Benemérita Universidad Autónoma de Puebla

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS
LICENCIATURA EN MATEMÁTICAS

MODELADO Y PREDICCIÓN DEL PIB TRIMESTRAL DE MÉXICO A TRAVÉS SERIES DE TIEMPO MÚLTIPLES

TESIS

QUE PARA OBTENER EL TÍTULO DE
LICENCIADA EN MATEMÁTICAS

PRESENTA
Arely Maldonado Azcona

DIRECTOR DE TESIS
Dr. Víctor Hugo Vázquez Guevara

PUEBLA, PUE. FECHA: Junio 2021

A mis padres.

Agradecimientos

Quiero agradecer a mis padres, por el apoyo incondicional en este largo camino; por su lucha incomparable para salir siempre adelante, los admiro mucho.

A mi hermana Fer, por su cariño, por estar en los buenos y malos momentos, y por soportarme tanto tiempo.

A mi director de tesis, Dr.Víctor Hugo Vázquez Guevara, por la confianza, apoyo y sobre todo paciencia que me brindó durante este proceso, muchas gracias.

A Luis, por apoyarme siempre y compartir estos últimos años conmigo. Gracias por tu motivación y cariño en momentos difíciles.

A mis amigos, por motivarme a nunca darme por vencida. Mire, gracias por tu bonita amistad, por enseñarme lo bonito de la vida que a veces no logro ver y por compartir tus conocimientos conmigo; eres una gran persona. Julio Cesar, gracias por tantos años de amistad, por creer en mí y apoyarme en este proceso. Oscar, gracias por tu cariño y amistad sin condiciones, por la motivacion día a día y el tiempo compartido. Yuli y Bren, gracias por su amistad tan bonita y duradera, me inspiran a seguir siempre adelante.

A mis sinodales, Dra. Hortensia Josefina Reyes Cervantes, Dr. Francisco Solano Tajonar Sanabria y Dr. Bulmaro Juárez Hernández, por aceptar el cargo y por tomarse el tiempo para la revisión de mi tesis.

Introducción

A menudo se necesitan hacer predicciones de interés que servirán para conocer algo que se desea, para tomar decisiones que dependen de lo que pase en el futuro, etc. Si las observaciones de series temporales están disponibles para una variable de interés y los datos del pasado contienen información sobre el desarrollo futuro de una variable, es posible utilizar como pronóstico alguna función de los datos recopilados en el pasado.

Este enfoque para pronósticar puede ser expresado como sigue. Sea y_t el valor de la variable de interés en el periodo t. Entonces un pronóstico para el periodo $T+h$, puede tener la forma

$$\widehat{y}_{T+h} = f(y_T, y_{T-1}, \ldots, y_0), \tag{0.0.1}$$

donde $f(\cdot)$ es una función adecuada que depende de las observaciones pasadas $y_T, y_{T-1}, \cdots, y_0$.

Frecuentemente la variable de interés no está relacionada sólo con sus valores pasados en el tiempo sino que, además, depende de los valores pasados de otras variables. En este caso, si las variables relacionadas se denotan por $y_{1t}, y_{2t}, \ldots y_{Kt}$, el pronóstico de $y_{k,T+h}$ puede ser de la forma
$\widehat{y}_{k,T+h} = f_1(y_{1,T}, y_{2,T}, \cdots, y_{K,T}, y_{1T-1}, y_{2,T-1}, \cdots, y_{K,T-1}, y_{1,T-2}, \cdots, y_{K,T-2}, \cdots, y_{1,0}, y_{2,0}, \cdots, y_{K,0})$.

Un conjunto de series de tiempo $y_{k,t}$, $k = 1, \ldots, K$, $t = 1, \cdots, T$ es llamada una serie de tiempo múltiple y la fórmula anterior expresa el pronóstico $\widehat{y}_{k,T+h}$ como una función de una serie de tiempo múltiple. Uno de los principales objetivos del análisis de series de tiempo múltiples es determinar funciones adecuadas $f_1, \ldots, f_K$ que pueden usarse para obtener pronósticos.

Algunas de las definiciones básicas para entender más, son las siguientes.

Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad, donde Ω es un espacio muestral, $\mathcal{F}$ una sigma álgebra de eventos de Ω, y $\mathcal{P}$ es una medida de probabilidad definida en $\mathcal{F}$. Una variable aleatoria (y) es una función de variables reales definida en Ω tal que, para cada número real c, $A_c = \{\omega \in \Omega \mid y(\omega) \leq c\} \in \mathcal{F}$. Esto es, A_c es un evento para el cual la probabilidad está definida en terminos de $\mathcal{P}$. La función $F_y : \mathbb{R} \longrightarrow [0, 1]$ definida como $F_y(c) = \mathcal{P}(A_c)$, es la función de distribución de y.

Un vector aleatorio K-dimensional o un vector de variables aleatorias es una función y de Ω al espacio Euclideano K-dimensional $\mathbb{R}^k$, esto es, y mapea a $\omega \in \Omega$ en $y(\omega) = (y_1(\omega), ..., y_k(\omega))$ tal que, para cada $c = (c_1, ..., c_k)' \in \mathbb{R}^k$ $A_c = \{\omega | y_1(\omega) \leq c_1, ..., y_k(\omega) \leq c_k(\omega)\} \in \mathcal{F}$.
La función $F : \mathbb{R}^k \longrightarrow [0, 1]$ definida por $F(c) = \mathcal{P}(A_c)$ es la función de distribución conjunta de y.
Si Z es un conjunto indexado a lo más numerable, por ejemplo, el conjunto de todos los enteros o todos los enteros positivos; un proceso estocástico (discreto) es una función de valores reales

$$y : Z \times \Omega \longrightarrow \mathbb{R},$$

tal que, para cada $t \in Z$ fijo, $y(t, \omega)$ es una variable aleatoria. Un proceso estocástico puede ser descrito por las funciones de distribución de todas las subcolecciones finitas de las y_t, $t \in S \subset Z$.
De manera análoga, un proceso estocástico múltiple es una función

$$y : Z \times \Omega \longrightarrow \mathbb{R}^K,$$

donde para cada $t \in Z$ fijo, $y(t, w)$ es un vector aleatorio $K-$dimensional.

Una realización de un proceso estocástico K-dimensional es una sucesión de vectores $y_t(\omega)$, $t \in Z$, para ω fijo. En otras palabras, una realización de un proceso estocástico es una función $Z \longrightarrow \mathbb{R}^k$, donde $t \longrightarrow y_t(w)$. Una serie de tiempo múltiple es considerada como una realización o posiblemente una parte finita de una realización, esto es, de vectores de la forma $(y_1(\omega), ...y_k(\omega))$.

Se puede comenzar con predicciones que son funciones lineales de observaciones pasadas. Consideremos una serie de tiempo univariada y_t y un pronóstico de periodo $h = 1$

$$\hat{y}_{T+1} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \dots$$

Asumiendo que únicamente un número finito, digamos p, de valores pasados son usados en la fórmula de predicción, obtenemos

$$\hat{y}_{T+1} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \dots + \alpha_p y_{T-p+1}. \tag{0.0.2}$$

Por supuesto, los valores verdaderos de y_{T+1} no son exactamente iguales a los del pronóstico $\hat{y}_T$. Si denotamos el error de pronóstico por $u_{T+1} := y_{T+1} - \hat{y}_{T+1}$, se obtiene que

$$y_{T+1} = \hat{y}_{T+1} + u_{T+1} = \nu + \alpha_1 y_T + ... + \alpha_p y_{T-p+1} + u_{T+1}. \tag{0.0.3}$$

Ahora, si el número de realizaciones de vectores aleatorios y la generación de datos se mantiene en cada periodo T, (0.0.3) tiene la forma de un proceso autorregresivo

$$y_t = \nu + \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + u_t, \tag{0.0.4}$$

donde $y_t, y_{t-1}, ..., y_{t-p}$ y u_t son variables aleatorias. Para obtener un proceso autorregresivo se asume que los errores de pronóstico u_t para diferentes periodos son no correlacionados, esto es, u_t y u_s son no relacionados para $s \neq t$.

