esta ocasión se tratará de modelar directamente la varianza a través de los Modelos ARCH-GARCH o sus extensiones asimétricas.

En la Tabla 4.1 y en las figuras 4.1 y 4.2 se presentan la información estadística y residual para el Modelo 3 de la SVM:

Tabla 1.1: Información sobre los coeficientes del Modelo 3 para la SVM

Variable	Coefficient	Std. Error	<i>t</i> -Statistic	Prob.
C	156.1661	34.82541	4.484257	0.0000
AR(1)	0.325909	0.107494	3.031876	0.0033
AR(12)	-0.335945	0.114892	-2.924008	0.0045
MA(13)	0.480540	0.111556	4.307611	0.0000

Figura 1.1: FAC y FACP estimadas residuales del Modelo 3 para la SVM

Figura 1.2: FAC y FACP estimadas de los residuos cuadráticos del Modelo 3 para la SVM

Se observan fuertes correlaciones entre los residuos cuadráticos estandarizados estimados, por lo cual se hace necesaria la modelación de la varianza condicional del Modelo 3.

En general, es difícil establecer el orden para los modelos ARCH-GARCH. Lo usual es probar los modelos con parámetros (1,0), (1,1), (1,2) o (2,2). Para este caso se empezó probando con el modelo $ARCH(1)$; los resultados aparecen en la Tabla 4.2 y en las Figuras 4.3 y 4.4.

Tabla 1.2: nformación estadística para el Modelo 3-ARCH(1) para la SVM

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	115.3958	36.57140	3.155355	0.0016
AR(1)	0.483776	0.103019	4.695993	0.0000
AR(12)	-0.449722	0.078830	-5.704935	0.0000
MA(13)	0.798959	0.041313	19.33915	0.0000
Variance Equation				
C	16129.50	3530.889	4.568113	0.0000
RESID(-1) ²	0.610436	0.249248	2.449115	0.0143
R-squared	0.335639	Mean dependent var	157.1548	
Adjusted R-squared	0.310726	S.D. dependent var	258.2531	
S.E. of regression	214.4082	Akaike info criterion	13.23093	
Sum squared resid	3677671.	Schwarz criterion	13.40456	
Log likelihood	-549.6992	Hannan-Quinn criter.	13.30073	
Durbin-Watson stat	2.201103			
Inverted AR Roots	.96 - .24i	.96 + .24i	.71 + .65i	.71 - .65i
	.28 - .89i	.28 + .89i	-.21 + .90i	-.21 - .90i
	-.63 - .66i	-.63 + .66i	-.87 + .24i	-.87 - .24i
Inverted MA Roots	.95-.24i	.95+.24i	.74-.65i	.74+.65i
	.35 + .92i	.35 - .92i	-.12 - .98i	-.12 + .98i
	-.56 - .81i	-.56 + .81i	-.87 + .46i	-.87 - .46i
	.98			

Figura 1.3: FAC y FACP estimadas residuales del Modelo 3-ARCH(1) para la SVM**Figura 1.4:** FAC y FACP estimadas de los residuos cuadráticos del Modelo 3-ARCH(1) para la SVM

Las Figuras 4.3 y 4.4 evidencian que existen problemas ya no solo en los residuos cuadráticos, sino también en los residuos simples. En la figura 4.4, la FACP en el orden 13 es significativo (y cercano a la estacionalidad 12); por lo cual, se decidió incluir un término AR(13) en el Modelo 3; esto tampoco solucionó totalmente la falta de independencia de los residuos cuadráticos. Luego, de algunas pruebas se encontró como modelo final aquel que contiene términos c, SAR(12), MA(12) y AR(13) para la media (se lo llamará Modelo 4) y ARCH(1) para la varianza. Los resultados se muestran en la Tabla 4.3 y en las Figuras 4.5 y 4.6.

Capítulo 2

Modelos Multivariantes de Series Temporales

Una serie temporal multivariante es un proceso estocástico $(X_t)_{t \in \mathbb{Z}}$, con X_t un vector donde cada componente se define como una serie temporal univariante. En este documento se utilizará la notación de vectores como columnas. Así se denota:

$$X_t = (X_{1t}, \dots, X_{kt})' \text{ el vector de } k \text{ series univariantes en el instante } t$$

Lo importante de tratar series multivariantes es que, a más de considerar simultáneamente observaciones de dos o más series univariantes, también se puede analizar las correlaciones existentes entre ellas; esto evidentemente enriquece el análisis, aunque los procesos operativos serán más complejos que en el caso univariante.

2.1 Procesos Estacionarios

Para poder estimar las características de los procesos se necesita suponer que son estables a lo largo del tiempo; esto implica, que son estacionarios.

2.1.1 Proceso estrictamente estacionario

Un proceso estocástico multivariado $(X_t)_{t \in \mathbb{Z}}$, con $X_t = (X_{1t}, \dots, X_{kt})'$, es estrictamente estacionario (o fuertemente estacionario) si las distribuciones conjuntas de cualquier conjunto finito de variables se mantienen por saltos.

Es decir, si:

$$F_{t_1+l, \dots, t_k+l}(x_{t_1+l}, \dots, x_{t_k+l}) = F_{t_1, \dots, t_k}(x_{t_1}, \dots, x_{t_k})$$

Para todo $k \in \mathbb{N}$ y para todo $t_1, \dots, t_k, l \in \mathbb{Z}$

Donde, F_{t_1, \dots, t_k} denota la distribución conjunta de X_{t_1}, \dots, X_{t_k} .

2.1.2 Proceso débilmente estacionario

Un proceso estocástico multivariado $(X_t)_{t \in \mathbb{Z}}$, con $X_t = (X_{1t}, \dots, X_{kt})'$, se dice que es *débilmente estacionario* si sus momentos de primer y segundo orden son invariantes en el tiempo

(no dependen de t); es decir:

1. $E(X_t) = \mu \quad \forall t$ (el vector media es constante).
2. $Cov(X_t, X_{t-l}) = E[(X_t - \mu)(X_{t-l} - \mu)'] = \Gamma_l \quad \forall t$; es decir, la matriz de *covarianzas cruzadas* entre X_t y X_{t-l} es independiente de t (solo depende del salto l).

La media μ es un vector k -dimensional compuesto por las esperanzas de las componentes de X_t . La matriz de covarianzas cruzadas es de orden $k * k$.

El i -ésimo elemento de la diagonal de Γ_0 es la varianza de X_{it} ; mientras que, el elemento (i, j) de Γ_0 es la covarianza entre X_{it} y X_{jt} . El elemento (i, j) de Γ_l es la covarianza entre X_{it} y $X_{j,t-l}$.

Observación 2.1. Se puede demostrar que si un proceso $(X_t)_{t \in \mathbb{Z}}$ es débilmente estacionario entonces también lo será cada una de sus componentes.

2.2 Matrices de Correlación Cruzada (Cross-Correlation)

En lo que sigue se considera que $(X_t)_{t \in \mathbb{Z}}$ es estacionaria.

Sea D una matriz diagonal de orden $k * k$ compuesta por las desviaciones estándar de X_{it} para $i = 1, \dots, k$, que se denota por: $D = \text{diag} \{ \sqrt{\Gamma_{11}(0)}, \dots, \sqrt{\Gamma_{kk}(0)} \}$. La matriz de correlaciones cruzadas de X_t se define como:

$$\rho_0 \equiv [\rho_{ij}(0)] = D^{-1} \Gamma_0 D^{-1}$$

De manera particular, el elemento (i, j) de ρ_0 es:

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{Cov(X_{it}, X_{jt})}{de(X_{it})de(X_{jt})},$$

donde, $de(\cdot)$ es la desviación estándar.

$\rho_{ij}(0)$ es el coeficiente de correlación lineal entre X_{it} y X_{jt} . En el análisis de series de tiempo, dicho coeficiente se conoce como de concurrencia (en el mismo instante). Es fácil ver que:

- i) $\rho_{ij}(0) = \rho_{ji}(0)$
- ii) $-1 \leq \rho_{ij}(0) \leq 1$
- iii) $\rho_{ii}(0) = 1$

Así, $\rho(0)$ es una matriz simétrica con 1 en la diagonal.

Hay que mencionar que las matrices Γ_l contienen las **relaciones en retardo** entre las componentes de las series. Por lo tanto, las matrices de correlación cruzada se utilizan para medir la fuerza de la dependencia lineal entre las series de tiempo.

La matriz de correlación cruzada de X_t con X_{t-l} se define como:

$$\rho_l \equiv [\rho_{ij}(l)] = D^{-1} \Gamma_l D^{-1}$$

donde, D es la matriz diagonal de las desviaciones estándar de las series individuales. De la definición se tiene:

$$\rho_{ij}(l) = \frac{\Gamma_{ij}(l)}{\sqrt{\Gamma_{ii}(0) \Gamma_{jj}(0)}} = \frac{\text{Cov}(X_{it}, X_{jt-l})}{\text{de}(X_{it}) \text{de}(X_{jt})} = \frac{\text{Cov}(X_{it}, X_{jt-l})}{\text{de}(X_{it}) \text{de}(X_{jt-l})}$$

que es el coeficiente de correlación lineal entre X_{it} y X_{jt-l} . Cuando $l > 0$, este coeficiente de correlación mide la dependencia lineal de X_{it} con respecto a X_{jt-l} (X_{t-l} ocurre con anterioridad al instante t). Consecuentemente, si $\rho_{ij}(l) \neq 0$ y $l > 0$, se dice que la serie X_{jt} **conduce** a la serie X_{it} con retardo l .

Similarmente, $\rho_{ji}(l)$ mide la dependencia lineal de X_{jt} con respecto a X_{it-l} y se puede decir que la serie X_{it} **conduce** a la serie X_{jt} , con retardo l , si $\rho_{ji}(l) \neq 0$ y $l > 0$.

Se pueden mencionar las siguientes propiedades cuando $l > 0$:

1. En general, $\rho_{ij}(l) \neq \rho_{ji}(l)$ para $i \neq j$, porque los dos coeficientes de correlación miden diferentes relaciones lineales entre las series. Por lo tanto, Γ_l y ρ_l son, generalmente, no simétricas.
2. Utilizando $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ y suponiendo que las series son estacionarias, se tiene:

$$\text{Cov}(X_{it}, X_{jt-l}) = \text{Cov}(X_{jt-l}, X_{it}) = \text{Cov}(X_{jt}, X_{it+l}) = \text{Cov}(X_{jt}, X_{it-(-l)})$$

Así que $\Gamma_{ij}(l) = \Gamma_{ji}(-l)$, donde $\Gamma_{ji}(-l)$ es el elemento (j, i) de Γ_{-l} ; la igualdad se cumple para $1 \leq i, j \leq k$. Es decir, $\Gamma_l = \Gamma'_{-l}$.

2.2.1 Dependencia Lineal

Considérense las matrices de correlación cruzada $\{\rho(l) | l = 0, 1, 2, \dots\}$ de una serie temporal vectorial estacionaria; éstas contienen la siguiente información:

1. Los elementos de la diagonal de la matriz de correlación cruzada $\rho_{ii}(l)$ son las funciones de autocorrelación de X_{it} .
2. El elemento fuera de la diagonal $\rho_{ij}(0)$ mide la relación lineal de concurrencia entre X_{it} y X_{jt} .
3. Para $l > 0$, el elemento fuera de la diagonal $\rho_{ij}(l)$ mide la dependencia lineal de X_{it} con respecto a X_{jt-l} .

Por lo tanto, si $\rho_{ij}(l) = 0$ para todo $l > 0$, X_{it} no depende linealmente de ningún valor del pasado X_{jt-l} .

Resumen e interpretación.

En general, la relación lineal entre dos series de tiempo X_{it} y X_{jt} puede resumirse en la siguiente forma:

1. X_{it} y X_{jt} no tienen relación lineal si $\rho_{ij}(l) = \rho_{ji}(l) = 0, \quad \forall l \geq 0$.
2. X_{it} y X_{jt} están al mismo tiempo correlacionadas si $\rho_{ij}(0) \neq 0$.
3. X_{it} y X_{jt} no tienen relación de avance-retardo si $\rho_{ij}(l) = 0$ y $\rho_{ji}(l) = 0, \quad \forall l > 0$. En este caso, se dice que las series son desacopladas.
4. Existe una *relación unidireccional* desde X_{it} hacia X_{jt} si $\rho_{ij}(l) = 0, \quad \forall l > 0$, pero $\rho_{ji}(v) \neq 0$ para algún $v > 0$. En este caso, X_{it} no depende de ningún valor del pasado de X_{jt} , pero X_{jt} depende de algún valor del pasado de X_{it} .
5. Existe una *relación de retroalimentación* entre X_{it} y X_{jt} si $\rho_{ij}(l) \neq 0$ para algún $l > 0$ y $\rho_{ji}(v) \neq 0$ para algún $v > 0$.

Las formulaciones anteriores son suficientes para analizar la dependencia lineal entre series temporales. Un enfoque más informativo para estudiar las relaciones entre las series temporales es construir un modelo multivariante para las series, porque un modelo correctamente especificado considera simultáneamente el número de series y las correlaciones cruzadas de las mismas.

En la práctica se utilizan los estimadores de las matrices antes mencionadas; en particular para $\Gamma(l)$:

$$\hat{\Gamma}(l) = \frac{1}{T} \sum_{t=l+1}^T (X_t - \bar{X})(X_{t-l} - \bar{X})', \quad l \geq 0$$

donde,

T: número de observaciones

$$\bar{X} = \frac{\left(\sum_{t=1}^T X_t \right)}{T} : \text{vector de medias muestrales}$$

y para $\rho(l)$:

$$\hat{\rho}(l) = \hat{D}^{-1} \hat{\Gamma}_X(l) \hat{D}^{-1}, \quad l \geq 0$$

donde,

\hat{D} : Es la matriz diagonal de orden $(k * k)$ que contiene las desviaciones estándar muestrales del vector X_t en la diagonal.

Ejemplo 2.1. Se consideran tres series de datos económicos de un país sudamericano: el producto interno bruto (PIB), denotada por (X_{1t}) ; el consumo interno (CI), denotada por (X_{2t}) y la demanda final interna (DFI), denotada por (X_{3t}) . Se dispone de 56 datos trimestrales, desde noviembre de 2010 hasta junio de 2015 (Ver Anexo D.1). Para efectos de comparaciones se trabajará únicamente con los primeros 50 datos y se guardarán los 6 restantes para comparar con predicciones posteriores (enero 2015 - junio de 2015). Se desea estimar las matrices de correlaciones cruzadas de las series.

Resolución.

Las matrices de correlación cruzada se las construyen de manera manual, considerando cada escenario de posibles combinaciones entre las variables; así, en este caso, se obtiene:

a) Estadísticos descriptivos de x_{1t} , X_{2t} y X_{3t} .

	Media	Mediana	Máximo	Mínimo	Desv. Est.	Asimetría	Curtosis
X_{1t}	96,52	89,79	165,31	49,79	34,15	0,35	1,92
X_{2t}	78,10	73,72	126,01	43,05	24,89	0,31	1,84
X_{3t}	98,71	92,34	166,66	50,51	34,48	0,30	1,87

b) Matrices de correlación cruzada

	retardo 1			retardo 2			retardo 3		
X_{1t}	0,94	0,94	0,94	0,87	0,88	0,87	0,81	0,82	0,81
X_{2t}	0,93	0,94	0,94	0,87	0,88	0,87	0,80	0,81	0,81
X_{3t}	0,93	0,94	0,94	0,87	0,88	0,87	0,80	0,81	0,81

c) Representación simplificada

+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+

Tabla 2.1: Resumen de estadísticas y matrices de correlación cruzada para X_{1t} , X_{2t} y X_{3t}

Para representar a las matrices de correlación cruzada, se utiliza la forma gráfica simplificada, que utiliza el hecho que $2/\sqrt{T}$ (0,28 en este caso) es el valor crítico de la correlación muestral con nivel de significación del 5 %, bajo la suposición que X_t es un ruido blanco:

- “+” representa a los coeficientes de correlación que son mayores o iguales a $2/\sqrt{T}$.
- “-” representa a los coeficientes de correlación que son menores o iguales que $-2/\sqrt{T}$.
- “.” Representa a los coeficientes que se encuentran entre a $-2/\sqrt{T}$ y $2/\sqrt{T}$.

Es fácil ver que las correlaciones cruzadas son significativas en los primeros tres retardos. En algunos paquetes estadísticos se puede encontrar el cálculo de las matrices de correlación cruzada. En Eviews, por ejemplo, se presenta la siguiente salida:

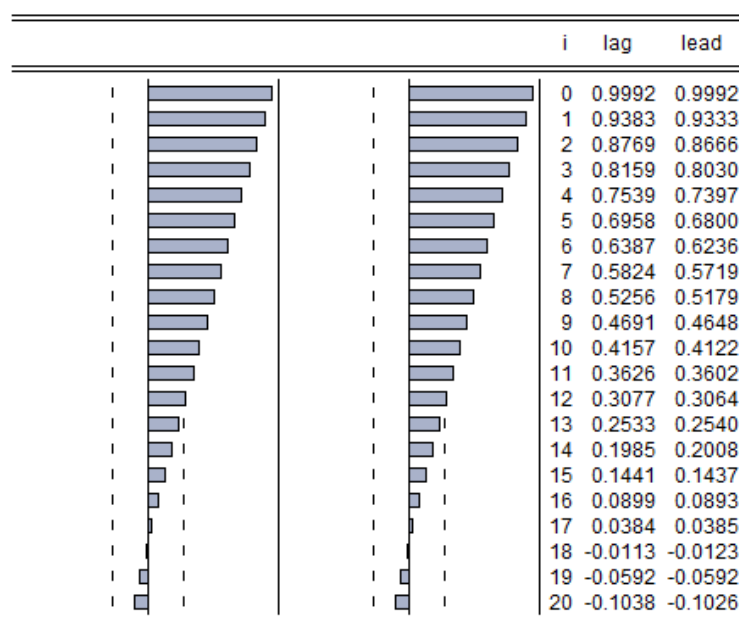


Figura 2.1: Correlaciones cruzadas entre X_{1t} y X_{2t}

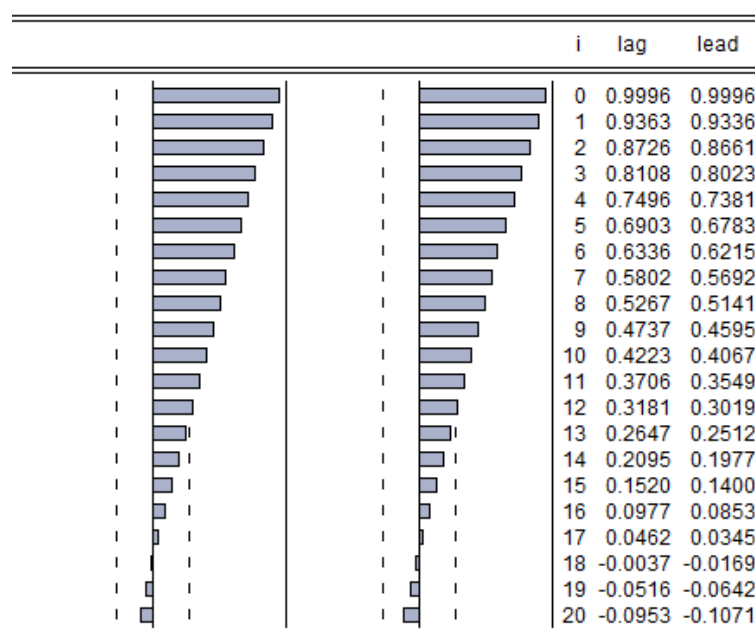


Figura 2.2: Correlaciones cruzadas entre X_{1t} y X_{3t}

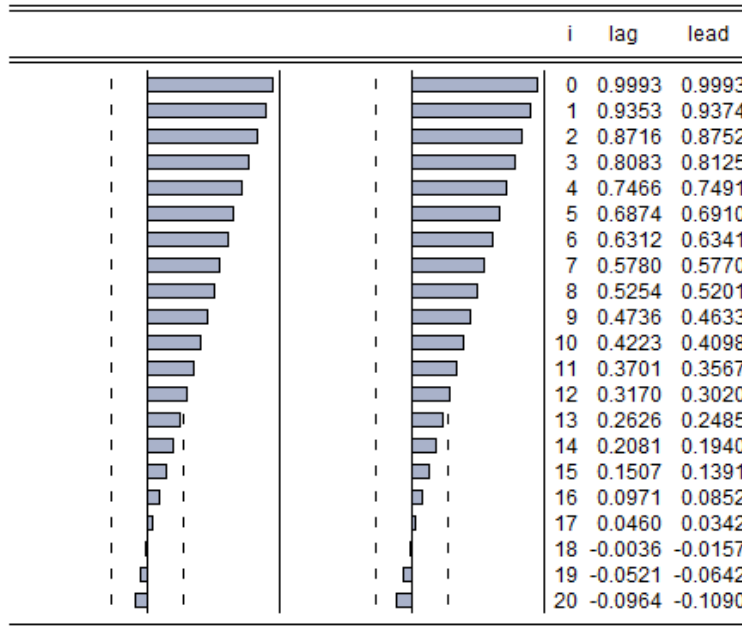


Figura 2.3: Correlaciones cruzadas entre X_{2t} y X_{3t}

Como se puede observar, los valores calculados por el paquete son aquellos que están en la diagonal secundaria de las matrices calculadas manualmente. Para poder completar la matriz, se puede, ver a partir de la fórmula de cálculo que las $\hat{\rho}_{ii}(l)$ corresponden a las autocorrelaciones simples de orden l de cada serie univariante dentro de X_t .