Si la serie de tiempo es considerada múltiple, una extensión de (0.0.3) es

$$\begin{aligned} \widehat{y}_{k,T+1} =& \nu + \alpha_{k1,1} y_{1,T} + \alpha_{k2,1} y_{2,T} + ... + \alpha_{kK,1} y_{K,T} + ... + \\ & \alpha_{k1,p} y_{1,T-p+1} + ... + \alpha_{kK,p} y_{K,T-p+1}, \quad k = 1, 2, ..., K. \end{aligned} \tag{0.0.5}$$

Para simplificar la notación, sea $y_t := (y_{1t}, ..., y_{Kt})'$, $\widehat{y}_t := (\widehat{y}_{1t}, ..., \widehat{y}_{Kt})', \nu = (\nu_1, \nu_2, ..., \nu_K)'$ y

$$A_i := \begin{bmatrix} \alpha_{11,i} & \cdots & \alpha_{1K,i} \\ \vdots & \ddots & \vdots \\ \alpha_{K1,i} & \cdots & \alpha_{KK,i} \end{bmatrix},$$

entonces, (1.3.4) puede escribirse compactamente como

$$\widehat{y}_{T+1} = \nu + A_1 y_T + ... + A_p y_{T-p+1}. \tag{0.0.6}$$

Si las variables aleatorias y_t son consideradas como vectores aleatorios, 0.0.6 es el pronóstico óptimo del modelo de un vector autorregresivo de la forma

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t \tag{0.0.7}$$

donde $u_t = (u_{1t}, ..., u_{kt})'$ forma una sucesión de vectores aleatorios de dimensión K, independientes identicamente distribuidas, con media cero.
En la práctica, para una serie temporal múltiple dada, primero debe especificarse un modelo y estimarse sus parámetros. Después, la adecuación del modelo se verifica mediante diversas herramientas estadísticas y luego el modelo estimado se puede utilizar para pronósticos y análisis dinámicos o estructurales.

Índice general

MODELADO Y PREDICCIÓN DEL PIB TRIMESTRAL DE MÉXICO A TRAVÉS SERIES DE TIEMPO MÚLTIPLES

Arely Maldonado Azcona

Junio 2021

Capítulo 1

Procesos autorregresivos multidimensionales

1.1. Propiedades del proceso VAR

En este capítulo se estudiarán características y propiedades de los procesos auterregresivos vectoriales (VAR), además se conocerán suposiciones que son necesarias para la modelación y estudio de estos.

1.1.1. Procesos estables VAR

El objeto de estudio en lo siguiente es el modelo autorregresivo vectorial de orden p (VAR(p))

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \qquad t = 0 \pm 1 \pm 2, \ldots, \tag{1.1.1}$$

donde $(y_t) = (y_{1t}, ..., y_{kt})^{'}$ es un vector aleatorio, las matrices A_i son la matrices de coeficientes, de dimensión $(k \times k)$; $\nu = (\nu_1, ...\nu_k)^{'}$ es un vector fijo de dimensión $(k \times 1)$ y $u_t = (u_{1t}, ..., u_{kt})$ es un proceso de ruido blanco, esto es, $E(u_t) = 0$, $E(u_t, u_t^{'}) = \Sigma_u$ y $E(u_t, u_s^{'}) = 0$ para $s \neq t$.

Para investigar las condiciones del modelo descrito en (1.1.1), primero se considera el modelo autorregresivo vectorial de orden uno

$$y_t = \nu + A_1 y_{t-1} + u_t. \tag{1.1.2}$$

Si este proceso de generación comienza en el instante $t = 1$,

$$
\begin{aligned}
y_1 &= \nu + A_1 y_0 + u_1, \\
y_2 &= \nu + A_1 y_1 + u_2 = \nu + A_1(\nu + A_1 y_0 + u_1) + u_2 \\
&= (I_k + A_1)\nu + {A_1}^2 y_0 + A_1 u_1 + u_2, \\
y_3 &= \nu + A_1 y_2 + u_3 \\
&= \nu + A_1[(I_k + A_1)\nu + {A_1}^2 y_0 + A_1 u_1 + u_2] + u_3 \\
&= (I_k + A_1 + {A_1}^2)\nu + {A_1}^3 y_0 + {A_1}^2 u_1 + A_1 u_2 + u_3, \\
&\vdots
\end{aligned}
$$

$$
y_t = (I_k + A_1 + {A_1}^2 + ... + {A_1}^{t-1})\nu + {A_1}^t y_0 + \sum_{n=1}^{t-1} {A_1}^i u_{t-i}. \tag{1.1.3}
$$

Por lo tanto, los vectores $y_1, ..., y_t$ están determinados únicamente por $y_0, u_1, \ldots, u_t$. También la distribución conjunta de $y_1, \ldots, y_t$ está determinada por la distribución conjunta de $y_0, u_1, \ldots, u_t$.