2.3 Modelos de Vectores Autoregresivos (VAR)

Este tipo de modelos no pertenecen a los modelos estocásticos desarrollados por Box y Jenkins; sin embargo, la representación VAR se puede considerar como la generalización de los modelos autoregresivos al caso multivariante.

2.3.1 El caso bivalente

Una representación VAR bivalente es aquella que consideran dos variables X_{1t} y X_{2t} . Cada una de ellas se expresa en función de sus propios valores del pasado y de los del presente y del pasado de la otra variable. Por ejemplo, se va a representar el modelo VAR bivalente de orden $p = 3$ [VAR (3)]; se escribe:

$$X_{1t} = v_1 + \sum_{i=1}^3 b_{1i}X_{1t-i} + \sum_{i=1}^3 c_{1i}X_{2t-i} - d_1X_{2t} + u_{1t}$$

$$X_{2t} = v_2 + \sum_{i=1}^3 b_{2i}X_{1t-i} + \sum_{i=1}^3 c_{2i}X_{2t-i} - d_2X_{1t} + u_{2t}$$

Las variables X_{1t} y X_{2t} son estacionarias; las perturbaciones u_{1t} y u_{2t} son ruidos blancos de varianzas constantes y no correlacionados. Se puede ver inmediatamente la gran cantidad

de parámetros a estimar (aquí 16 coeficientes), con los problemas típicos de pérdida de grados de libertad. Hay que tomar en cuenta que X_{1t} tiene un efecto inmediato en X_{2t} y recíprocamente. Este sistema inicial se denomina *forma estructural* de la representación VAR. Su *forma matricial*, se expresa como:

$$BX_t = v + \sum_{i=1}^3 \tilde{A}_i X_{t-i} + u_t$$

con:

$$B = \begin{bmatrix} 1 & d_1 \\ d_2 & 1 \end{bmatrix} \quad X_t = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \tilde{A}_i = \begin{bmatrix} b_{1i} & c_{1i} \\ b_{2i} & c_{2i} \end{bmatrix} \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Para obtener la *forma estándar* de un modelo VAR, se multiplica la ecuación anterior por B^{-1} (que se supone existe); es decir, se expresa por:

$$X_{1t} = v_1^0 + \sum_{i=1}^3 a_{1i}^1 X_{1t-i} + \sum_{i=1}^3 a_{1i}^2 X_{2t-i} + \vartheta_{1t}$$

$$X_{2t} = v_2^0 + \sum_{i=1}^3 a_{2i}^1 X_{1t-i} + \sum_{i=1}^3 a_{2i}^2 X_{2t-i} + \vartheta_{2t}$$

En esta especificación, los errores ϑ_{1t} y ϑ_{2t} son funciones de las innovaciones u_{1t} y u_{2t} ; en efecto, de $\vartheta = B^{-1}u$, se obtiene:

$$\vartheta_{1t} = \frac{(u_{1t} - d_1 u_{2t})}{(1 - d_1 d_2)} \text{ y } \vartheta_{2t} = \frac{(u_{2t} - d_2 u_{1t})}{(1 - d_1 d_2)}$$

Se puede ver que:

$$E(\vartheta_{1t}) = 0; \quad E(\vartheta_{2t}) = 0; \quad E(\vartheta_{1t} \vartheta_{1t-i}) = 0; \quad E(\vartheta_{2t} \vartheta_{2t-i}) = 0$$

Por lo tanto, los elementos de cada familia de errores tienen esperanza nula y son no correlacionados. Además:

$$E(\vartheta_{1t}^2) = \frac{(\sigma_{u_1}^2 + d_1^2 \sigma_{u_2}^2)}{(1 - d_1 d_2)^2}; \quad E(\vartheta_{2t}^2) = \frac{(\sigma_{u_2}^2 + d_2^2 \sigma_{u_1}^2)}{(1 - d_1 d_2)^2}$$

Donde $\sigma_{u_1}^2$ y $\sigma_{u_2}^2$ son las varianzas de u_1 y u_2 , respectivamente. Así, la varianza de los errores es constante (independiente del tiempo). Además:

$$E(\vartheta_{1t} \vartheta_{2t}) = -\frac{(d_2 \sigma_{u_1}^2 + d_1 \sigma_{u_2}^2)}{(1 - d_1 d_2)^2}$$

Si $d_1 = d_2 = 0$, las variables X_{1t} y X_{2t} no tienen ninguna influencia sincrónica entre sí, pues los errores ϑ_{1t} y ϑ_{2t} serían no correlacionados. En caso contrario, los errores ϑ_{1t} y ϑ_{2t} estarían correlacionados y por tanto, una variación de uno de estos errores en un instante dado tiene impacto en el otro.

Proposición 2.1. *El modelo VAR no permite distinguir entre variables endógenas (variables propias del fenómeno estudiado) y exógenas (variables externas que ayudan a explicar las*

variables endógenas).

2.3.2 Representación general de un VAR

Notación. Un modelo VAR a k variables con p retardos se denota $VAR(p)$.

La generalización de la representación VAR a k variables con p retardos se escribe en su forma estándar como:

$$X_t = v_0 + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + u_t$$

donde,

$$X_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{k,t} \end{bmatrix}; \quad v_0 = \begin{bmatrix} v_1^0 \\ v_2^0 \\ \vdots \\ v_k^0 \end{bmatrix}; \quad A_i = \begin{bmatrix} a_{1i}^1 & \dots & a_{1i}^k \\ \vdots & \ddots & \vdots \\ a_{ki}^1 & \dots & a_{ki}^k \end{bmatrix}; \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{kt} \end{bmatrix}$$

u_t es el vector compuesto por los ruidos blancos de cada una de las k ecuaciones del modelo.

Se denota por: $\Sigma_u = E(u_t u_t')$, la matriz desconocida, de dimensión k , de varianzas-covarianzas de los errores.

Esta representación puede escribirse mediante el operador de retardo B , como:

$$(I - A_1 B - A_2 B^2 - \dots - A_p B^p) X_t = v_0 + u_t, \quad \text{o también:} \quad A(B) X_t = v_0 + u_t$$

donde, el operador de retardo B se define de la siguiente manera:

$$B^i X_t = X_{t-i}, \quad i = 1, 2, \dots$$

$$B^0 X_t = X_t$$

2.3.2.1 Estabilidad de un VAR

Considérese un modelo $VAR(1)$:

$$X_t = v_0 + A_1 X_{t-1} + u_t$$

Se dice que un $VAR(1)$ es estable si todos los valores propios de A_1 son de valor absoluto menor que 1; lo que se puede expresar también por:

$$\det(I_k - A_1 z) \neq 0, \quad \text{para} \quad |z| \leq 1$$

Esto implica que todas las raíces del polinomio característico están fuera del círculo unidad.

2.3.2.2 Representación de un proceso VAR(p) en la forma de VAR(1)

Un proceso $VAR(p)$ se puede escribir como un proceso $VAR(1)$ si se plantea en la siguiente forma:

$$X_t = A_0 + AX_{t-1} + U_t$$

donde,

$$X_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix}_{kp*1} \quad X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{kt} \end{pmatrix}_{k*1} \quad A_0 = \begin{pmatrix} v_0 \\ 0_{k*1} \\ \vdots \\ 0_{k*1} \end{pmatrix}_{kp*1} \quad U_t = \begin{pmatrix} (u_t)_{k*1} \\ 0_{k*1} \\ \vdots \\ 0_{k*1} \end{pmatrix}_{kp*1}$$

$$A = \begin{pmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_k & 0 \end{pmatrix}_{kp*kp}$$

Lo importante de esta representación es que para obtener las propiedades de los procesos VAR, es suficiente con probarlas para una VAR (1)

2.3.2.3 Procesos VAR(p) estables

Se dice que un $VAR(p)$ es estable si:

$$\det(I_{kp} - Az) \neq 0, \text{ para } |z| \leq 1$$

Además, se puede demostrar que:

$$\det(I_{kp} - Az) = \det(I - A_1z - A_2z^2 - \dots - A_pz^p)$$

Observación 2.2. Se puede demostrar que si un proceso $VAR(1)$ es estable, entonces es estacionario.

En general, se puede demostrar que un proceso $VAR(p)$ es estacionario si el polinomio definido a partir de la expresión: $\det(I - A_1z - A_2z^2 - \dots - A_pz^p)$ tiene sus raíces fuera del círculo unidad del plano complejo; es decir:

$$\det(I - A_1z - A_2z^2 - \dots - A_pz^p) \neq 0 \forall z \text{ tal que } |z| \leq 1$$

Ejemplo 2.2. Determine si el siguiente modelo es estacionario.

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} + \begin{bmatrix} 0,8 & 0,9 \\ 0,7 & 0,7 \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Resolución.

Se tiene que:

$$\det \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,8 & 0,9 \\ 0,7 & 0,7 \end{bmatrix} z \right) = 1 - 1,5z - 0,07z^2, \quad \text{entonces} \quad \begin{matrix} z_1 = -16,17 \\ z_2 = -15,97 \end{matrix}$$

Las dos raíces son superiores a 1 en valor absoluto; por lo tanto, el proceso es estable; lo que implica que es estacionario.

2.4 Representación VARMA de una Serie Multivariante

2.4.1 La representación VMA

Un modelo media móvil vectorial de orden q ($VMA(q)$ por sus siglas en inglés), tiene la siguiente forma:

$$X_t = m_0 + u_t - M_1 u_{t-1} - \dots - M_q u_{t-q} \quad \text{o} \quad X_t = m_0 + M(B)u_t$$

donde,

m_0 : Es un vector de dimensión k constante.

M_i : Son matrices de dimensión $k * k$.

$M(B) = I - M_1 B - \dots - M_q B^q$ es el polinomio matriz MA en términos del operador de retardo B .

$\{u_t\}_{t \in \mathbb{Z}}$: Es un ruido blanco multidimensional.

De manera similar al caso univariante, los procesos $VMA(q)$ son débilmente estacionarios, siempre que la matriz de covarianzas (Σ_u) de u_t exista. Si se toma la esperanza de X_t , se tiene:

$$\mu = E(X_t) = m_0$$

Así, el vector constante m_0 es el vector media de X_t para un modelo VMA.

Se define $\tilde{X}_t = X_t - m_0$ como el proceso corregido en media $VAR(q)$. Cuando se tiene un proceso $VMA(q)$ y considerando el hecho de que los $\{u_t\}$ no están correlacionados, se obtiene:

1. $Cov(\tilde{X}_t, u_t) = \Sigma_u$
2. $\Gamma_0 = \Sigma_u + M_1 \Sigma_u M_1' + \dots + M_q \Sigma_u M_q'$
3. $\Gamma_l = 0$ si $l > q$
4. $\Gamma_l = \sum_{j=l}^q M_j \Sigma_u M_{j-l}'$ si $1 \leq l \leq q$, donde $M_0 = -I$

Dado que $\Gamma_l = 0$ para $l > q$, las matrices de correlación cruzada de un proceso $VMA(q)$ satisfacen:

$$\rho_l = 0, \quad l > q$$

2.4.2 Representación lineal de un VAR(p)

Cuando se analizaron las series temporales univariantes, se mostró que bajo ciertas condiciones un proceso $AR(1)$ se puede representar como un proceso lineal. De la misma manera,

para las series multivariantes se puede representar, en particular, un $VAR(1)$ como un proceso lineal (se dice que es la representación lineal del proceso). Un modelo con esta forma permite medir el impacto en los valores presentes de una variación de innovaciones (o choques).

Sea X_t un $VAR(1)$ estable:

$$X_t = v_0 + A_1 X_{t-1} + u_t$$

Si se realizan sustituciones repetidas en el proceso hasta el i -ésimo paso, se obtiene:

$$\begin{aligned} X_t &= v_0 + A_1 (v_0 + A_1 X_{t-2} + u_{t-1}) + u_t = (I + A_1) v_0 + A_1^2 X_{t-2} + (A_1 u_{t-1} + u_t) \\ X_t &= v_0 + A_1 \left((I + A_1) v_0 + A_1^2 X_{t-3} + A_1 u_{t-2} + u_{t-1} \right) + u_t \\ &= (I + A_1 + A_1^2) v_0 + A_1^3 X_{t-3} + (A_1^2 u_{t-2} + A_1 u_{t-1} + u_t) \\ &\vdots \\ X_t &= (I + A_1 + \dots + A_1^i) v_0 + A_1^{i+1} X_{t-i} + \sum_{j=0}^i A_1^j u_{t-j}, \quad i = 0, 1, 2, \dots \end{aligned}$$

Por definición, $A^0 = I$.

Como el VAR es estable, se cumple que:

$$(I + A_1 + \dots + A_1^i) v_0 \rightarrow (I - A_1)^{-1} v_0 \quad \text{si } i \rightarrow \infty$$

Además, $A_1^{i+1} \rightarrow 0$ rápidamente; así, se lo puede ignorar. Por lo tanto, se obtiene la siguiente representación:

$$X_t = (I - A_1)^{-1} v_0 + \sum_{i=0}^{\infty} A_1^i u_{t-i}$$

La generalización a un proceso $VAR(p)$ se la realiza aplicando la representación de un $VAR(p)$ como un $VAR(1)$. Así, se obtiene:

$$X_t = \mu + \sum_{i=0}^{\infty} M_i u_{t-i}$$

donde,

$$\begin{aligned} \mu &= (I - A_1 - A_2 - \dots - A_p)^{-1} v_0 \\ M_i &= \sum_{j=1}^{\min(p,i)} A_j M_{i-j} \quad i = 1, 2, \dots \quad \text{y} \quad M_0 = I \end{aligned}$$

Las matrices M_i aparecen como un “*factor de impacto*”, a través de las cuales se analiza el efecto de un choque a lo largo de todo el proceso.

Observación 2.3.

1. Así, se obtiene que si un proceso $VAR(p)$ es estable, tiene una representación lineal estacionaria.
2. No se profundiza sobre la modelación $VMA(q)$ porque no está implementada en los

programas comerciales usuales.

Ejemplo 2.3. Considérese el proceso VMA (1):

$$X_t = \mu + u_t - M_1 u_{t-1} = \mu + u_t - M u_{t-1}$$

donde, por simplicidad, se ha quitado el subíndice de M_1 . Este modelo puede escribirse explícitamente como:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} - \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}$$

Se dice que la serie de retardos (X_t) solo depende del presente y del pasado de $\{u_t\}$. Por lo tanto, el modelo es de memoria finita.

El parámetro m_{12} denota la dependencia lineal de X_{1t} con $u_{2,t-1}$ en la presencia de $u_{1,t-1}$. Si $m_{12} = 0$, X_{1t} no depende de los retardos de u_{2t} y, entonces tampoco, de los retardos de X_{2t} . De manera similar, si $m_{21} = 0$, X_{2t} no depende de los valores pasados de X_{1t} . Los elementos fuera de la diagonal de M muestran la dependencia entre las componentes de las series.

Para este ejemplo, se pueden clasificar las relaciones entre X_{1t} y X_{2t} así:

1. Son series desacopladas si $m_{12} = m_{21} = 0$.
2. Hay una relación dinámica unidireccional de X_{1t} sobre X_{2t} si $m_{12} = 0$, pero $m_{21} \neq 0$ y viceversa.
3. Hay una relación de retroalimentación entre X_{1t} y X_{2t} si $m_{12} \neq 0$ y $m_{21} \neq 0$.

Finalmente, la correlación actual entre los m_{ij} (coeficientes estimados para el modelo VMA) es la misma que entre los u_{it} . La descripción previa se puede generalizar para un modelo $VMA(q)$.

2.4.3 La representación VARMA

La representación VAR puede generalizarse (es una aplicación multivariante del teorema de descomposición de Wold (1954)), por analogía con los procesos $ARMA(pq)$.

$$X_t = A_0 + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + u_t + M_1 u_{t-1} + M_2 u_{t-2} + \dots + M_q u_{t-q}$$

Se trata de un proceso ARMA multivariante que se denota: VARMA.

Las condiciones de estacionariedad son análogas a las de un proceso ARMA univariante:

- Un proceso VAR es siempre invertible; es lineal (por ende estacionario) cuando es estable.
- Un proceso VMA es siempre estacionario. Es invertible si las raíces del polinomio característico asociado a $M(z)$ están fuera del círculo unitario complejo.
- Las condiciones de estacionariedad e invertibilidad de un VARMA están dadas, respectivamente, por la parte VAR y la parte VMA del VARMA.

La generalización de los modelos ARMA encuentra nuevos temas que no ocurren en el desarrollo de los modelos VAR y VMA. Uno de ellos es el *problema de identificación*. A diferencia de los modelos ARMA, los modelos VARMA pueden no estar definidos de manera única.

Ejemplo 2.4. Considere un modelo bivalente $VMA(1)$:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}$$

Es *idéntico* al modelo bivalente $VAR(1)$:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

La equivalencia de los modelos se puede examinar fácilmente componente a componente. Es decir, para el modelo $VMA(1)$ se tiene:

$$X_{1t} = u_{1t} - 2u_{2,t-1} \quad y \quad X_{2t} = u_{2t}$$

Por otro lado, para el modelo $VAR(1)$ se tiene:

$$X_{1t} + 2X_{2,t-1} = u_{1t} \quad y \quad X_{2t} = u_{2t}$$

De los modelos se puede ver que:

$$X_{2,t-1} = u_{2,t-1}$$

Luego, los modelos para X_{1t} son idénticos. Este tipo de problema de identificación es inofensivo porque cualquiera de los modelos puede ser utilizado en una aplicación real. Sin embargo, existen casos en los que esta situación si se convierte en un problema y hay que tener en cuenta muchas restricciones para poder estimar un modelo VARMA.

2.5 Formulación de un modelo VAR

Los parámetros de un proceso VAR pueden estimarse solamente en las series temporales estacionarias. Se conoce que muchas series pueden volverse estacionarias a través de un proceso de diferenciación (en el caso de una tendencia determinista o una estacionalidad) o a través de una transformación de las variables (por ejemplo, una transformación logarítmica) en ciertos casos con heteroscedasticidad.

2.5.1 Estimación

En el caso de un proceso VAR, las ecuaciones pueden estimarse por MCO independientemente una de la otra (o por un método de máxima verosimilitud).

Sea el modelo $VAR(p)$ estimado:

$$X_t = \hat{A}_0 + \hat{A}_1 X_{t-1} + \hat{A}_2 X_{t-2} + \dots + \hat{A}_p X_{t-p} + \hat{u}_t$$

siendo, \hat{u}_t el vector de dimensión $(k, 1)$ de componentes $\hat{u}_{1t}, \hat{u}_{2t}, \dots, \hat{u}_{kt}$.

Se denotará por $\hat{\Sigma}_{u,p}$ la matriz de varianzas covarianzas estimada de los residuos del modelo. Para cualquier orden p , se define por:

$$\hat{\Sigma}_{u,p} = \frac{1}{T - kp - 1} \sum_{t=p+1}^T \hat{u}_t (\hat{u}_t)', \quad p \geq 0$$

2.5.2 Determinación del número de retardos

El mayor problema que debe enfrentarse a la hora de estimar los modelos VAR es el de la determinación del número de retardos a incluir en la estimación; suele realizarse en forma cuantitativa, analizando los resultados de la estimación y comparando los resultados obtenidos entre distintos modelos alternativos, ya que no es frecuente encontrar evidencias teóricas al respecto.

Los criterios comúnmente utilizados para la selección entre modelos alternativos son el criterio informativo de Akaike (AIC), el criterio de información bayesiano (BIC), que también se conoce como el criterio de Schwarz (SC) o el criterio de Hanan-Quinn (HQ).

Para el caso de la representación VAR, estos criterios se pueden utilizar para determinar el orden p del modelo. EL proceso de selección del orden de la representación consiste en estimar todos los modelos VAR para retardos de 0 a p_0 (p_0 es el máximo retardo admisible por la teoría económica o por los datos disponibles y se fija de antemano). Los estadísticos AIC(p), SC(p) y HQ(p) para el caso multivariante tienen las siguientes expresiones:

$$AIC(p) = \ln \left[\left| \sum_u \right| \right] + \frac{2k^2 p}{T}$$

$$SC(p) = \ln \left[\left| \sum_u \right| \right] + \frac{k^2 p \ln(T)}{T}$$

$$HQ(p) = \ln \left[\left| \sum_u \right| \right] + \frac{2k^2 p \ln(\ln(T))}{T}$$

donde,

k = número de variables del sistema.

T = número de observaciones.

p = número de retardos.

\sum_u = matriz de varianzas covarianzas de residuos del modelo con retardo p (fijo).

Otro criterio utilizado para determinar el retardo del modelo, es la razón de máxima verosimilitud. Para utilizar este criterio, es necesario que el vector de las innovaciones tenga una distribución normal; el logaritmo de la función de verosimilitud tiene la siguiente expresión:

$$l = -\frac{Tp}{2} (1 + \ln(2\pi)) - \frac{T}{2} \ln \left[\left| \hat{\Sigma}_u \right| \right]$$

Observación 2.4. El retardo p que minimice la mayor cantidad de los criterios de AIC, HQ, BIC; o, maximice el logaritmo de la función de verosimilitud, se retiene. En la práctica se aconseja que $p \leq 5$, debido a que valores superiores implican incorporar una gran cantidad de parámetros.

También se puede utilizar el estadístico $M(p)$ para probar la hipótesis nula H_0 : El modelo es un $VAR(p)$ contra la alternativa, H_1 : El modelo es un $VAR(p-1)$.

Este estadístico se define por:

$$M(p) = - \left(T - k - p - \frac{3}{2} \right) \ln \left[\frac{|\hat{\Sigma}_{u,p}|}{|\hat{\Sigma}_{u,p-1}|} \right]$$

donde,

$\hat{\Sigma}_{u,j}$ =matriz de varianzas covarianzas del modelo con retardo j .

$M(p)$ sigue asintóticamente una distribución Ji-cuadrado con k^2 grados de libertad.

2.5.3 Diagnóstico y validación del modelo

Un buen punto de partida para la verificación de que el modelo estimado es el adecuado, es la significación de los parámetros estimados, para no tener parámetros no deseados o parámetros que no aportan al modelo. Por otro lado, esto puede ser engañoso, porque los parámetros estimados de un modelo pobre pueden ser también significativos. Por lo tanto, no se debe depender exclusivamente de la significación de los parámetros para evaluar el modelo.

Como en la mayoría de situaciones de modelación, la forma de evaluación se realiza a través del comportamiento de los residuos. Si el modelo es una representación adecuada de un proceso generado por las series de tiempo, los residuos no deben tener ninguna tendencia significativa ni patrón.

Una forma de observar esto es considerar los elementos individuales de la matriz de autocorrelación de los vectores de residuos. Otra forma es el uso del estadístico *Portmanteau*, que se analizará posteriormente.

2.5.3.1 Matrices de autocorrelación multivariante

Sea $\{u_t\}$ un ruido blanco k -dimensional con matriz de covarianza Σ_u y su correspondiente matriz de correlación R_u . La matriz de autocovarianza y la matriz de autocorrelación muestral de $\{u_t\}$ con respecto al retardo i están dadas por:

$$\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^T \hat{u}_t \hat{u}_{t-i}' \quad i = 0, 1, \dots; i < T$$

$$\hat{R}_i = V_u^{-\frac{1}{2}} \hat{C}_i V_u^{-\frac{1}{2}} \quad i = 0, 1, \dots; i < T$$

donde, T es el número de observaciones de las series de tiempo y $V_u^{-\frac{1}{2}}$ es una matriz diagonal $(k * k)$ con el inverso de la raíz cuadrada de los elementos de la diagonal de C_0 en su

diagonal.

Sea $R_l^* = (R_1, \dots, R_l)'$.

2.5.3.2 La prueba "Portmanteau"

La prueba de bondad de ajuste para los residuos de Box-Pierce (1970), la prueba *Portmanteau*, fue extendida a modelos VAR multivariante por Hosking (1980) y Li-McLeod (1981). Esta prueba determina si las autocorrelaciones residuales, sobre un retardo específico, son estadísticamente nulos.

La hipótesis que se prueba es:

$$H_0 : R_l^* = (R_1, \dots, R_l)' = 0 \quad \text{contra} \quad H_a : R_l^* = (R_1, \dots, R_l)' \neq 0$$

Si no se rechaza la hipótesis nula, se puede asumir que los residuos se comportan como un ruido blanco y, por lo tanto, es adecuado el modelo ajustado.

La prueba multivariante *Portmanteau* propuesta por Hosking (1980) considera el estadístico:

$$Q(l) = T \sum_{i=1}^l \text{tr} \left(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1} \right)$$

Este estadístico tiene aproximadamente una distribución Ji-Cuadrada con $k^2(l-p)$ grados de libertad bajo la hipótesis nula, donde p es el orden estimado del modelo VAR (p) y l es el número de retardos incluidos en la prueba para la significación total. Ljung-Box (1978) propusieron una modificación que conduce a propiedades mejores en el caso univariante; Hosking considera una modificación similar para el caso multivariante. El estadístico modificado de la prueba *Portmanteau* está dado por:

$$Q'(l) = T^2 \sum_{i=1}^l (T-i)^{-1} \text{tr} \left(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1} \right)$$

2.5.3.3 Prueba de Breusch - Godfrey o Prueba del Multiplicador de Lagrange (LM)

Se utiliza para detectar autocorrelación de cualquier orden, especialmente en aquellos modelos con o sin variables dependientes retardadas. Permite determinar si existe correlación en los residuos hasta un determinado orden.

Se realiza la siguiente prueba de hipótesis:

$$H_0 : \rho_l = 0, \quad \text{contra} \quad H_a : \rho_l \neq 0,$$

donde, l es el orden del modelo VAR ajustado.

El estadístico utilizado para la prueba es:

$$LM = TR^2$$

donde, T el número de observaciones y R^2 corresponde a la bondad de ajuste de la regresión auxiliar entre las variables y los residuos.

Este estadístico, bajo H_0 , sigue asintóticamente una distribución Ji-cuadrado con l grados de libertad, χ_l^2 .

2.5.3.4 Prueba de Jarque-Bera

Es una prueba asintótica de normalidad para grandes muestras. La prueba de Jarque-Bera (JB) considera la relación entre los coeficientes de asimetría y apuntamiento de los residuos de la ecuación estimada y los correspondientes de una distribución normal, de forma tal que si estas relaciones son suficientemente diferentes se rechazará la hipótesis nula de normalidad.

Se realiza la siguiente prueba de hipótesis:

H_0 : los residuos siguen una distribución normal multivariante

H_1 : los residuos no siguen una distribución normal multivariante

Este estadístico se basa en las medidas de apuntamiento (*curtosis*) y la asimetría a través de la transformación de Mahalanobis.

La i -ésima componente del vector de asimetría estimado, se calcula de la siguiente manera:

$$as_i = \frac{\frac{1}{T} \sum_{j=1}^T \hat{\sigma}_{ij}^3}{\frac{1}{T} \sum_{j=1}^T (\hat{\sigma}_{ij}^2)^{3/2}} = \frac{\sum_{j=1}^T \hat{\sigma}_{ij}^3}{\sum_{j=1}^T (\hat{\sigma}_{ij}^2)^{3/2}}$$

La i -ésima componente del vector de apuntamiento estimado se calcula de la siguiente manera:

$$k_i = \frac{\frac{1}{T} \sum_{j=1}^T \hat{\sigma}_{ij}^4}{\frac{1}{T} \sum_{j=1}^T (\hat{\sigma}_{ij}^2)^2} = \frac{\sum_{j=1}^T \hat{\sigma}_{ij}^4}{\sum_{j=1}^T (\hat{\sigma}_{ij}^2)^2}$$

$\hat{\sigma}_{ij}$ son los elementos de la matriz \hat{V} , que se define de la siguiente manera:

$$\hat{V} = \hat{U} S_{\hat{U}}^{-1}$$

donde, \hat{U} es la matriz de los residuos obtenidos a través de la estimación de las variables utilizando el método de mínimos cuadrados; mientras que $S_{\hat{U}}$ es una matriz triangular superior tal que:

$$\hat{U}' \hat{U} = S_{\hat{U}}' S_{\hat{U}} \quad \text{y} \quad (\hat{U}' \hat{U})^{-1} = S_{\hat{U}}^{-1} (S_{\hat{U}}^{-1})'$$

En este caso, \hat{V} es la matriz ortogonalizada de los residuos estimados; es decir, as_i y k_i corresponden a la asimetría y el apuntamiento individual estimados, respectivamente.

Entonces, se define a la asimetría y al apuntamiento estimados de la distribución de la serie multivariante como:

$$AS = (as_1, \dots, as_T)' (as_1, \dots, as_T)$$

$$K = (k_1 - 3, \dots, k_T - 3)' (k_1 - 3, \dots, k_T - 3)$$

El estadístico utilizado para la prueba es:

$$JB = T \left[\frac{AS}{6} + \frac{K}{24} \right]$$

Este estadístico se compara con una distribución Ji-Cuadrada con $2T$ grados de libertad.

2.6 Predicción

Con los coeficientes estimados del modelo, se puede calcular la predicción para un horizonte h , dada la información hasta el período T ; por ejemplo, para un VAR (1) se tiene:

$$\hat{X}_T(1) = \hat{v}_0 + \hat{A}_1 X_T$$

Al horizonte de 2 períodos, la predicción es:

$$\hat{X}_T(2) = \hat{v}_0 + \hat{A}_1 \hat{X}_T(1) = \hat{v}_0 + \hat{A}_1 \hat{v}_0 + \hat{A}_1^2 X_T$$

Al horizonte de 3 períodos, la predicción se escribe:

$$\hat{X}_T(3) = \hat{v}_0 + \hat{A}_1 \hat{X}_T(2) = (I + \hat{A}_1 + \hat{A}_1^2) \hat{v}_0 + \hat{A}_1^3 X_T$$

$$\hat{X}_T(h) = \hat{v}_0 + \hat{A}_1 \hat{X}_T(h-1) = (I + \hat{A}_1 + \dots + \hat{A}_1^{h-1}) \hat{v}_0 + \hat{A}_1^h X_T, \quad h \geq 0$$

Cuando $h \rightarrow \infty$, la previsión tiende a un valor constante (estado estacionario) puesto que $\hat{A}_1^i \rightarrow 0$ si $i \rightarrow \infty$ y existe el límite de $\sum_{j=0}^{\infty} \hat{A}_1^j$, que es igual a $(I - \hat{A}_1)^{-1}$. Por tanto:

$$\hat{X}_T(h) \rightarrow (I - \hat{A}_1)^{-1} \hat{v}_0 \quad \text{cuando } h \rightarrow \infty$$

El error de predicción al horizonte h viene dado por:

$$e_T(h) = X_{T+h} - \hat{X}_T(h)$$

En particular, para $h=1$ y $h=2$, se tiene:

$$e_T(1) = u_{T+1}$$

$$e_T(2) = u_{T+2} + \hat{A}_1 u_{T+1}$$

En general, para el horizonte h , se tiene:

$$e_T(h) = \sum_{i=0}^{h-1} \hat{A}_1^i u_{T+h-i}$$

La esperanza del error de predicción es nula. La matriz de varianza-covarianza del error de predicción es:

$$\sum_u(h) = E \left[\left(\sum_{i=0}^{h-1} A_1^i u_{T+h-i} \right) \left(\sum_{i=0}^{h-1} A_1^i u_{T+h-i} \right)' \right]$$

La varianza-covarianza estimada viene dada por:

$$\hat{\sum}_u(h) = E \left[\left(\sum_{i=0}^{h-1} \hat{A}_1^i \hat{u}_{T+h-i} \right) \left(\sum_{i=0}^{h-1} \hat{A}_1^i \hat{u}_{T+h-i} \right)' \right]$$

Luego,

$$\hat{\sum}_T(h) = M_0 \hat{\sum}_u M'_0 + M_1 \hat{\sum}_u M'_1 + \dots + M_{h-1} \hat{\sum}_u M'_{h-1}$$

donde M_i son las matrices de la representación VMA.

Por lo tanto, se tiene:

$$M_1 = \hat{A}_1; \quad M_2 = \hat{A}_1 M_1 + \hat{A}_2 M_0 = \hat{A}_1^2 + \hat{A}_2;$$

$$M_3 = \hat{A}_1 M_2 + \hat{A}_2 M_1 + \hat{A}_3 M_0 = \hat{A}_1^3 + \hat{A}_1 \hat{A}_2 + \hat{A}_2 \hat{A}_1 + \hat{A}_3$$

La varianza del error de predicción para cada una de las predicciones de las k variables $(\hat{\sigma}_i^2(h))$ se lee sobre la primera diagonal de la matriz $\hat{\sum}_u(h)$. El intervalo de predicción al nivel $(1 - \alpha)$ está dado por: $\hat{X}_{iT}(h) \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_i(h)$ donde $z_{1-\frac{\alpha}{2}}$ es el cuantil de orden $(1-\alpha/2)$ de la ley normal.

Ejemplo 2.5. Considerando las series del ejemplo 5.1, se busca modelar en su forma VAR. Sin embargo, las series en el ejemplo 5.1 están en niveles y como se pudo observar no son estacionarias; por esta razón, se trabajará con las variaciones trimestrales de las series. Así, se tendrá: la variación trimestral del PIB (Y_{1t}), del CI (Y_{2t}) y de la DFI (Y_{3t}) de un país sudamericano. Nótese que se ha denotado a las variaciones de las series con Y ; para las variables originales se dejará la notación con X . Se trabajará con 49 datos, dado que al calcular las variaciones, se pierde el primer dato. La serie Y iniciará en la observación 2 hasta la observación 50. (Ver Anexo D.1).

Determinar:

- El orden del modelo VAR.
- Los parámetros del modelo.
- La predicción para las 6 siguientes observaciones y dar el intervalo de confianza al 95 %.

Resolución.

Se inicia presentando el gráfico de las series:

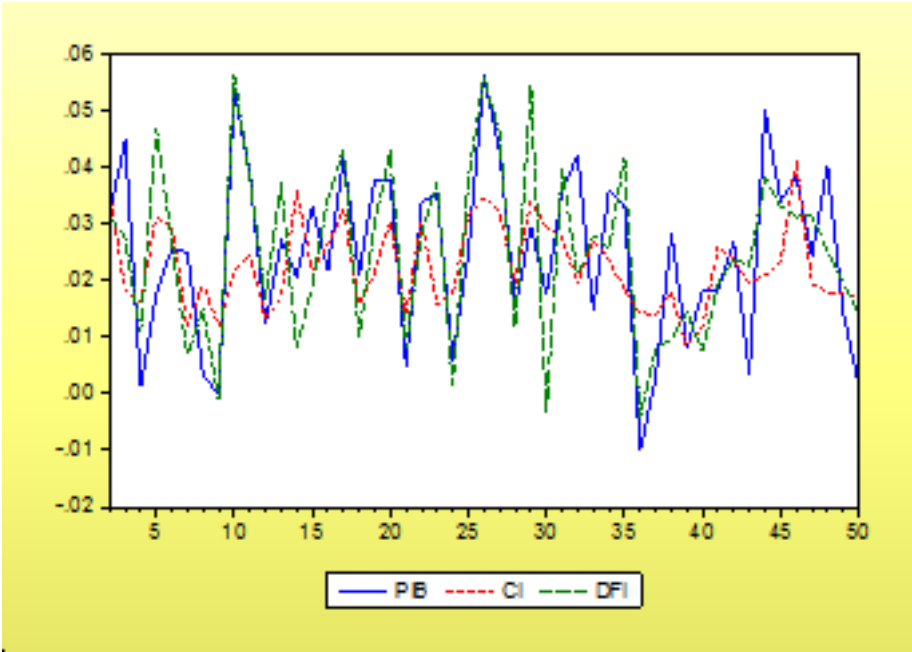


Figura 2.4: Gráfico de las variaciones trimestrales de las series PIB, CI y DFI

Antes de realizar los procedimientos para estimar los modelos, se debe verificar si las series a ser analizadas son estacionarias; para ello se realiza la prueba de raíces unitarias para cada serie utilizando el programa EViews:

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6,689257	0,0000
Test critical values:	1 % level	-4,161144	
	5 % level	-3,506374	
	10 % level	-3,183002	

Tabla 2.2: Prueba DFA para Y_{1t}

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6,231136	0,0000
Test critical values:	1 % level	-4,161144	
	5 % level	-3,506374	
	10 % level	-3,183002	

Tabla 2.3: Prueba DFA para Y_{2t}

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7,729086	0,0000
Test critical values: 1 % level	-4,161144	
5 % level	-3,506374	
10 % level	-3,183002	

Tabla 2.4: Prueba DFA para Y_{3t}

Como se puede ver en las tablas 2.2, 2.3 y 2.4, las tres series son estacionarias.

- a) Se utilizarán los criterios de Akaike, Schwarz y el logaritmo de máxima verosimilitud para determinar el retardo p entre 1 y 4. Se deben estimar cuatro modelos diferentes y retener aquel que satisfaga la mayor cantidad de criterios óptimos.

Inicialmente, se tiene un modelo de la forma, para $p=1$:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ Y_{3t} \end{bmatrix} = \begin{bmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \end{bmatrix} + \begin{bmatrix} a_{11}^1 & a_{11}^2 & a_{11}^3 \\ a_{21}^1 & a_{21}^2 & a_{21}^3 \\ a_{31}^1 & a_{31}^2 & a_{31}^3 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ Y_{3t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

Algunos paquetes econométricos utilizan la estimación de MCO ecuación por ecuación ya que sólo los valores rezagados de las variables endógenas aparecen en el lado derecho de la ecuación, lo que hace que los estimadores sean eficientes.

Se obtiene lo siguiente:

$$\hat{Y}_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115$$

$$\hat{Y}_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197$$

$$\hat{Y}_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185$$

Con la ayuda del paquete EViews 7, se realiza la estimación de los 4 modelos. Así, se obtuvieron los siguientes resultados:

Criterio / Retardo	1	2	3	4
Log likelihood	484,96	479,76	476,41	477,71
Akaike	-18,92	-18,72	-18,60	-18,67
Schwarz	-18,46	-17,91	-17,43	-17,13

Tabla 2.5: Criterios para escoger el retardo del VAR

Como se puede observar en la tabla 5.5, es en el retardo 1 ($p = 1$) donde los criterios de Akaike y Schwarz se minimizan y el valor del *log de verosimilitud* es el máximo. Por lo tanto se realiza la estimación del VAR(1).

En el caso del programa EViews, para calcular el VAR se debe ingresar el número de retardos como un rango; por ejemplo, para un VAR(1) se debe especificar como "1 1", para un

VAR(4) se lo especifica como "1 4", "2 4", "3 4" o "4 4", dependiendo del rango de retardos que se requiera en el modelo.

Figura 2.5: Especificación de un modelo VAR en EViews

b) El modelo VAR estimado se escribe:

$$Y_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115 + \hat{u}_{1t}$$

(-0,55)
(2,05)
(-0,09)
(-1,66)

$R^2 = 0,09$; $n = 50$; $(.)$ = estadístico correspondiente a la distribución t de Student.

$$Y_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197 + \hat{u}_{2t}$$

(-0,56)
(0,40)
(0,83)
(5,85)

$R^2 = 0,03$; $n = 50$; $(.)$ = estadístico correspondiente a la distribución t de Student.

$$Y_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185 + \hat{u}_{3t}$$

(0,92)
(1,18)
(-1,52)
(2,61)

$R^2 = 0,06$; $n = 50$; $(.)$ = estadístico correspondiente a la distribución t de Student.

Antes de realizar las predicciones, se debe verificar si el modelo cumple con el criterio de estabilidad (las raíces del polinomio característico están fuera del círculo unidad). Con ayuda

del programa EViews se pueden calcular las inversas de las raíces del polinomio característico autoregresivo, las que se espera que se encuentren dentro del círculo unidad. Así se obtiene:

Root	Modulus
-0,246672 - 0,062904i	0,254566
-0,246672 + 0,062904i	0,254566
0,097725	0,097725

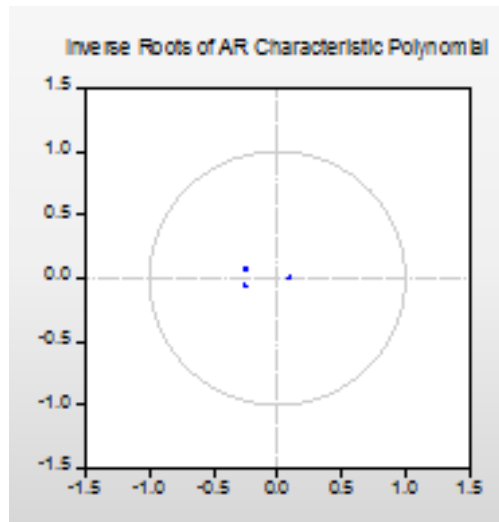


Figura 2.6: Criterio de estabilidad para el VAR(1) estimado

Analítica y gráficamente, se concluye que las inversas de las raíces del polinomio característico se encuentran dentro del círculo unidad; por lo tanto, se concluye que el modelo es estable y, por tanto, es estacionario.

- c) Ahora, se necesita verificar que los residuos del modelo sean ruidos blancos; en general, se prueba la independencia. Para ello, se utilizará el paquete EViews para obtener las pruebas sobre los residuos que se describieron anteriormente. Así se obtiene:

Prueba Portmanteau

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	0,628382	NA*	0,641751	NA*	NA*
2	9,516110	0,3911	9,915902	0,3573	9
3	16,74629	0,5406	17,62809	0,4804	18
4	25,33301	0,5558	26,99543	0,4640	27
5	35,85563	0,4754	38,74160	0,3470	36
6	41,19646	0,6338	44,84542	0,4784	45
7	51,95004	0,5539	57,43497	0,3491	54
8	62,57741	0,4913	70,18781	0,2494	63
9	72,52842	0,4604	82,43521	0,1879	72
10	80,22987	0,5033	92,16336	0,1863	81
11	88,06759	0,5380	102,3312	0,1763	90
12	89,89910	0,7324	104,7732	0,3264	99
13	94,98418	0,8100	111,7471	0,3831	108
14	99,40624	0,8789	117,9900	0,4570	117
15	103,7800	0,9263	124,3517	0,5248	126
16	115,1927	0,8904	141,4708	0,3343	135
17	118,4536	0,9412	146,5199	0,4259	144
18	127,4406	0,9348	160,8991	0,3149	153
19	130,5767	0,9668	166,0899	0,3965	162
20	136,6630	0,9751	176,5237	0,3701	171

Tabla 2.6: Prueba de autocorrelación Pormanteau

En la tabla 5.6 se observa que los p-valores (Prob.) para los retardos 2 al 10 son no significativos; de esto se concluye que los residuos no están autocorrelacionados.

Prueba LM

Lags	LM-Stat	Prob
1	6,309247	0,7086
2	8,996282	0,4376
3	7,384406	0,5972
4	8,563089	0,4785
5	11,86241	0,2212
6	5,615736	0,7777
7	13,14360	0,1562
8	12,15696	0,2046
9	13,95395	0,1240
10	9,385154	0,4025
11	9,988141	0,3514
12	2,713061	0,9746
13	5,772031	0,7625
14	6,030074	0,7369
15	6,067111	0,7332
16	16,42881	0,0584
17	5,150654	0,8210
18	13,63067	0,1361
19	4,863835	0,8460
20	7,838041	0,5505

Tabla 2.7: Prueba LM

Por los valores en la columna Prob. (ver tabla 5.7), se puede concluir que no existe correlación serial. Esto confirma que no hay que reformular el modelo planteado.

Pruebas de Normalidad

Component	Jarque-Bera	df	Prob.
1	1,142571	2	0,5648
2	2,337811	2	0,3107
3	0,393216	2	0,8215
Joint	3,873599	6	0,6938

Tabla 2.8: Prueba de Normalidad de los residuos

De la prueba de Jarque-Bera se concluye que la distribución de los residuos es una distribución normal multivariante.

Observación 2.5. De los resultados obtenidos en los literales b y c se concluye que el modelo VAR(1) es adecuado para los datos.

d) La predicción calculada por el modelo, de manera recurrente es:

$$\hat{Y}_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115$$

$$\hat{Y}_{1:51} = -0,1168 * 0,0015 + 0,7168 * 0,0167 - 0,0198 * 0,0142 + 0,0115$$

$$\hat{Y}_{1:51} = 0,023$$

$$\hat{Y}_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197$$

$$\hat{Y}_{2:51} = -0,0575 * 0,0015 + 0,0674 * 0,0167 + 0,0900 * 0,0142 + 0,0197$$

$$\hat{Y}_{2:51} = 0,022$$

$$\hat{Y}_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185$$

$$\hat{Y}_{3:51} = 0,1972 * 0,0015 + 0,4192 * 0,0167 - 0,3462 * 0,0142 + 0,0185$$

$$\hat{Y}_{3:51} = 0,021$$

donde, $\hat{Y}_{i:j}$ significa, la previsión de la variable Y_i para el período j .

Observación 2.6. A pesar que se trabaja con 49 datos, la última observación es la número 50, por los motivos explicados en el enunciado del ejemplo; es por esto, que la primera observación a predecir es la de orden 51 aunque en realidad correspondería al dato 50 de una serie temporal que inicie en el instante $t=1$.

De la misma manera se obtiene:

$$\hat{Y}_{1:52} = 0,024$$

$$\hat{Y}_{2:52} = 0,022$$

$$\hat{Y}_{3:52} = 0,025$$

$$\hat{Y}_{1:53} = 0,023$$

$$\hat{Y}_{2:53} = 0,022$$

$$\hat{Y}_{3:53} = 0,023$$

\vdots

Para calcular la varianza del error de predicción, si fuera un VAR (1) se tiene:

$$M_1 = \hat{A}_1; \quad M_2 = \hat{A}_1 M_1 = \hat{A}_1^2; \quad etc \dots$$

Para el caso del ejemplo, se obtiene:

$$\hat{A}_1 = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,04192 & -0,3462 \end{bmatrix}$$

La matriz de varianza covarianza estimada de la predicción, para el horizonte $h=1$, es:

$$\hat{\Sigma}_T(1) = \hat{\Sigma}_u = \begin{bmatrix} 0,0002 & 0,0001 & 0,0002 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0002 & 0,0001 & 0,0002 \end{bmatrix}$$

Así, la varianza del error de predicción para $\hat{Y}_{1:51}$ es igual a 0,0002, la varianza del error de predicción para $\hat{Y}_{2:51}$ es igual a 0,0001 y la varianza del error de predicción para $\hat{Y}_{3:51}$ es igual a 0,0002.

Los intervalos de confianza para $Y_{1:51}$, $Y_{2:51}$ y $Y_{3:51}$ vienen dados, respectivamente, por:

$$0,024 \pm 1,96 * \sqrt{0,0002} = [-0,007; 0,053]$$

$$0,022 \pm 1,96 * \sqrt{0,0001} = [0,007; 0,037]$$

$$0,025 \pm 1,96 * \sqrt{0,0002} = [-0,009; 0,051]$$

Para los horizontes $h = 2$, $h = 3$ se utilizan las siguientes fórmulas:

$$\hat{\Sigma}_T(2) = \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' = \begin{bmatrix} 0,0003 & 0,0001 & 0,0002 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0002 & 0,0001 & 0,0003 \end{bmatrix}$$

$$\hat{\Sigma}_T(3) = \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' + \hat{A}_1^2 \hat{\Sigma}_u \hat{A}_1^{2'} = \begin{bmatrix} 0,0005 & 0,0001 & 0,0004 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0004 & 0,0001 & 0,0005 \end{bmatrix}$$

Entonces los intervalos de confianza son:

$$IC(Y_{1:52}) = 0,024 \pm 1,96 * 0,016 = [-0,007; 0,056]$$

$$IC(Y_{2:52}) = 0,022 \pm 1,96 * 0,008 = [0,007; 0,037]$$

$$IC(Y_{3:52}) = 0,025 \pm 1,96 * 0,016 = [-0,006; 0,056]$$

$$IC(Y_{1:52}) = 0,024 \pm 1,96 * 0,023 = [-0,020; 0,068]$$

$$IC(Y_{1:52}) = 0,022 \pm 1,96 * 0,011 = [0,001; 0,043]$$

$$IC(Y_{1:52}) = 0,024 \pm 1,96 * 0,023 = [-0,020; 0,068]$$

\vdots

Comparación con modelos univariantes

La teoría VAR sugiere que las predicciones logradas son de mejor calidad que si se realiza la modelación de las series de manera univariante. Para comprobar esto, se realizó un modelo univariante para cada una de las series analizadas en este documento. Así, para las variaciones PIB se encontró el modelo $Y_{1t} = 0,94Y_{1t-5} + \hat{u}_t - 0,87\hat{u}_{1t-5}$; para la variación del CI se estimó el modelo $Y_{2t} = 0,97Y_{2t-7} + \hat{u}_t - 0,88\hat{u}_{2t-7}$ y para la variación de la DFI se obtuvo el modelo $Y_{3t} = 0,36Y_{3t-2} + 0,53Y_{3t-6} + \hat{u}_t$.

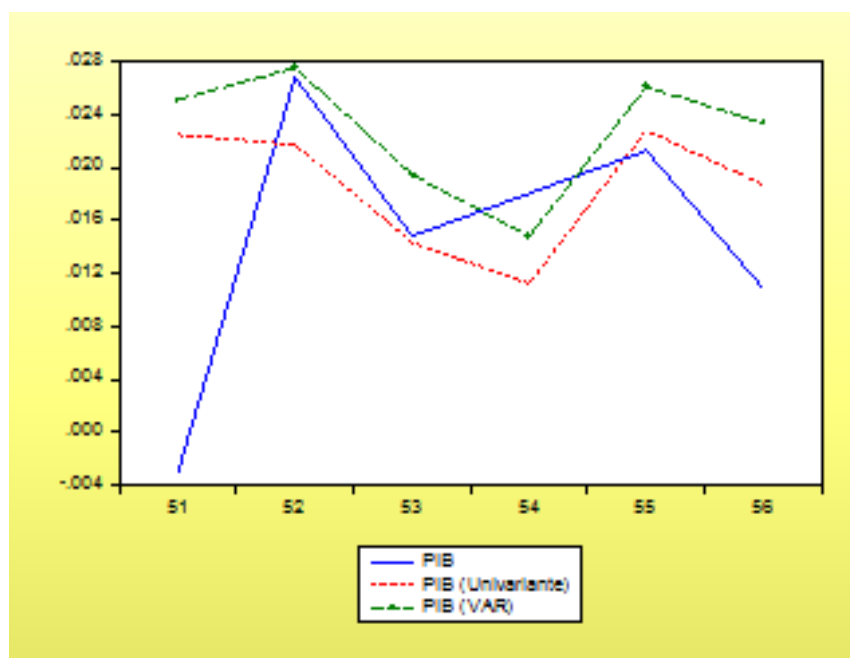


Figura 2.7: Comparación de las predicciones VAR y UNIVARIANTE para Y_{1t} (PIB)

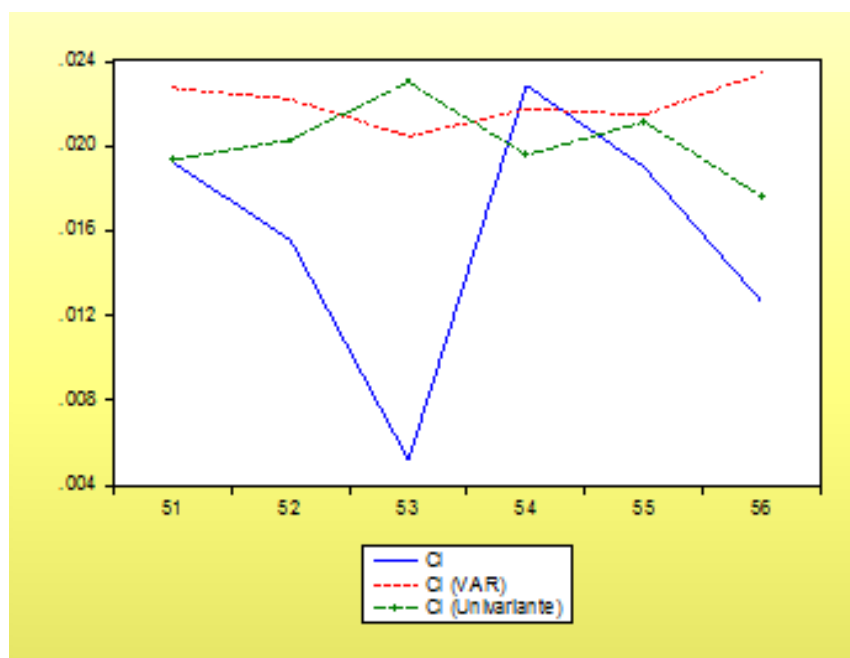


Figura 2.8: Comparación de las predicciones VAR y UNIVARIANTE para Y_{2t} (CI)

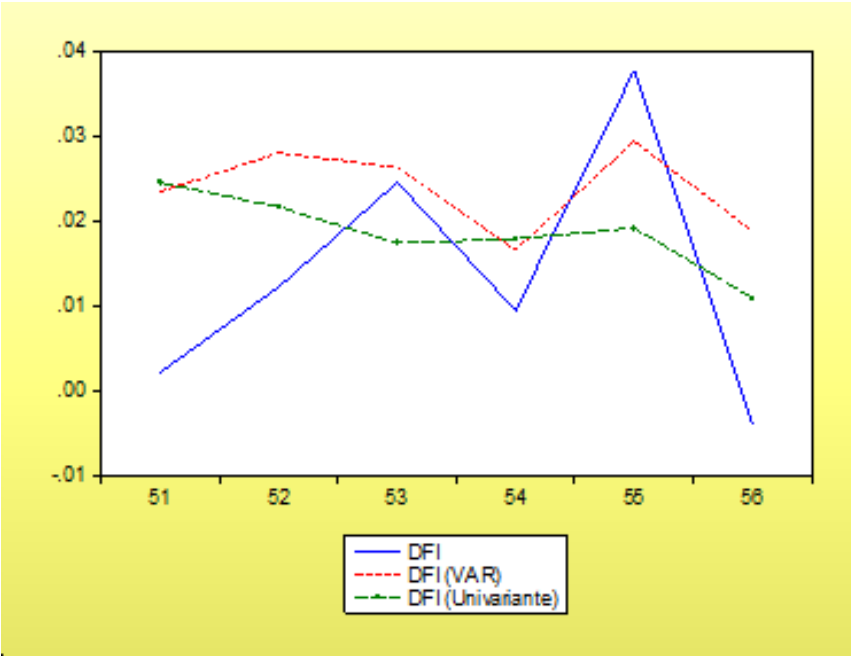


Figura 2.9: Comparación de las predicciones VAR y UNIVARIANTE para Y_{3t} (DFI)

Se puede observar que el ajuste que tienen las predicciones del modelo VAR; pero, para comparar analíticamente se calcula del error cuadrático medio estimado, para determinar el mejor ajuste. Así se tiene:

	PIB	CI	DFI
VAR	0,0005	0,0005	0,0006
UNIVARIANTE	0,0004	0,0004	0,0004

Tabla 2.9: Error Medio Cuadrático estimado para los modelos VAR y univariante

Como se puede observar, los errores son muy pequeños para ambos casos; la diferencia existente es estadísticamente no significativa entre los dos tipos de predicciones. Esto se explica porque las correlaciones cruzadas entre las series son muy poco significativas; sin embargo, el objetivo de esta presentación, es más bien, didáctico; posteriormente se mostrará un ejemplo más completo.

2.7 La Causalidad

En la teoría, la demostración de relaciones causales entre las variables de análisis proporciona los elementos de reflexión propicios para una mejor comprensión de los fenómenos, sobre todo los económicos. De manera práctica, “*the causal knowledge*” (“el conocimiento de la causalidad”) es necesario en la formulación correcta de la política económica. En efecto, saber la dirección de la causalidad es también importante en cuanto a poner un enlace entre las variables económicas.

2.7.1 La causalidad según Granger

Granger (1969) propuso los conceptos de causalidad y de exogeneidad: en el sentido de series de tiempo, se diría que la variable X_{2t} es la causa de X_{1t} , si la predicción de X_{1t} mejora si la información relativa a X_{2t} se incorpora al análisis (*el término predicción parece preferible en el marco de la causalidad; en efecto, decir que Y_t causa X_t , solo significa que es preferible para predecir X_t conocer Y_t , que no conocerla*).

Sea el modelo VAR(p) para el cual las variables X_{1t} y X_{2t} son estacionarias:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} a_1^1 & b_1^1 \\ a_1^2 & b_1^2 \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \end{bmatrix} + \begin{bmatrix} a_2^1 & b_2^1 \\ a_2^2 & b_2^2 \end{bmatrix} \begin{bmatrix} X_{1t-2} \\ X_{2t-2} \end{bmatrix} + \dots + \\ \dots + \begin{bmatrix} a_p^1 & b_p^1 \\ a_p^2 & b_p^2 \end{bmatrix} \begin{bmatrix} X_{1t-p} \\ X_{2t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

El bloque de variables ($X_{2t-1}, X_{2t-2}, \dots, X_{2t-p}$) se considera como exógeno en comparación del bloque de variables ($X_{1t-1}, X_{1t-2}, \dots, X_{1t-p}$) si el hecho de añadir el bloque X_{2t} no mejora significativamente la determinación de las variables X_{1t} . Se trata de efectuar una prueba de restricciones sobre los coeficientes de las variables X_{2t} de la representación VAR (Se denotará por RVAR al modelo VAR restringido). La determinación del retardo p se efectúa por los criterios de Akaike, Shwarz o Hanan-Quinn.

- X_{2t} no causa X_{1t} , si se acepta la siguiente hipótesis:

$$H_0 : b_1^1 = b_2^1 = \dots = b_p^1 = 0$$

- X_{1t} no causa X_{2t} , si se acepta la siguiente hipótesis:

$$H_0 : a_1^2 = a_2^2 = \dots = a_p^2 = 0$$

Si se llegan a aceptar las dos hipótesis: X_{1t} causa a X_{2t} y viceversa, se habla de efectos de retroalimentación ("*feedback effect*").

Estas pruebas pueden llevarse a cabo con la ayuda de la prueba clásica de Fisher de la nulidad de los coeficientes (prueba de Wald), ecuación por ecuación o bien directamente comparando un modelo VAR sin restricciones (UVAR) y el modelo VAR restringido (RVAR).

Se calcula el siguiente estadístico:

$$L^* = (T - c) (\ln |\Sigma_{RVAR}| - \ln |\Sigma_{UVAR}|) \sim \chi^2(2p)$$

donde:

Σ_{RVAR} = matriz de varianzas-covarianzas de los residuos del modelo restringido,

Σ_{UVAR} = matriz de varianzas-covarianzas de los residuos del modelo sin restricciones,

T = número de observaciones,

c = número de parámetros estimados de cada ecuación del modelo sin restricciones.

Si $L^* > \chi^2_{1-\alpha}(2p)$, entonces se rechaza la hipótesis de la validez de la restricción con un nivel de significación α .

También se suele utilizar la prueba tradicional de Fisher:

$$F^* = \frac{\frac{SCRR - SCRU}{c}}{\frac{SCRU}{n-k-1}}$$

SRU: es la suma de cuadrados de los residuos del modelo sin restricciones.

SRRR: es la suma de cuadrados de los residuos del modelo restringido.

C= número de restricciones (número de coeficientes que pone a prueba la hipótesis nula).

Si $F^* > F_{c;n-k-1}^\alpha$ se rechaza la hipótesis nula.

2.7.2 La causalidad según Sims

Sims (1980) presenta una especificación de prueba ligeramente diferente. Se considera que si los valores futuros de X_{1t} permiten explicar los valores presentes de X_{2t} , entonces X_{2t} es la causa de X_{1t} .

Esto se representa de la siguiente manera:

$$X_{1t} = a_1^0 + \sum_{i=1}^p a_{1i}^1 X_{1t-i} + \sum_{i=1}^p a_{1i}^2 X_{2t-i} + \sum_{i=1}^p b_i^2 X_{2t+i} + u_{1t}$$

$$X_{2t} = a_2^0 + \sum_{i=1}^p a_{2i}^1 X_{1t-i} + \sum_{i=1}^p a_{2i}^2 X_{2t-i} + \sum_{i=1}^p b_i^1 X_{1t+i} + u_{2t}$$

- X_{1t} no causa X_{2t} , si se aceptan la siguiente hipótesis:

$$H_0 : b_1^2 = b_2^2 = \dots = b_p^2 = 0$$

- X_{2t} no causa X_{1t} , si se aceptan la siguiente hipótesis:

$$H_0 : b_1^1 = b_2^1 = \dots = b_p^1 = 0$$

Se sigue utilizando la prueba clásica de Fisher de nulidad de coeficientes.

Observación 2.7. La prueba de Sims, se deja de como referencia dado que no se puede estimar directamente en EViews. Se puede crear un código de programación dentro del programa para poder realizar la estimación, pero esto está fuera del alcance de este documento.

Ejemplo 2.6. A partir de la representación VAR (1) estimada en el ejemplo 5.5, se va a proceder a realizar las pruebas de Granger.

Resolución.

Se procede con una prueba de Fisher, ecuación por ecuación.

Prueba de Granger

- H_0 : Y_{2t} y Y_{3t} no causan Y_{1t}

Se han estimado los siguientes modelos:

$$Y_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115 + \hat{u}_{1t}$$

$$R^2 = 0,09; \quad n = 48; \quad SCR U = 0,010381 \text{ (sin restricciones)}$$

$$Y_{1t} = -0,0197Y_{1t-1} + 0,025 + \hat{u}_{1t}$$

$$R^2 = 0,0004; \quad n = 48; \quad SCRR = 0,011474 \text{ (restringido)}$$

donde,

SCR U: es la suma de cuadrados de los residuos del modelo sin restricciones.

SCRR: es la suma de cuadrados de los residuos del modelo restringido.

$$F^* = \frac{\frac{SCR R - SCR U}{c}}{\frac{SCR U}{n-k-1}} = \frac{\frac{0,011474 - 0,010381}{1}}{\frac{0,010381}{48-3-1}} = 4,6326$$

c= número de restricciones (número de coeficientes que pone a prueba la hipótesis nula); en este caso c=1, dado que se elimina un coeficiente de cada ecuación.

$F^* > F_{1;48}^{0,05} \approx 4,05$; por tanto, se rechaza la hipótesis nula; Y_{2t} y Y_{3t} explica significativamente la variable Y_{1t} ; existe causalidad según Granger.

- Ho: Y_{1t} y Y_{3t} no causan Y_{2t}

Se han estimado los siguientes modelos:

$$Y_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197 + \hat{u}_{2t}$$

$$R^2 = 0,0026; \quad n = 48; \quad SCR U = 0,002594 \text{ (sin restricciones)}$$

$$Y_{2t} = 0,1014Y_{2t-1} + 0,0199 + \hat{u}_{2t}$$

$$R^2 = 0,011; \quad n = 48; \quad SCRR = 0,002638 \text{ (restringido)}$$

$$F^* = \frac{\frac{SCR R - SCR U}{c}}{\frac{SCR U}{n-k-1}} = \frac{\frac{0,002638 - 0,002594}{1}}{\frac{0,002594}{48-3-1}} = 0,7463$$

$F^* < F_{1;104}^{0,05} \approx 4,05$; por tanto, se acepta la hipótesis nula; Y_{1t} y Y_{3t} no explican significativamente la variable Y_{2t} ; no existe causalidad según Granger.

- Ho: Y_{1t} y Y_{2t} no causan Y_{3t}

Se han estimado los siguientes modelos:

$$Y_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185 + \hat{u}_{3t}$$

$$R^2 = 0,0110; \quad n = 48; \quad SCR U = 0,011036 \text{ (sin restricciones)}$$

$$Y_{3t} = -0,1394Y_{3t-1} + 0,0285 + \hat{u}_{3t}$$

$$R^2 = 0,011; \quad n = 48; \quad SCRR = 0,011487 \text{ (restringido)}$$

$$F^* = \frac{\frac{SCRR - SCRU}{c}}{\frac{SCRU}{n-k-1}} = \frac{\frac{0,011487 - 0,011036}{1}}{\frac{0,011036}{48-3-1}} = 1,7981$$

$F^* < F_{1;104}^{0,05} \approx 4,05$; por tanto, se acepta la hipótesis nula; Y_{1t} y Y_{2t} no explican significativamente la variable Y_{3t} ; no existe causalidad según Granger.

El siguiente gráfico muestra la salida del paquete EViews para la Prueba de Granger. Hay que considerar que el paquete considera la prueba de Wald y compara el valor obtenido con el estadístico Chi-cuadrado.

Dependent variable: Y_{1t}			
Excluded	Chi-sq	df	Prob.
Y_{2t}	3.972118	1	0.0463
Y_{3t}	0.007279	1	0.9320
All	4.631706	2	0.0987

Dependent variable: Y_{2t}			
Excluded	Chi-sq	df	Prob.
Y_{2t}	0.557515	1	0.4553
Y_{3t}	0.661477	1	0.4160
All	0.746494	2	0.6885

Dependent variable: Y_{3t}			
Excluded	Chi-sq	df	Prob.
Y_{2t}	0.236725	1	0.6266
Y_{3t}	1.415327	1	0.2342
All	1.799419	2	0.4067

Tabla 2.10: Prueba de causalidad de Granger

Como se puede ver, el p-valor (Prob.) es mayor que 0,05 en todos los casos, esto implica que no existe causalidad entre las variables. Sin embargo, al realizar el cálculo inicial se concluyó que si existe causalidad de Y_{3t} y Y_{2t} hacia Y_{1t} . La diferencia se origina en los algoritmos de cálculo que tienen los paquetes implementados.

2.8 Análisis de los “choques”

El análisis de los choques consiste en medir el impacto de la variación de una innovación sobre las variables.

Considérese el modelo estimado del ejemplo 5.5:

$$Y_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115 + \hat{u}_{1t}$$

$$Y_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197 + \hat{u}_{2t}$$

$$Y_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185 + \hat{u}_{3t}$$

Una variación en un instante dado de \hat{u}_{1t} tiene una consecuencia inmediata sobre Y_{1t} , y entonces sobre $Y_{1,t+1}$ y $Y_{2,t+1}$; por ejemplo, si se produce en t un choque sobre \hat{u}_{1t} igual a 1 y de orden 0 sobre \hat{u}_{2t} (las otras variables permanecen iguales que en el tiempo $(t-1)$; para los valores subsiguientes, u_{t+i} retorna a su valor u_{t-1}), se tiene los impactos siguientes:

En el período t :

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \\ \Delta Y_{3t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

En el período $t+1$:

$$\begin{bmatrix} \Delta \hat{Y}_{1,t+1} \\ \Delta \hat{Y}_{2,t+1} \\ \Delta \hat{Y}_{3,t+1} \end{bmatrix} = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,4192 & -0,3462 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0,1170 \\ -0,0575 \\ 0,1972 \end{bmatrix}$$

En el período $t+2$:

$$\begin{bmatrix} \Delta \hat{Y}_{1,t+2} \\ \Delta \hat{Y}_{2,t+2} \\ \Delta \hat{Y}_{3,t+2} \end{bmatrix} = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,4192 & -0,3462 \end{bmatrix} \begin{bmatrix} -0,1170 \\ -0,0575 \\ 0,1972 \end{bmatrix} = \begin{bmatrix} -0,0315 \\ 0,0206 \\ -0,1154 \end{bmatrix}$$

Por otro lado, utilizando la representación lineal del VAR, se logra realizar el análisis de las funciones de impulso-respuesta, ya que un choque en Y_{1t} se reflejará como la primera columna de $M_i = A_1^i$ (recuerde que $M_0 = I$); de forma similar, un impacto sobre Y_{2t} se reflejará en la segunda columna de la matriz M_i . La generalización a un VAR con k variables, es inmediata.

Considere el modelo estimado en el ejemplo 5.5:

$$\hat{A}_1 = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,4192 & -0,3462 \end{bmatrix}$$

$$\hat{A}_1^2 = \begin{bmatrix} -0,0315 & -0,0437 & 0,0737 \\ 0,0206 & 0,0011 & -0,0240 \\ -0,1154 & 0,0245 & 0,1567 \end{bmatrix}$$

Así, los elementos de \hat{A}_1^i representan los efectos de los choques unitarios de las variables del sistema luego de i periodos. Es por esta razón que, se les conoce como **multiplicadores dinámicos o respuestas al impulso**.

En las tablas siguientes se muestran los resultados de las funciones de impulso-respuesta que presenta el paquete EViews para las variables Y_{1t} , Y_{2t} y Y_{3t} , que son los cálculos hechos manualmente al inicio de esta sección. Esto también se muestra en los gráficos siguientes:

Período	Y_{1t}	Y_{2t}	Y_{3t}
1	1,000000	0,000000	0,000000
2	-0,193069	-0,080383	0,108038
3	-0,016093	0,019189	-0,090439
4	0,014628	-0,005392	0,036103
5	-0,005860	0,001663	-0,012563
6	0,002041	-0,000533	0,004198
7	-0,000682	0,000173	-0,001384
8	0,000225	-5,66E-05	0,000454
9	-7,38E-05	1,85E-05	-0,000149
10	2,42E-05	-6,06E-06	4,88E-05

Tabla 2.11: Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{1t}

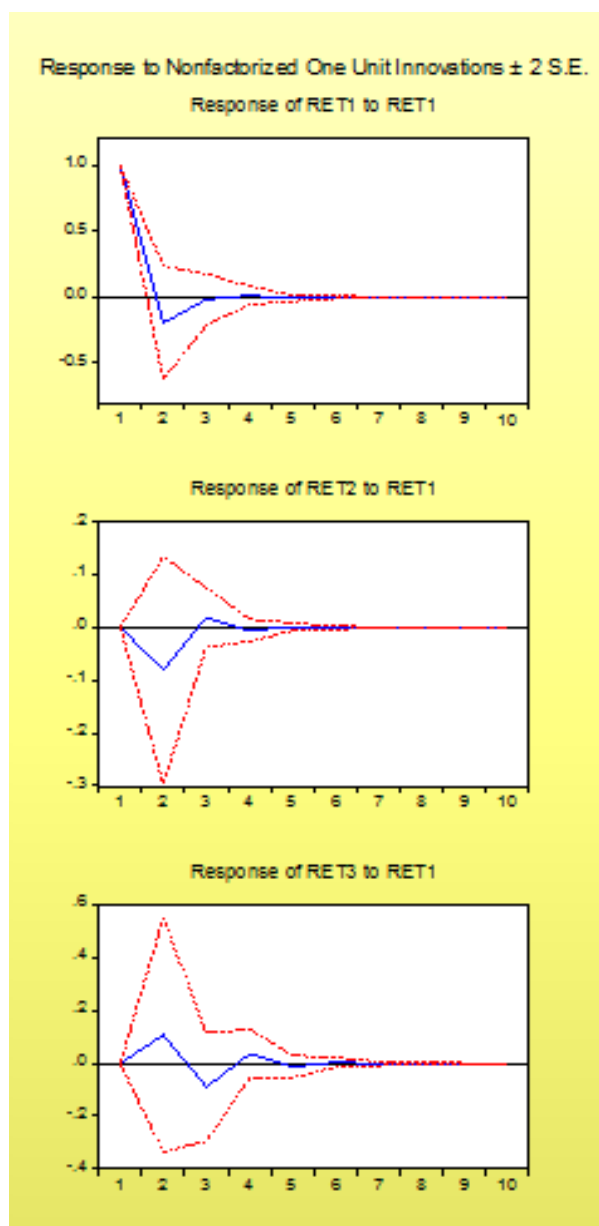


Figura 2.10: Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{1t}

Período	Y_{1t}	Y_{2t}	Y_{3t}
1	0.000000	1.000000	0.000000
2	0.689275	0.075128	0.424214
3	-0.073293	-0.011642	-0.032978
4	0.005504	0.002053	-0.002027
5	0.000315	-0.000470	0.002131
6	-0.000345	0.000131	-0.000865
7	0.000140	-4.02E-05	0.000303
8	-4.91E-05	1.29E-05	-0.000101
9	1.65E-05	-4.18E-06	3.34E-05
10	-5.43E-06	1.36E-06	-1.10E-05

Tabla 2.12: Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{2t}

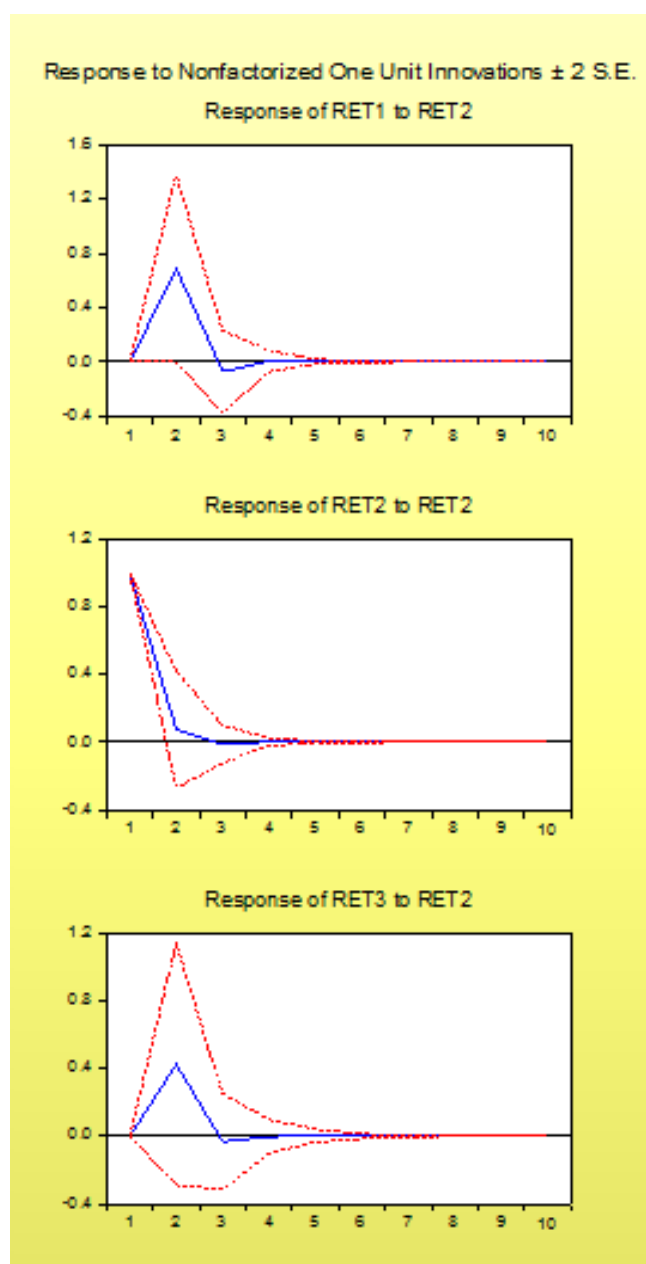
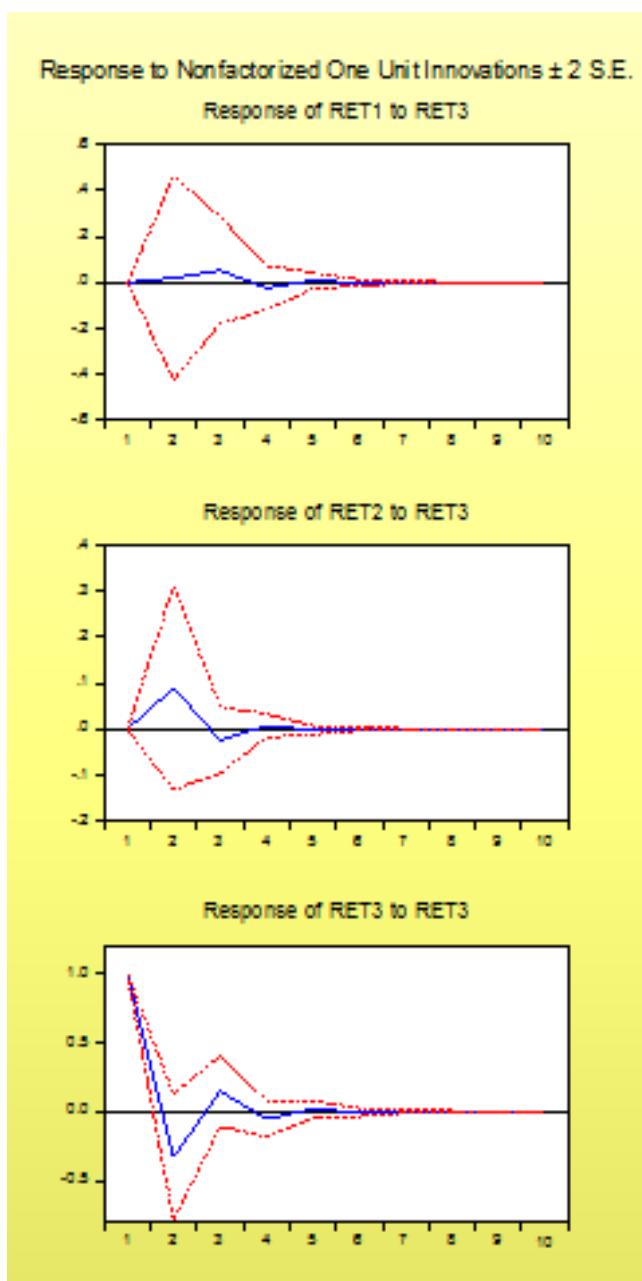


Figura 2.11: Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{2t} .

Período	Y_{1t}	Y_{2t}	Y_{3t}
1	0.000000	0.000000	1.000000
2	0.018858	0.089860	-0.328412
3	0.052104	-0.024276	0.148011
4	-0.024001	0.007288	-0.053278
5	0.008653	-0.002311	0.017996
6	-0.002924	0.000748	-0.005955
7	0.000968	-0.000244	0.001957
8	-0.000318	7.98E-05	-0.000642
9	0.000104	-2.61E-05	0.000210
10	-3.42E-05	8.55E-06	-6.88E-05

Tabla 2.13: Respuesta de las variables Y_{1t} , Y_{2t} y Y_{3t} ante un choque unitario de Y_{3t}



La elección de la dirección del impacto es muy importante y determina los valores obtenidos. Se puede observar que el efecto de la innovación se desvanece con el tiempo; esto caracteriza a un proceso VAR estacionario.

- i. Si las variables están medidas en escalas diferentes es común considerar las innovaciones iguales a su desviación típica en lugar de los choques unitarios; es decir:

$$\hat{u}_{t,0} = \sqrt{\text{Var}(\hat{u}_t)}$$

Esto constituye únicamente un reescalamiento de las funciones de impulso-respuesta.

- Si una variables no causa (en el sentido de Granger) al resto de variables en el sistema, entonces las respuestas al impulso sobre las otras variables serán cero.

2.9 Descomposición de la varianza

2.9.1 Representación de errores ortogonales

Un problema a considerar en el análisis de las funciones de impulso-respuesta es el de la correlación contemporánea de errores y, por lo tanto, el impacto de un choque sobre una variable puede acompañarse de un impacto en otra variable; ignorarla, puede distorsionar la verdadera relación dinámica entre las variables. Es por esto que se trata de manera general de realizar el análisis a través de la búsqueda de una representación de errores ortogonales.

Considerando la representación lineal del VAR, se puede obtener lo siguiente:

1. Dado que Σ_u es simétrica y definida positiva, entonces existe P no singular tal que:

$$\Sigma_u = PP'$$

donde, P es una matriz triangular obtenida a través de la descomposición de Cholesky.

2. Luego, se puede expresar lo siguiente:

$$v_t = Pu_t$$

donde, v_t es un vector aleatorio con $E(v_t) = 0$ y $V(v_t) = I$.

Entonces, v_t son las *innovaciones ortogonales* de u_t . Luego, reemplazando:

$$X_t = \sum_{i=1}^{\infty} \theta_i Pu_{t-i} = \sum_{i=1}^{\infty} M_i u_{t-i}, \quad M_i \equiv \theta_i P$$

Entonces, $\theta_0 = P$; por lo que, salvo el caso que Σ_u sea diagonal, Σ_v no será diagonal y sus elementos recogerán las respuestas inmediatas del sistema de choques unitarios. Es por esto que se los conoce como **multiplicadores de impacto**. Además, el hecho que $\theta_0 = P$ sea una matriz triangular, implica que el orden de las variables en el vector es importante (órdenes diferentes de descomposición pueden producir funciones de impluso respuesta diferentes).

2.9.2 Descomposición de la Varianza

La descomposición de la varianza del error de predicción tiene como objetivo calcular para cada una de las innovaciones su contribución a la varianza del error. Por la técnica matemática de la descomposición de Cholesky de la matriz Σ_u que es simétrica y definida positiva, se puede escribir la varianza del error de predicción para un horizonte h en función de la varianza del error atribuida a cada una de las variables; es suficiente dividir cada una de estas variaciones por la varianza total, para obtener su peso relativo en porcentaje.

Se retoma el modelo VAR (1) de dos variables X_{1t} y X_{2t} . La varianza del error de predicción para $X_{1,t+h}$ se puede escribir:

$$\sigma_{x_h}^2 = \sigma_{u_1}^2 \left[m_{11}^2(0) + \dots + m_{11}^2(h-1) \right] + \sigma_{u_2}^2 \left[m_{22}^2(0) + \dots + m_{22}^2(h-1) \right]$$

Donde los $m_{ii}^{(j)}$ son los términos de las matrices M_i (representación lineal del proceso).

Al horizonte h , la descomposición de la varianza, en porcentaje, de las innovaciones propias de X_{1t} sobre X_{1t} , están dadas por:

$$\frac{\sigma_{u_1}^2 \left[m_{11}^2(0) + \dots + m_{11}^2(h-1) \right]}{\sigma_{X_1}^2(h)} * 100$$

Y la descomposición de la varianza, en porcentaje, de X_{2t} sobre X_{1t}

$$\frac{\sigma_{u_2}^2 \left[m_{22}^2(0) + \dots + m_{22}^2(h-1) \right]}{\sigma_{X_1}^2(h)} * 100$$

La interpretación de los resultados es importante:

- Si un choque sobre u_{1t} no afecta la varianza del error de X_{2t} independientemente del horizonte de predicción, entonces X_{2t} puede considerarse como exógena porque X_{2t} evoluciona independientemente de u_{1t} .
- En caso contrario, si un choque sobre u_{1t} afecta fuertemente, en realidad totalmente, la varianza del error de X_{2t} , entonces X_{2t} se considera como endógena.

2.9.3 Elección del orden de descomposición

Nótese que el problema de la correlación contemporánea de los errores y, por lo tanto, el impacto de un choque sobre una variable, implica una elección de descomposición que proporciona resultados asimétricos en función del orden de las variables. El problema es más complejo si el número de las variables es importante.

Ejemplo 2.7. Considérese un modelo VAR con 4 variables X_1, X_2, X_3 y X_4 .

Supóngase que se elige el siguiente orden para la descomposición de Cholesky: $X_2 X_3 X_1 X_4$ y esto lleva a obtener la siguiente tabla hipotética:

Ordenación de Cholesky: $X_2 X_3 X_1 X_4$				
Respuesta de X_2				
Período	X_1	X_2	X_3	X_4
1	0,000000	4,583291	0,000000	0,000000
2	0,775767	-0,251545	0,815017	-0,905811
Respuesta de X_3				
Período	X_1	X_2	X_3	X_4
1	0,000000	2,439203	4,54469	0,000000
2	-0,441257	-0,754324	-3,564595	-0,566602
Respuesta de X_1				
Período	X_1	X_2	X_3	X_4
1	4,603662	3,022459	1,802923	0,000000
2	-0,795786	-0,101091	0,388095	-0,255476
Respuesta de X_4				
Período	X_1	X_2	X_3	X_4
1	1,568803	0,486351	1,415711	3,191513
2	-0,329818	-4,104756	-0,527459	-3,186928

Tabla 2.14: Descomposición hipotética de choques

La interpretación de la tabla 5.14 se haría de la siguiente manera:

- Un choque para el período 1 sobre X_2 tiene un impacto solamente sobre X_2 y ausencia de correlación contemporánea con X_3 , X_1 y X_4 .
- Un choque para el período 1 sobre X_3 tiene un impacto sobre X_2 y X_3 y ausencia de correlación contemporánea con X_1 y X_4 .
- Un choque para el período 1 sobre X_1 tiene un impacto sobre X_2 , X_3 y X_1 y ausencia de correlación contemporánea con X_4 .
- Finalmente, un choque para el período 1 sobre X_4 tiene un impacto sobre todas las variables.

Ahora, si se realiza el análisis considerando un orden diferente de las variables para la descomposición, se tendrá:

Ordenación de Cholesky: $X_3X_4X_2X_1$				
Respuesta de X_3				
Período	X_1	X_2	X_3	X_4
1	0,000000	0,000000	2,805461	0,000000
2	-0,527459	-3,186928	0,815017	-0,905811
Respuesta de X_4				
Período	X_1	X_2	X_3	X_4
1	0,000000	0,000000	-2,231567	1,256043
2	-0,441257	-0,754324	0,388095	-0,255476
Respuesta de X_2				
Período	X_1	X_2	X_3	X_4
1	0,000000	-0,022459	1,256923	-2,256123
2	0,815017	-0,905811	0,388095	-0,255476
Respuesta de X_1				
Período	X_1	X_2	X_3	X_4
1	1,457895	-2,145627	1,711415	3,564281
2	-0,441257	-0,754324	-0,524784	-3,968741

Tabla 2.15: Otra descomposición hipotética de choques

La interpretación de la tabla 5.15, se realiza de manera similar a la tabla 5.10.

El orden de descomposición se debería efectuar desde la variable que se supone es más exógena hasta la variable menos exógena. En caso de duda, es necesario realizar diferentes combinaciones del orden de descomposición y analizar la robustez de los resultados.

Ejemplo 2.8. A partir de la representación VAR(1) estimada en el ejemplo 5.5, calcular e interpretar las funciones de impulso-respuesta ortogonales, y la descomposición de la varianza.

Resolución.

Dado que las variables del ejemplo son variaciones del PIB, el CI y el DFI, es lógico pensar que un choque sobre la variable variación del PIB influya en la variación del CI y el DFI, más que si el choque fuera al revés. Esto sería: una innovación sobre Y_{1t} (variación del PIB) influye de manera instantánea a Y_{2t} (variación del CI) y a Y_{3t} (variación del DFI); por otro lado, una innovación sobre Y_{2t} o Y_{3t} no influye de manera contemporánea a Y_{1t} .

La matriz de varianza-covarianza estimada de los residuos es igual (ejemplo 5.5) a:

$$\hat{\Sigma}_u = \begin{bmatrix} 0,0002 & 0,0001 & 0,0002 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0002 & 0,0001 & 0,0002 \end{bmatrix}$$

Con el programa EViews, las salidas de las funciones de impulso-respuesta y la des-

composición de varianza serían:

Período	Variance Decomposition of Y_1 :			
	S.E.	Y_1	Y_2	Y_3
1	0,015360	43,03994	21,09375	35,86631
2	0,016092	40,67934	25,73053	33,59013
3	0,016102	40,63720	25,71162	33,65118
4	0,016104	40,63653	25,70823	33,65524
5	0,016104	40,63654	25,70806	33,65540
6	0,016104	40,63655	25,70805	33,65540
7	0,016104	40,63655	25,70805	33,65540
8	0,016104	40,63655	25,70805	33,65540
9	0,016104	40,63655	25,70805	33,65540
10	0,016104	40,63655	25,70805	33,65540

Período	Variance Decomposition of Y_2 :			
	S.E.	Y_1	Y_2	Y_3
1	0,007678	0,000000	100,0000	0,000000
2	0,007778	1,084622	98,63426	0,281119
3	0,007784	1,144767	98,54459	0,310640
4	0,007784	1,149508	98,53672	0,313774
5	0,007784	1,149959	98,53593	0,314109
6	0,007784	1,150005	98,53585	0,314145
7	0,007784	1,150010	98,53584	0,314149
8	0,007784	1,150010	98,53584	0,314149
9	0,007784	1,150011	98,53584	0,314149
10	0,007784	1,150011	98,53584	0,314149

Período	Variance Decomposition of Y_3 :			
	S.E.	Y_1	Y_2	Y_3
1	0,015837	0,000000	34,49091	65,50909
2	0,016226	0,450218	33,21246	66,33733
3	0,016293	0,759358	33,02585	66,21479
4	0,016303	0,808238	33,01068	66,18109
5	0,016304	0,814151	33,00932	66,17653
6	0,016305	0,814812	33,00919	66,17600
7	0,016305	0,814883	33,00917	66,17594
8	0,016305	0,814891	33,00917	66,17594
9	0,016305	0,814892	33,00917	66,17594
10	0,016305	0,814892	33,00917	66,17594
Cholesky Ordering: Y_2 Y_2 Y_1				

Tabla 2.16: Descomposición de la Varianza (Orden de Cholesky Y_3 Y_2 Y_1)

La descomposición de la varianza indica que la varianza del error de predicción de Y_{1t} representa un 40,64 % con sus propias innovaciones, un 25,71 % con las de Y_{2t} y un 33,66 % con las de Y_{3t} . La varianza del error de predicción de Y_{2t} es de un 1,15 % con Y_{1t} , un 98,54 % con Y_{2t} y un 0,31 % con Y_{3t} . Finalmente, la varianza del error de predicción de Y_{3t} es de 0,81 % con Y_{1t} , un 33,01 % con Y_{2t} y un 66,18 % con sus propias innovaciones. Este efecto de asimetría se estudió en la parte de la causalidad (ejemplo 5.5); lo que también se corrobora ahora.

Por último, cabe señalar que la tabla anterior muestra la desviación estándar del error de previsión para Y_{1t} , Y_{2t} y Y_{3t} , que se calculó de manera tediosa en el ejemplo 5.5.

Observación 2.8. La modelización VAR se realiza siempre sobre series estacionarias; sin embargo, el método se puede aplicar a series que se las vuelve estacionarias a través de diferenciación, con la idea de que se debe recuperar la serie original una vez calculadas las predicciones (esto se puede hacer dado que la diferenciación es una transformación lineal). En algunos paquetes econométricos como EViews o Stata, entre otros, el programa da la opción automática de recuperar la serie original de los datos.

2.10 Ejemplo Práctico

Se consideran dos series de datos del Ecuador: las variaciones del índice de precios al productor (IPP), denotada por (X_{1t}) y del índice de actividad económica (IAE), denotada por (X_{2t}). Se dispone de 132 datos mensuales desde enero de 2004 hasta junio de 2015 (Ver Anexo D.2). Para efectos de comparaciones se trabajará únicamente con los datos hasta diciembre de 2014 y se guardarán los del año 2015. Los datos se tomaron de la página oficial del INEC. Se desea estimar un modelo para realizar las predicciones de ambas series.

Resolución.

1. Matrices de correlación cruzada

a) Estadísticos descriptivos de X_{1t} y X_{2t}

	Media	Mediana	Máximo	Mínimo	Desv. Est.	Asimetría	Curtosis
X_{1t}	0,64	1,16	15,17	-14,86	5,02	-0,40	3,74
X_{2t}	0,07	-0,30	15,18	-11,45	4,42	0,51	3,70

b) Matrices de correlación cruzada

	retardo 1		retardo 2		retardo 3		retardo 4		retardo 5	
X_{1t}	0,17	0,12	0,08	-0,02	0,01	0,08	-0,12	-0,03	-0,03	-0,09
X_{2t}	0,21	-0,20	0,05	0,08	0,14	0,02	0,13	-0,09	0,01	0,08

c) Representación simplificada

$$\begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline + & - \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array}$$

Tabla 2.17: Resumen de estadísticas y matrices de correlación cruzada para X_{1t} y X_{2t} .

Es fácil ver que las correlaciones cruzadas son significativas al 5 % en el retardo 1 ($2/\sqrt{T} = 0,1747$ en este caso). Así, X_{2t} depende de los valores en el primer retardo de X_{1t} en el primer retardo y del suyo propio. En EViews, se presenta la siguiente salida:

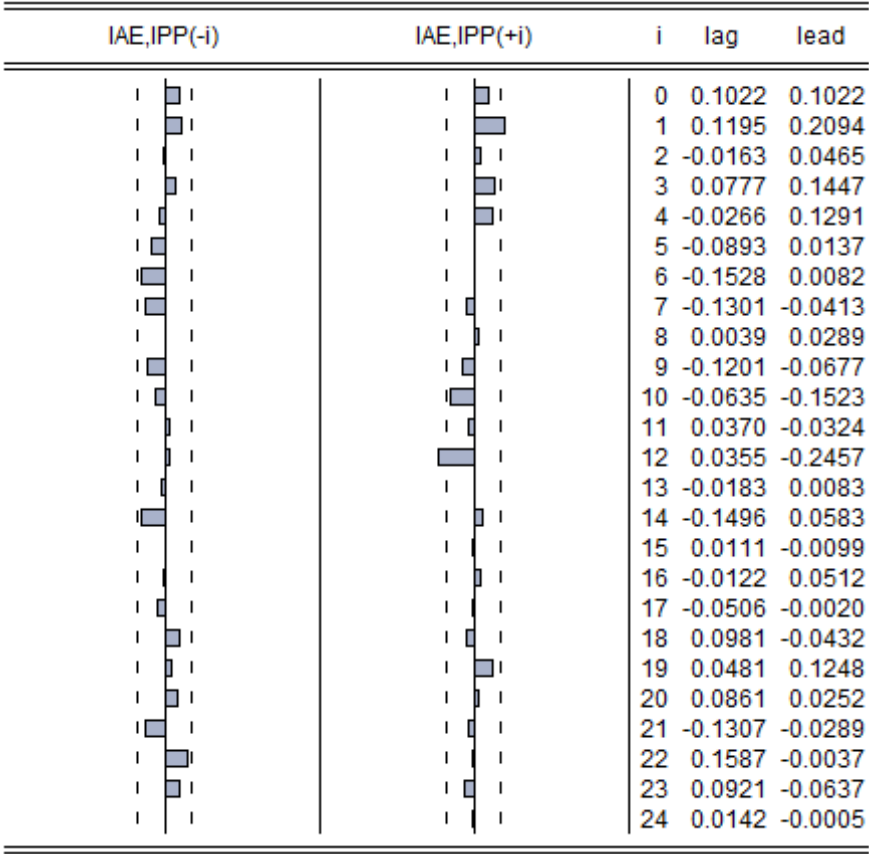


Figura 2.12: Correlaciones cruzadas entre X_{1t} y X_{2t}

2. Estimación y validación del VAR

Se inicia presentando el gráfico de las series:

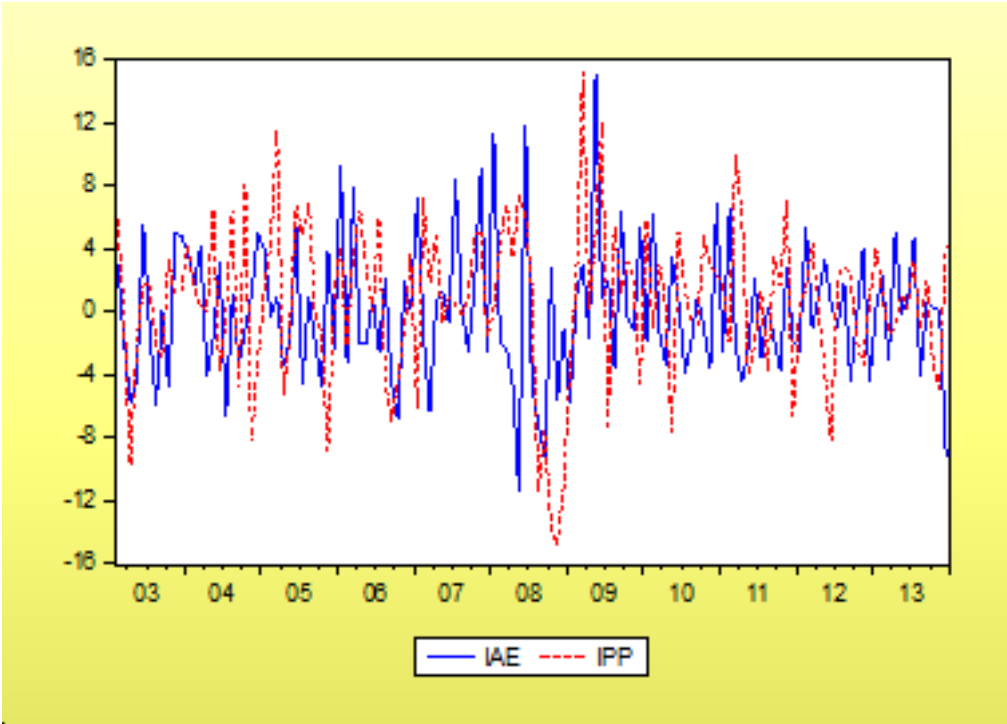


Figura 2.13: Gráfico de secuencia de las series

Antes de realizar los procedimientos para estimar los modelos, se debe verificar si las series a ser analizadas son estacionarias; para ello se realiza la prueba de raíces unitarias para cada serie utilizando el programa EViews:

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.001098	0.0000
Test critical values:	1 % level	-4.037668	
	5 % level	-3.448348	
	10 % level	-3.149326	

*MacKinnon (1996) one-sided p-values.

Tabla 2.18: Prueba DFA para X_{1t}

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.00154	0.0000
Test critical values:	1 % level	-4.037668	
	5 % level	-3.448348	
	10 % level	-3.149326	

*MacKinnon (1996) one-sided p-values.

Tabla 2.19: Prueba DFA para X_{2t}

Como se puede ver en las figuras 5.18 y 5.19, las dos series son estacionarias.

- a) Se utilizarán los criterios de Akaike, Schwarz y el logaritmo de máxima verosimilitud para determinar el retardo p entre 1 y 4. Se deben estimar cuatro modelos diferentes y retener aquel que satisfaga la mayor cantidad de criterios óptimos.

Se obtiene lo siguiente:

$$\hat{X}_{1t} = 0,1479X_{1t-1} + 0,2268X_{2t-1} + 0,4928$$

$$\hat{X}_{2t} = 0,1274X_{1t-1} - 0,2187X_{2t-1} - 0,0772$$

Con la ayuda del paquete EViews 7, se realiza la estimación de los 4 modelos. Así, se obtuvieron los siguientes resultados:

Criterio / Retardo	1	2	3	4
Log likelihood	-761,42	-756,05	-746,56	-736,01
Akaike	11,81	11,87	11,88	11,87
Schwarz	11,93	12,10	12,20	12,28

Tabla 2.20: Criterios para escoger el retardo del VAR

Como se puede observar en la tabla 5.18, es en el retardo 1 ($p = 1$) donde los criterios de Akaike y Schwarz se minimizan aunque el valor del log de verosimilitud es mínimo (en la práctica, muy pocas veces se utiliza este criterio como decisivo para escoger el retardo del VAR). Por lo tanto se realiza la estimación del VAR(1).

- i. El modelo VAR estimado se escribe:

$$X_{1t} = 0,1479X_{1t-1} + 0,2268X_{2t-1} + 0,4928 + \hat{u}_{1t}$$

$$(1,71) \quad (2,28) \quad (1,14)$$

$R^2 = 0,07$; $n = 130$; $(.)$ = estadístico correspondiente a la distribución t de Student.

$$X_{2t} = 0,1274X_{1t-1} - 0,2187X_{2t-1} - 0,0772 + \hat{u}_{2t}$$

$$(1,67) \quad (-2,49) \quad (-0,20)$$

$R^2 = 0,06$; $n = 130$; $(.)$ = estadístico correspondiente a la distribución t de Student.

Antes de realizar las predicciones, se debe verificar si el modelo cumple con el criterio de estabilidad. Con ayuda del programa EViews se obtiene:

Raíces	Módulo
-0,285373	0,285373
0,214590	0,214590

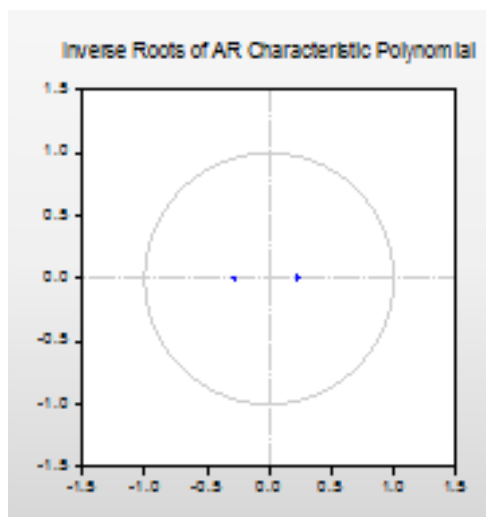


Figura 2.14: Criterio de estabilidad para el VAR(1) estimado

Analítica y gráficamente, se concluye que las inversas de las raíces del polinomio característico se encuentran dentro del círculo unidad; por lo tanto, se concluye que el modelo es estable y, por tanto, es estacionario.

- ii. Ahora, se necesita verificar que los residuos del modelo sean ruidos blancos; en general, se prueba la independencia. Para ello, se utilizará el paquete EViews para obtener las pruebas sobre los residuos que se describieron anteriormente. Así se obtiene:

Prueba Portmanteau

Lags	Q-Stat	Prob	Adj Q-Stat	Prob	df
1	0,050722	NA*	0,051115	NA*	NA*
2	1,429068	0,9641	1,450997	0,9627	6
3	6,825196	0,7418	6,974593	0,7278	10
4	14,04034	0,4467	14,41879	0,4190	14
5	15,45813	0,6303	15,89329	0,6000	18
6	24,47227	0,3230	25,34360	0,2809	22
7	26,60556	0,4302	27,59829	0,3785	26
8	34,51039	0,2609	36,02148	0,2074	30
9	37,10211	0,3279	38,80597	0,2619	34
10	51,10065	0,0760	53,97106	0,0447	38
11	53,74575	0,1057	56,86065	0,0627	42
12	68,78477	0,0164	73,42907	0,0062	46
13	72,41862	0,0207	77,46668	0,0076	50
14	76,83463	0,0223	82,41566	0,0077	54
15	77,52667	0,0443	83,19797	0,0167	58
16	80,19277	0,0599	86,23825	0,0226	62
17	82,43977	0,0832	88,82330	0,0321	66
18	86,47531	0,0883	93,50741	0,0318	70
19	90,20916	0,0968	97,88039	0,0330	74
20	93,29485	0,1141	101,5271	0,0380	78

Tabla 2.21: Prueba de autocorrelación Pormanteau

En la tabla 5.21 se observa que los p-valores (Prob.) para los retardos de 1 al 9 son no significativos; sin embargo, a partir del retardo 10 se vuelven significativos, esto sugiere que los residuos están autocorrelacionados.

Prueba LM

Lags	LM-Stat	Prob
1	0,600602	0,9630
2	1,492304	0,8280
3	5,496276	0,2401
4	7,679129	0,1041
5	1,468612	0,8322
6	9,278191	0,0545
7	2,213578	0,6965
8	8,222604	0,0838
9	2,633746	0,6209
10	14,83753	0,0051
11	2,909367	0,5731
12	16,83962	0,0021
13	4,069454	0,3967
14	4,565627	0,3348
15	0,731412	0,9474
16	2,790990	0,5934
17	2,346395	0,6723
18	4,123520	0,3895
19	3,978227	0,4090
20	3,335910	0,5033

Tabla 2.22: Prueba LM

Por los valores en la columna Prob. (ver tabla 5.20), se puede concluir que existe autocorrelación entre los residuos (retardos 10 y 12). Esto confirma que se hay que reformular el modelo planteado.

Reformulación del modelo

Se puede ver en las figuras 5.3 y 5.4 que a partir del retardo 10 existe autocorrelación de los residuos, por lo que se prueba un nuevo modelo VAR (10) para corregir este inconveniente.

Se estimó un VAR(10); sin embargo, al retardo 12 y 6 se tenía autocorrelación de residuos. Luego, se agregaron los retardos de orden 6 al VAR pero se encontró autocorrelación al retardo 12. Finalmente, se agregó el retardo 12 y se consiguió que los residuos no estén correlacionados, pero el problema era que no seguían una distribución normal multivariante.

Luego, se procedió a agregar un retardo de orden 14 y a quitar el retardo 12; con esto se consiguió que los residuos sean ruidos blancos, aunque presenta algo de correlación en el retardo 12, se decidió conservar este modelo ya que es el que cumple con más pruebas de independencia de residuos. Se debe mencionar que solamente se retienen los coeficientes significativos. Así, se obtuvo el siguiente modelo:

$$X_{1t} = 0,1416X_{1t-1} - 0,2403X_{1t-6} + 0,0889X_{1t-10} - 0,1047X_{1t-14} + 0,2241X_{2t-1} \\ + 0,0326X_{2t-6} - 0,1393X_{2t-10} + 0,1082X_{2t-14} + 0,7716 + \hat{u}_{1t}$$

$$X_{2t} = 0,0703X_{1t-1} - 0,1787X_{1t-6} - 0,1133X_{1t-10} - 0,1893X_{1t-14} - 0,2875X_{2t-1} \\ - 0,0605X_{2t-6} - 0,2330X_{2t-10} - 0,0422X_{2t-14} + 0,3525 + \hat{u}_{2t}$$

iii. La predicción calculada por el modelo, de manera recurrente es:

$$\hat{X}_{1t} = 0,1416X_{1t-1} - 0,2403X_{1t-6} + 0,0889X_{1t-10} - 0,1047X_{1t-14} + 0,2241X_{2t-1} \\ + 0,0326X_{2t-6} - 0,1393X_{2t-10} + 0,1082X_{2t-14} + 0,7716$$

$$\hat{X}_{1,15:1} = 0,1416 * 4,15 - 0,2403 * (3,26) + 0,0889 * (-1,34) - 0,1047 * (-3,37) \\ + 0,2241 * (-9,23) + 0,0326 * (4,71) - 0,1393 * (-3,12) + 0,1082 * (4,03) \\ + 0,7716$$

$$\hat{X}_{1,15:1} = -0,235$$

$$\hat{X}_{2t} = 0,0703X_{1t-1} - 0,1787X_{1t-6} - 0,1133X_{1t-10} - 0,1893X_{1t-14} - 0,2875X_{2t-1} \\ - 0,0605X_{2t-6} - 0,2330X_{2t-10} - 0,0422X_{2t-14} + 0,3525$$

$$\hat{X}_{2,15:1} = 0,0703 * 4,15 - 0,1787 * (-3,26) - 0,1133 * (-1,34) - 0,1893 * (-3,37) \\ - 0,2875 * (-9,23) - 0,0605 * (4,71) - 0,2330 * (-3,12) - 0,0422 * (4,03) \\ + 0,3525$$

$$\hat{X}_{2,15:1} = 3,777$$

donde, $\hat{X}_{i,15:j}$ significa, la previsión de la variable X_i para el mes j del año 2015 (15:j).

De la misma manera se obtiene:

$$\hat{X}_{1,15:2} = 0,400$$

$$\hat{X}_{2,15:2} = -1,281$$

$$\hat{X}_{1,15:3} = -0,221$$

$$\hat{X}_{2,15:3} = -0,536$$

Para calcular la varianza del error de predicción, si fuera un VAR (1) se tiene:

$$M_1 = \hat{A}_1; \quad M_2 = \hat{A}_1 M_1 = \hat{A}_1^2; \quad \text{etc} \dots$$

Dado que el modelo es un VAR (14) con coeficientes 1, 6, 10 y 14, se tendrán las matrices \hat{A}_1 , \hat{A}_6 , \hat{A}_{10} , \hat{A}_{14} y estarán compuestas por los coeficientes de los retardos de las variables analizadas; así, se obtiene:

$$\hat{A}_1 = \begin{bmatrix} 0,1416 & 0,2241 \\ 0,0703 & -0,2871 \end{bmatrix}; \quad \hat{A}_6 = \begin{bmatrix} -0,2403 & 0,0326 \\ -0,1787 & -0,0605 \end{bmatrix}$$

$$\hat{A}_{10} = \begin{bmatrix} 0,0889 & -0,1393 \\ -0,1133 & -0,2330 \end{bmatrix}; \quad \hat{A}_{14} = \begin{bmatrix} -0,1047 & 0,1082 \\ -0,1893 & -0,0422 \end{bmatrix}$$

La matriz de varianza covarianza estimada de la predicción, para el horizonte $h = 1$, es:

$$\hat{\Sigma}_T(1) = \hat{\Sigma}_u = \begin{bmatrix} 23,821 & 0,422 \\ 0,422 & 16,602 \end{bmatrix}$$

Así, la varianza del error de predicción para $\hat{X}_{1,15:1}$ es igual a 23,821 y la varianza del error de predicción para $\hat{X}_{2,15:1}$ es igual a 16,602.

Los intervalos de confianza para $X_{1,15:1}$ y $X_{2,15:1}$ vienen dados, respectivamente, por:

$$-0,235 \pm 1,96 * \sqrt{23,821} = [-9,80; 9,33]$$

$$3,777 \pm 1,96 * \sqrt{16,602} = [-4,21; 11,76]$$

Para los horizontes $h = 2$, $h = 3$, se utilizan las siguientes fórmulas:

$$\begin{aligned} \hat{\Sigma}_T(2) &= \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' + \hat{A}_6 \hat{\Sigma}_u \hat{A}_6' + \hat{A}_{10} \hat{\Sigma}_u \hat{A}_{10}' + \hat{A}_{14} \hat{\Sigma}_u \hat{A}_{14}' \\ &= \begin{bmatrix} 27,78 & 1,51 \\ 0,93 & 20,44 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \hat{\Sigma}_T(3) &= \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' + \dots + \hat{A}_{12} \hat{\Sigma}_u \hat{A}_{12}' + \hat{A}_1^2 \hat{\Sigma}_u \hat{A}_1^{2'} + \dots + \hat{A}_{14}^2 \hat{\Sigma}_u \hat{A}_{14}^{2'} \\ &= \begin{bmatrix} 27,92 & 1,25 \\ 0,96 & 20,72 \end{bmatrix} \end{aligned}$$

Entonces los intervalos de confianza son:

$$IC(X_{1,15:2}) = -1,394 \pm 1,96 * 5,43 = [-9,93; 10,73]$$

$$IC(X_{2,15:2}) = 0,268 \pm 1,96 * 4,59 = [-10,14; 7,58]$$

$$IC(X_{1,15:3}) = -0,645 \pm 1,96 * 5,45 = [-10,58; 20,51]$$

$$IC(X_{2,15:3}) = -0,832 \pm 1,96 * 4,62 = [-9,46; 17,99]$$

Comparación con modelos univariantes

La teoría VAR sugiere que las predicciones logradas son de mejor calidad que si se realiza la modelación de las series de manera univariante. Para comprobar esto, se realizó un modelo univariante para cada una de las series analizadas en este ejemplo. Así, se encontró que para el IPP el modelo univariante es $X_{1t} = 0,18X_{1t-1} - 0,24X_{1t-6} + \hat{u}_t$; mientras que, para el IAE es $X_{2t} = -0,25X_{2t-1} - 0,27X_{2t-10} + \hat{u}_t - 0,32\hat{u}_{t-8}$. Con estos modelos se realizaron las predicciones para el año 2014 y se obtiene lo siguiente:

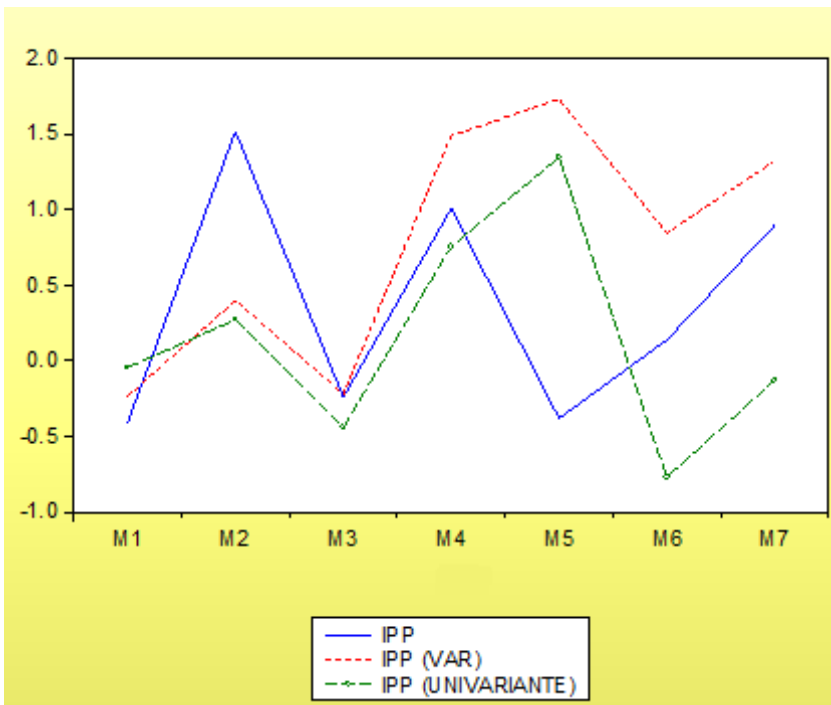


Figura 2.15: Comparación de las predicciones VAR y UNIVARIANTE para X_{1t} (IPP)

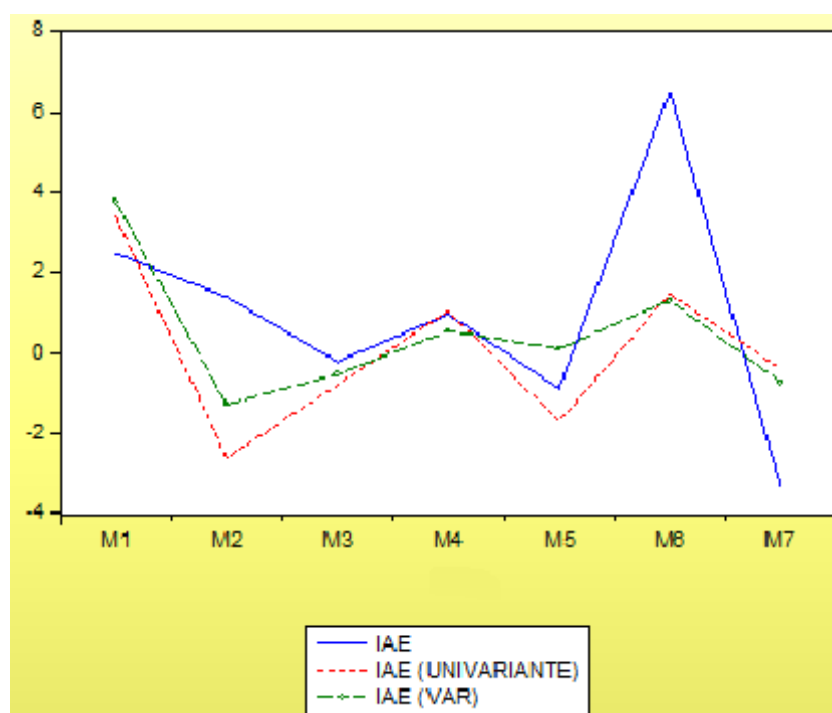


Figura 2.16: Comparación de las predicciones VAR y UNIVARIANTE para X_{2t} (IAE)

Se puede observar que el ajuste que tienen las predicciones del modelo VAR para el IPP es un tanto mejor que las del modelo univariante. Por otro lado, en el caso del IAE, las predicciones parecen bastante similares entre los dos modelos. Para poder determinar esto, se realiza el cálculo del error cuadrático medio para determinar el mejor ajuste. Así se tiene:

	IPP	IAE
VAR	0,95	6,29
UNIVARIANTE	0,95	7,41

Tabla 2.23: Error Medio Cuadrático estimado para los modelos VAR y UNIVARIANTE

Como se puede observar el modelo VAR es mejor en el caso del IAE; sin embargo, no lo es para el IPP. Se realizó una prueba t de medias para determinar si existe diferencia estadística entre las medias de los errores cuadráticos generados por los modelos y se comprobó que en el caso del IAE el error medio cuadrático es diferente estadísticamente entre los modelos (el modelo VAR ajusta mejor los datos); en el caso del IPP se determinó que no existe diferencia significativa entre los errores cuadráticos. Como conclusión, el modelo VAR predice de mejor manera que el modelo univariante.

3. La causalidad

El siguiente gráfico muestra la salida del paquete EViews para la Prueba de Granger.

Dependent variable: IPP			
Excluded	Chi-Square	df	Prob.
IAE	6.971411	4	0.1374
All	6.971411	4	0.1374

Dependent variable: IAE			
Excluded	Chi-Square	df	Prob.
IPP	11.44772	4	0.0220
All	11.44772	4	0.0220

Tabla 2.24: Prueba de causalidad de Granger

Como se puede ver, el p-valor (Prob.) es menor que 0,05 en el segundo caso y mayor que 0,05 en el primero; por lo que se concluye que X_{1t} (IPP) explica significativamente la variable X_{2t} (IAE), pero X_{2t} no explica significativamente la variable X_{1t} .

4. Análisis de los “choques”

En las tablas siguientes se muestran los resultados de las funciones de impulso-respuesta que presenta el paquete EViews para las variables X_{1t} y X_{2t} ; también se muestra los gráficos correspondientes:

Período	IPP	IAE
1	1,000000	0,000000
2	0,144733	0,078893
3	0,042822	-0,010459
4	0,003298	0,006279
5	0,002218	-0,001481
6	-8,96E-05	0,000586
7	-0,198874	-0,156816
8	-0,100462	0,001894
9	-0,022618	-0,014730
10	-0,007966	0,001527

Tabla 2.25: Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{1t}

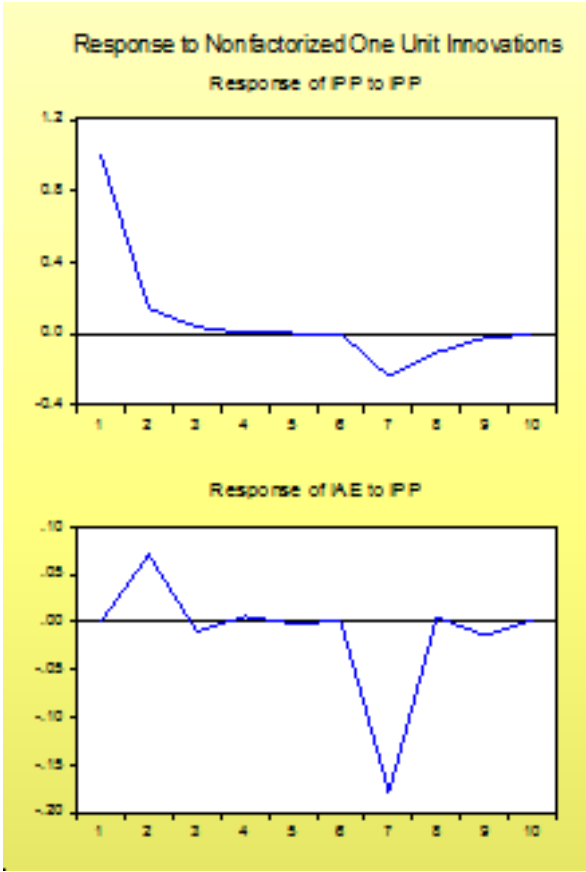


Figura 2.17: Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{1t}

Período	IPP	IAE
1	0.000000	1.000000
2	0.224126	-0.287467
3	-0.032697	0.098387
4	0.017422	-0.030581
5	-0.004387	0.010015
6	0.001624	-0.003187
7	0.032155	-0.059519
8	-0.072018	-0.003284
9	0.000135	-0.004230
10	-0.006113	-3.68E-05

Tabla 2.26: Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{2t}

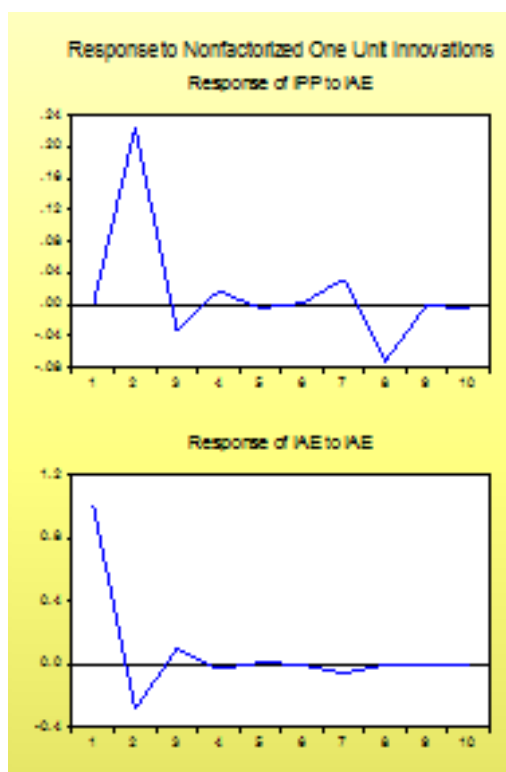


Figura 2.18: Respuesta de las variables X_{1t} y X_{2t} ante un choque unitario de X_{2t}

La elección de la dirección del impacto es muy importante y determina los valores obtenidos. Se puede observar que el efecto de la innovación se desvanece con el tiempo; esto caracteriza a un proceso VAR estacionario.

5. Descomposición de la varianza

A partir de la representación $VAR(14)$ estimada se calcula la descomposición de la varianza.

Dado que las variables del ejemplo son variaciones del IPP y el IAE, es lógico pensar que un choque sobre la variable variación del IPP influya la variación del IAE más que si el choque fuera al revés. Esto sería: una innovación sobre X_{1t} (variación del IPP) influya de manera instantánea a X_{2t} (variación del IAE); por otro lado, una innovación sobre X_{2t} no influya de manera contemporánea a X_{1t} .

La matriz de varianza-covarianza estimada de los residuos es igual (ejemplo 5.5) a:

$$\hat{\Sigma}_u = \begin{bmatrix} 23,82 & 0,42 \\ 0,42 & 16,60 \end{bmatrix}$$

Con el programa EViews, las salidas de las funciones de impulso-respuesta y la descomposición de varianza serían:

Descomposición de la varianza de X1:				Descomposición de la varianza de X2:			
Periodo	S.E.	X1	X2	Periodo	S.E.	X1	X2
1	4,880657	100,0000	0,000000	1	4,074543	0,045030	99,95497
2	5,015877	96,68675	3,313246	2	4,251402	0,601299	99,39870
3	5,020588	96,62258	3,377415	3	4,270454	0,605403	99,39460
4	5,021112	96,60331	3,396688	4	4,272338	0,608029	99,39197
5	5,021150	96,60210	3,397905	5	4,272537	0,608160	99,39184
6	5,021154	96,60193	3,398072	6	4,272558	0,608181	99,39182
7	5,157177	96,71430	3,285697	7	4,368618	4,623096	95,37690
8	5,191890	96,43880	3,561199	8	4,368704	4,625945	95,37406
9	5,193082	96,44043	3,559566	9	4,369329	4,651664	95,34834
10	5,193258	96,43838	3,561625	10	4,369338	4,652042	95,34796
Orden de Cholesky: X2 X1							

Tabla 2.27: Descomposición de la Varianza (Orden de Cholesky X1 X2)

La descomposición de la varianza indica que la varianza del error de predicción de X_{1t} representa un 96,63 % con sus propias innovaciones y un 3,56 % con las de X_{2t} . La varianza del error de predicción de X_{2t} es de un 4,65 % con X_{1t} y un 95,34 % con X_{2t} . Este efecto de asimetría se estudió en la parte de la causalidad; lo que también se corrobora ahora.

Por último, cabe señalar que la tabla anterior muestra la desviación estándar del error de previsión para X_{1t} y X_{2t} .

Capítulo 3

La Cointegración

El análisis de cointegración fue tratado por Granger (1983) y por Engle y Granger (1987); este concepto se considera uno de los más importantes dentro del análisis de series temporales.

La cointegración aparece cuando dos o más series presentan una relación de movimiento conjunto (tendencia) a largo plazo y las diferencias entre ellas son estables.

3.1 Propiedades del orden de Integración de una serie

Sean dos series afectadas de una tendencia (ver los gráficos 6.1 y 6.2); intuitivamente se puede decir lo siguiente:

- En el primer caso, las dos series tienen una tendencia de evolución similar en un primer lapso y después una tendencia de evolución divergente en un segundo período; entonces las series no están cointegradas.
- En el segundo caso, las dos series tienen una evolución similar en todo el período de análisis; las series están cointegradas, si existe una evolución a largo plazo similar entre las series.

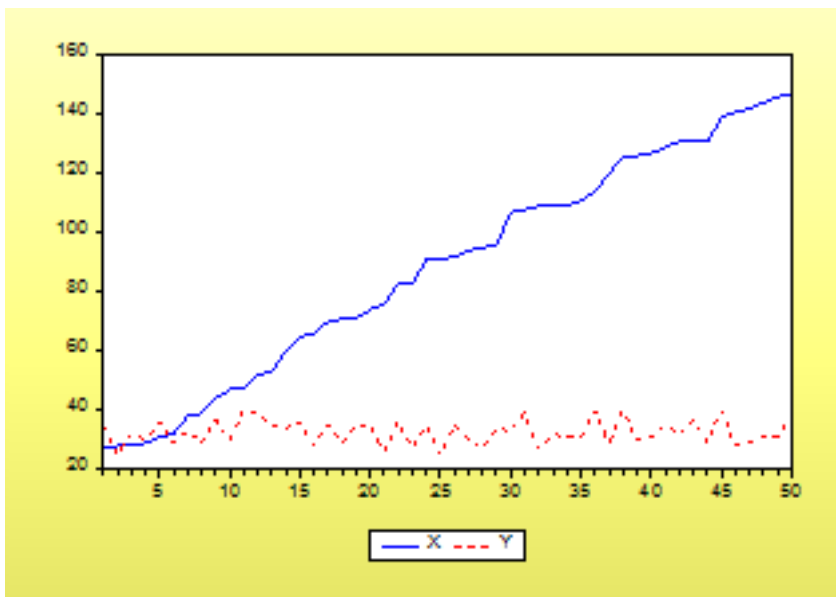


Figura 3.1: Las variables X_{1t} y X_{2t} no están cointegradas

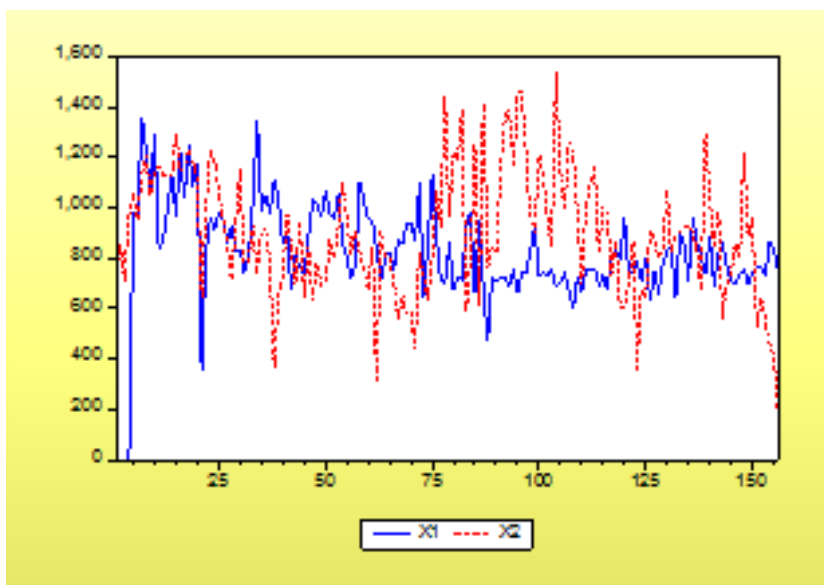


Figura 3.2: Las variables X_{1t} y X_{2t} están cointegradas

Recuérdese que una serie es integrada de orden d , si hay que diferenciarla d veces para volverla estacionaria.

Sea una serie X_{1t} estacionaria y una serie X_{2t} integrada de orden 1; entonces se obtiene:

$$\begin{matrix} X_{1t} \rightarrow I(0) \\ X_{2t} \rightarrow I(1) \end{matrix} \Rightarrow X_{1t} + X_{2t} \rightarrow I(1)$$

La serie $X_t = X_{1t} + X_{2t}$ no es estacionaria ya que es la suma de una serie afectada de una tendencia y una serie estacionaria.

En general:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d') \end{matrix}, d \neq d' \Rightarrow X_{1t} + X_{2t} \rightarrow I(?)$$

Es imposible llegar a una conclusión con respecto a la suma de dos series de orden de integración diferente.

Incluso si dos series X_{1t} y X_{2t} son integradas de orden d , en general se obtiene:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d) \end{matrix} \Rightarrow X_{1t} + X_{2t} \rightarrow I(?)$$

También, para la combinación lineal $\alpha X_{1t} + \beta X_{2t} \rightarrow I(?)$.

En efecto, el resultado depende de los signos de los coeficientes α , β y de la existencia de una dinámica no estacionaria común.

3.2 Condiciones de cointegración

Dos series X_{1t} y X_{2t} son cointegradas si se satisfacen las siguientes condiciones:

- Ambas están afectadas por una tendencia del mismo orden de integración d .
- Una combinación lineal de estas series permite reducir a una serie de orden de integración inferior.

Es decir, si:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d) \end{matrix}$$

Tal que, $\alpha X_{1t} + \beta X_{2t} \rightarrow I(d - b)$, donde $d \geq b > 0$.

Se denota: $X_{1t}, X_{2t} \rightarrow CI(d, b)$, con $[\alpha \quad \beta]'$ como el vector de cointegración.

En el caso general (k variables), se tiene:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d) \\ \dots \\ X_{kt} \rightarrow I(d) \end{matrix}$$

denótese por $X'_t = [X_{1t} \quad X_{2t} \dots X_{kt}]$.

Si existe un vector de cointegración $\beta' = [\beta_1 \ \beta_2 \dots \beta_k]$, tal que $\beta' X_t \rightarrow I(d - b)$, entonces las k variables están cointegradas y el vector de cointegración es β . Se denota $X_t \rightarrow CI(d - b)$ con $b > 0$.

Observación 3.1.

1. No se diferencian las series individualmente ya que, en general, esto produciría que se sobrediferencie el sistema. Es decir, cuando se encuentra un vector de cointegración ocurre que se elimina la tendencia común que existe entre las series (lo que correspondería a una diferenciación); si se diferencian individualmente se estaría realizando diferenciaciones adicionales a las que se logra con el vector de cointegración.
2. En la práctica, el caso más considerado es cuando $d = b = 1$, que es lo que se asume en este documento.

Ejemplo 3.1. Se consideran tres series económicas de un país sudamericano: Producto Interno Bruto, Consumo Interno y la Demanda Final Interna denotadas por X_{1t} , X_{2t} , X_{3t} , respectivamente. Los datos son trimestrales en un período que va desde noviembre 2010 hasta junio de 2015 (56 observaciones), los que se utilizarán para analizar una posible cointegración entre las variables; para efectos de ejemplo, se trabajará únicamente con 50 datos y se dejarán 6 (enero a junio de 2015) para poder realizar las comparaciones con las predicciones (ver anexo D.1). Se pide analizar una eventual cointegración entre las variables.

Resolución.

Para iniciar, se presentan los gráficos de las series temporales a analizarse:

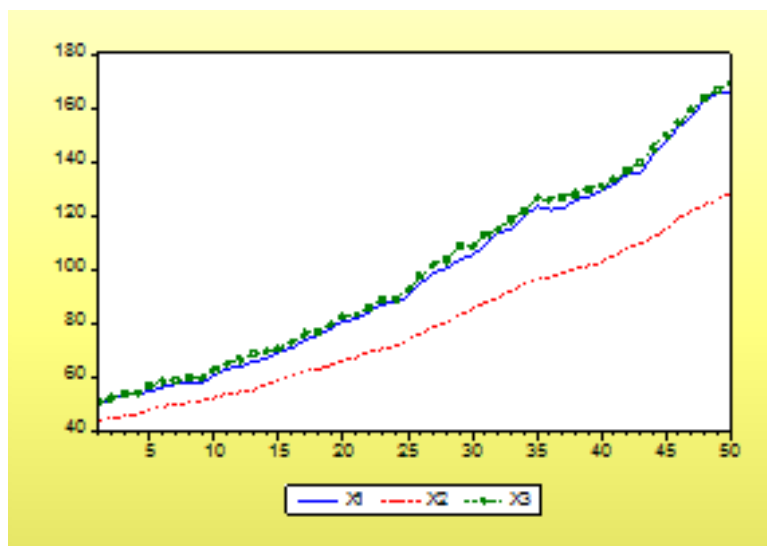


Figura 3.3: Series consideradas en el ejemplo

Ahora, para determinar si las series son estacionarias es necesario verificar si existe una raíz unitaria en las series. Se realiza la prueba DFA a nivel, para las tres series:

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.317447	0.8718
Test critical values:	1 % level	-4.156734
	5 % level	-3.504330
	10 % level	-3.181826

Tabla 3.1: Prueba de raíces unitarias para las X_{1t}

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.813922	0.9572
Test critical values:	1 % level	-4.156734
	5 % level	-3.504330
	10 % level	-3.181826

Tabla 3.2: Prueba de raíces unitarias para las X_{2t}

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.054405	0.9263
Test critical values:	1 % level	-4.156734
	5 % level	-3.504330
	10 % level	-3.181826

Tabla 3.3: Prueba de raíces unitarias para las X_{3t}

Se concluye que las tres series tienen raíz unitaria. A continuación, se presentan las pruebas de raíces unitarias de las series en primeras diferencias:

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.222194	0.0000
Test critical values:	1 % level	-4.161144
	5 % level	-3.506374
	10 % level	-3.183002

Tabla 3.4: Prueba de raíces unitarias para X_{1t} en primera diferencia

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.620775	0.0001
Test critical values:	1 % level	-4.161144
	5 % level	-3.506374
	10 % level	-3.183002

Tabla 3.5: Prueba de raíces unitarias para X_{2t} en primera diferencia

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.038579	0.0000
Test critical values:	1 % level	-4.161144	
	5 % level	-3.506374	
	10 % level	-3.183002	

Tabla 3.6: Prueba de raíces unitarias para X_{3t} en primera diferencia

De las pruebas se concluye que se necesita realizar solamente una diferenciación no estacional para volverlas estacionarias; por lo tanto, las series son integradas de orden 1, $I(1)$.

Capítulo 4

Modelo de Corrección del error (MCE)

El modelo de corrección del error (MCE) recibe su nombre debido a que mediante este, se corrigen los errores o desviaciones que las variables presentan en el corto plazo con respecto a su relación de equilibrio a largo plazo.

Del párrafo anterior se puede determinar que el MCE está relacionado con la cointegración de las variables; si las variables no están cointegradas, no se puede plantear el MCE.

Al utilizar este método, se considera el comportamiento dinámico de las series; además, presenta ventajas con respecto a los métodos en los que se utiliza la diferenciación de las series para convertirlas en estacionarias, pues pueden dar lugar a casos de sobrediferenciación y pérdida de información.

4.1 Planteamiento del MCE

Se considera el modelo VAR(p) de k variables, donde por simplicidad se omiten las posibles componentes deterministas:

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + u_t$$

con u_t ruido blanco.

Considérese,

$$A(B) = I - A_1 B - \dots - A_p B^p$$

Recuérdese que si las raíces del polinomio característico están fuera del círculo unidad, X_t es estacionario; es decir, es $I(0)$. Si $|A(1)| = 0$, se dice que X_t tiene una raíz

unitaria. Por simplicidad de la exposición, se asume que X_t es integrada de orden 1 ($I(1)$); esto significa que:

$$(1 - B)X_t$$

es estacionaria, si X_t no lo es.

Un modelo de corrección del error (MCE) para un proceso $VAR(p)$ es:

$$\Delta X_t = \Omega X_{t-1} + A_1^* \Delta X_{t-1} + \dots + A_{p-1}^* \Delta X_{t-p+1} + u_t$$

donde,

$$A_j^* = - \sum_{i=j+1}^p A_i, \quad j = 1, \dots, p-1$$

$$\Omega = \alpha \beta' = A_p + A_{p-1} + \dots + A_1 - I = -A(1)$$

α y β son matrices de dimensión $k \times r$ (r es el número de factores de cointegración y $r < k$), de rango completo.

Al término ΩX_{t-1} se lo conoce como el **término de corrección del error**, que desempeña un papel clave en el estudio de cointegración. La existencia del término $\beta' X_{t-1}$ es natural en la representación de corrección del error; además, es estacionario. Las columnas de β son los vectores de cointegración de X_t .

La estacionariedad de $\beta' X_{t-1}$ puede ser justificada de la siguiente manera: La teoría de series con raíces unitarias muestra que el coeficiente de correlación simple entre una serie no estacionaria con raíz unitaria y una estacionaria tiende a 0 cuando el tamaño de la muestra tiende a infinito ($T \rightarrow \infty$). En un MCE X_{t-1} es no estacionario, pero ΔX_t es estacionario; por lo tanto, solo existe una forma en que ΔX_t puede relacionarse significativamente con X_{t-1} : a través de la serie estacionaria $\beta' X_{t-1}$.

Para poder recuperar A_i a partir de MCE se tiene:

$$A_1 = I + \Omega + A_j^*$$

$$A_i = A_i^* - A_{i-1}^*, \quad i = 2, \dots, p$$

donde $A_p^* = 0$, la matriz cero. En lo que sigue se supone que X_t es al menos $I(1)$. Por lo tanto, se consideran tres casos en el MCE:

- Si $rg(\Omega) = 0$, implica que $\Omega = 0$ y las X_t son no cointegradas. Con esto, se estima un modelo VAR clásico en primeras diferencias a fin de eliminar la tendencia.
- Si $rg(\Omega) = k$, implica que X_t no contiene raíces unitarias; esto es, X_t es $I(0)$. El modelo de corrección del error no es informativo y se debe estudiar directamente X_t .
- Si $0 < rg(\Omega) = r < k$, existen r vectores (fila) linealmente independientes que recogen r relaciones de cointegración diferentes; entonces, se puede descomponer

la matriz Ω en el producto de dos submatrices α y β , de orden $(k \times r)$ de forma tal que:

$$\Omega = \alpha\beta'$$

donde β' contiene los coeficientes de los vectores de cointegración y α los parámetros de la velocidad de ajuste. Sustituyendo la matriz Ω en la expresión general del modelo, se tiene:

$$\Delta X_t = \Omega X_{t-1} + A_1^* \Delta X_{t-1} + \dots + A_{p-1}^* \Delta X_{t-p+1} + u_t$$

donde, el producto $\beta' X_{t-1}$ es estacionario.

4.2 Prueba de Johansen

Para determinar el número de relaciones de cointegración Johansen (1988) propuso dos pruebas fundamentadas en los valores propios de la matriz Ω .

En la práctica, el procedimiento de aplicación de la prueba de Johansen se realiza mediante las siguientes etapas:

1. Determinación del orden óptimo del modelo VAR con las series que son integradas del mismo orden. Esta etapa es muy importante, ya que los resultados del contraste son muy sensibles frente a una mala especificación de retardos. Posteriormente, se realizará un ajuste con respecto a la tendencia común.
En algunos paquetes estadísticos como el EViews, se escoge el retardo del VAR pero con las variables diferenciadas una vez. Es decir, el paquete realiza una diferenciación en el VAR de manera automática.
2. Estimación del modelo. Para realizar la descomposición de la matriz Ω en las submatrices α y β , se utiliza el método de máxima verosimilitud propuesto por Johansen (1988).
3. Determinación del rango de cointegración, que vendrá dado por el propio rango de la matriz Ω . Teniendo en cuenta que se trata de una matriz de coeficientes estimados, y por lo tanto aleatorios, el rango de dicha matriz no puede determinarse de forma absoluta, sino que habrá que definirlo dentro de un entorno probabilístico.

Se definen dos estadísticos alternativos, basados en los valores propios λ_i de la matriz $\hat{\Omega}$ y formulados como:

a. **Traza**

$$\lambda_{\text{traza}}(r) = -T \sum_{i=r+1}^k \ln(1 - \lambda_i)$$

donde:

λ_i = es el i -ésimo valor propio de la matriz $\hat{\Omega}$ (en orden creciente),
 r = rango de la matriz $\hat{\Omega}$,
 k = número de variables,
 T = número de observaciones.

Con este estadístico se contrasta la hipótesis nula de existencia de un número de vectores de cointegración menor o igual a r frente a la hipótesis alternativa de existencia de más de r relaciones de cointegración.

b. Máximo valor propio

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \lambda_{r+1})$$

donde:

λ_i = es el i -ésimo valor propio de la matriz $\hat{\Omega}$ (en orden creciente),
 r = rango de la matriz $\hat{\Omega}$,
 k = número de variables,
 T = número de observaciones.

Este estadístico contrasta la hipótesis nula de existencia de r vectores de cointegración frente a la hipótesis alternativa de existencia de $r+1$ vectores de cointegración.

Estos estadísticos tienen una ley de probabilidad (similar a una χ^2) tabulada mediante simulaciones por Johansen y Juselius (1990).

Ejemplo 4.1. A partir del ejemplo 6.1, se pide analizar una eventual cointegración entre las variables y estimar un modelo VAR o un modelo VEC si fuese el caso.

Resolución

- a. **Primera etapa:** Determinar el número de retardos de la representación VAR.

Para iniciar, se presentan los gráficos de las series temporales a analizarse:

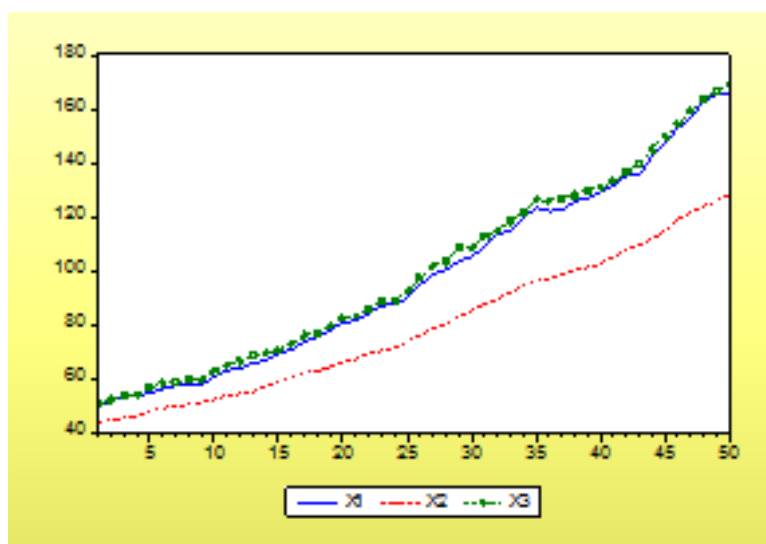


Figura 4.1: Series consideradas en el ejemplo

Como se vio en el ejemplo 6.1, las series son cointegradas de orden 1. Ahora, se determina el mejor retardo para el modelo VAR (considerando diferentes criterios):

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-170.1455	NA	1.441464	8.879257	9.007223*	8.925170*
1	-158.4805	20.93714	1.260025*	8.742591	9.254456	8.926244
2	-150.7736	12.64724	1.359430	8.808904	9.704668	9.130296
3	-146.6524	6.129033	1.787069	9.059096	10.33876	9.518228
4	-143.2875	4.486496	2.493999	9.348077	11.01164	9.944949
5	-124.5157	22.14106*	1.627928	8.846961	10.89442	9.581572
6	-119.0699	5.585502	2.196024	9.029224	11.46058	9.901575
7	-116.5909	2.161115	3.663332	9.363638	12.17890	10.37373
8	-98.07554	13.29311	2.931050	8.875669	12.07483	10.02350
9	-88.11999	5.615947	4.160914	8.826666	12.40972	10.11224
10	-76.43760	4.792776	6.776135	8.689108*	12.65606	10.11242

Tabla 4.1: Elección del retardo del VAR

Como se puede observar, los criterios difieren de cuál es el mejor retardo del VAR a estimar (cabe considerar que en este caso se consideran las series en primeras diferencias). Los valores dados en la tabla se pueden considerar como cotas superiores del retardo del VAR. Debido a que se tienen pocos datos se desechará el retardo 10, por lo que, se puede estimar un VAR(5).

Esto da lugar a 2 posibles modelos:

1. VAR(5), considerando solamente el retardo 5.
2. VAR(5), considerando los retardos del 1 al 5.

b. **Segunda Etapa:** Prueba de Johansen

Como ya se mencionó, las series son $I(1)$. Se realizan las pruebas sobre los supuestos de la tendencia de los datos (éstas están implementadas en EViews):

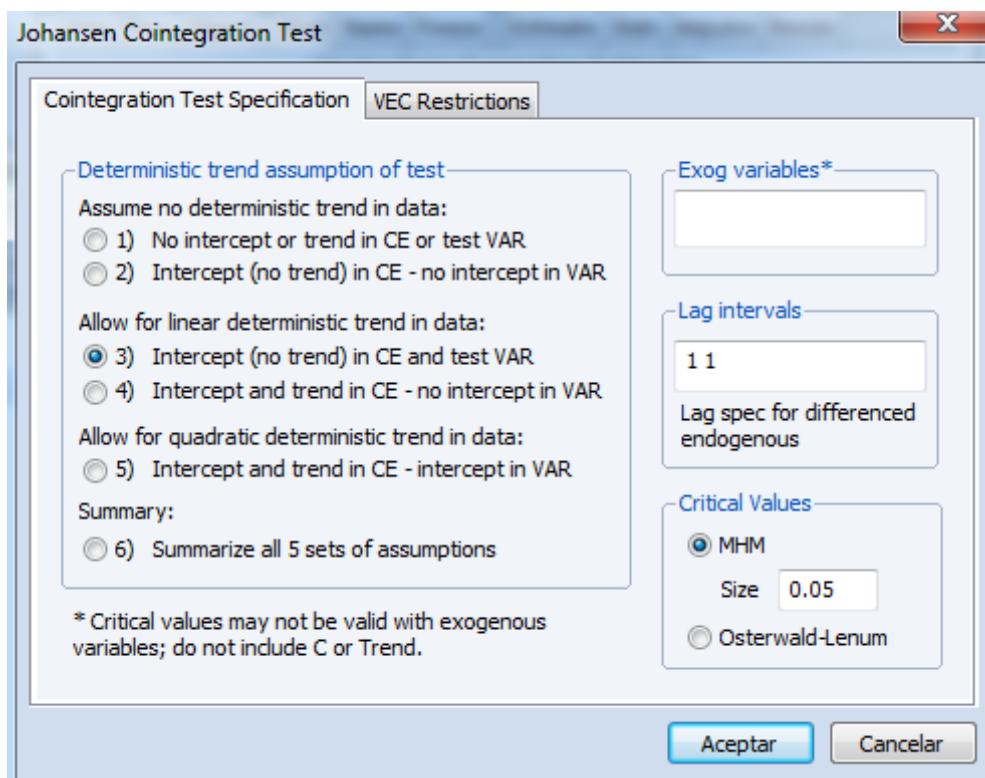


Figura 4.2: Prueba de cointegración de Johansen

Nótese, que esta prueba es válida solo para series que no son estacionarias. Además, los retardos se fijan para la series en primeras diferencias (ver figura 9.3, *lag intervals*) y no en niveles. Por ejemplo, si se escribe "1 2" la prueba hace la regresión de ΔX_t sobre ΔX_{t-1} , ΔX_{t-2} y sobre las variables exógenas especificadas, de ser el caso. Para X_t tendría 3 retardos. Si se desea correr la prueba con un retardo en niveles, se debe escribir "0 0".

Índice alfabético

ARCH

- Definición, 3
- Restricciones, 4

Causalidad, 40

- Granger, 41
- Sims, 42

Cointegración

- Condiciones, 75
- Definición, 73

Dependencia lineal multivariante, 13

Descomposición de Cholesky, 52

EGARCH

- Definición, 5

Factor de impacto, 22

GARCH

- Definición, 4

IGARCH

- Definición, 4

Innovación

- Ortogonal, 52

Matriz

- Correlaciones cruzadas, 12
- De correlación residual, 26
- De covarianza residual, 26

Modelos MCE, 79

- Planteamiento, 79
- Término de corrección del error, 80

Modelos VAR, 17

- Bivariante, 17
- Diagnóstico y validación, 26
- Estabilidad, 19
- Estimación de los coeficientes, 24
- Forma estándar, 18

Forma estructural, 18

Forma matricial, 18

Predicción, 29

Representación de VAR(1), 20

Representación general, 19

Representación lineal, 21

Modelos VARMA, 23

Modelos VMA

- Orden q , 21

Multiplicadores

- De impacto, 52
- Dinámicos, 45

PARCH

- Definición, 6

Predicción

- Modelos VAR, 29

Prueba

- Jarque-Bera multivariante, 28
- Johansen, 81
- Multiplicador de Lagrange multivariante, 27
- Portmanteau multivariante, 27

Relaciones en retardo, 12

TARCH

- Definición, 6