Se analizará ahora qué tipo de proceso es congruente con (1.1.1). Para esto se considera otra vez el proceso VAR(1). De (1.1.3) se tiene que

$$
\begin{aligned}
y_t &= \nu + A_1 y_{t-1} + u_t \\
&= (I_k + A_1 + \ldots + {A_1}^j)\nu + {A_1}^{j+1} y_{t-j-1} + \textstyle\sum_{i=0}^{j} {A_1}^i u_{t-i}.
\end{aligned}
$$

Si todos los eigenvalores de A_1 tienen módulo menor que uno, la sucesión A^i, $i = 0, 1, \ldots$, es absolutamente sumable (por regla 3 Apéndice A.2). Por lo tanto, la suma infinita $\sum_{n=1}^{\infty} A^i u_{t-i}$ converge en media cuadrática (por Proposición C.1). Por otra parte,

$$
(I_k + A_1 + \ldots + {A_1}^j)\nu \xrightarrow[j \longrightarrow \infty]{} (I_k - A_1)^{-1}\nu \tag{1.1.4}
$$

(por regla 3 Apéndice A.2). Además , ${A_1}^{j+1}$ converge a cero rápidamente cuando $j \to \infty$, así podemos ignorar al término ${A_1}^{j+1} y_{t-j-1}$, en el límite. Por lo tanto; y_t, en (1.1.2), es un proceso autorregresivo vectorial de orden uno si puede ser escrito de la siguiente manera

$$y_t = \mu + \sum_{n=1}^{\infty} A^i u_{t-i}, \qquad t = 0 \mp 1 \mp 2, \ldots, \tag{1.1.5}$$

donde $\mu = (I_k - A_1)^{-1}\nu$, $t = 0, \pm 1, \pm 2, \ldots$.
La distribución y las distribuciones conjuntas de las variables aleatorias y_t están determinadas únicamente por las distribuciones del proceso u_t. De la Proposición C.2, se tiene que el primer y segundo momentos del proceso (y_t) están dados por

$$\begin{aligned} E[y_t] &= E\left[\mu + \sum_{n=1}^{\infty} A^i u_{t-i}\right] \\ &= E[\mu] + E\left[\sum_{n=1}^{\infty} A^i u_{t-i}\right] \\ &= \mu + \lim_{n\to\infty} \sum_{i=0}^{n} A^i E(u_{t-i}) \\ &= \mu, \end{aligned} \tag{1.1.6}$$

pues $E(u_t) = 0$ para $t = 0\pm, 1 \mp 2, \ldots$, y

$$\begin{aligned} \Gamma_y(h) &= E\left[(y_t - \mu)(y_{t-h} - \mu)'\right] \\ &= E\left[\left(\sum_{i=0}^{\infty} {A_1}^i u_{t-i}\right)\left(\sum_{j=0}^{\infty} {A_1}^j u_{t-h-j}\right)'\right] \\ &= \lim_{n\to\infty} \sum_{i=0}^{n}\sum_{i=0}^{n} A^i E\left(u_{t-i}u'_{t-h-j}\right) {A_j}' \\ &= \lim_{n\to\infty} \sum_{i=0}^{n} A^{h+j}\Sigma_u {A_1^j}' \\ &= \sum_{i=0}^{\infty} A^{h+j}\Sigma_u A_1^j. \end{aligned} \tag{1.1.7}$$

Si todos los eigenvalores de A_1 tienen módulos menores que uno, se dice que (y_t) es un *proceso estable* . Por la regla 7 del Apéndice A.3, la condición

es equivalente a

$$det(I_k - A_1 z) \neq 0, \quad para \quad |z| \leq 1. \tag{1.1.8}$$

Es importante mencionar que el proceso (y_t) para $t = 0 \pm 1, \pm 2, \ldots$ puede también ser definido si la condición de estabilidad, en (1.1.8), no se satisface.

El caso del proceso $VAR(1)$ se puede extender para un proceso VAR(p) con $p > 1$. Esto es, si (y_t) es un proceso VAR(p) de dimensión K, puede escribirse como

$$Y_t = \boldsymbol{\nu} + \boldsymbol{A} Y_{t-1} + U_t, \tag{1.1.9}$$

donde

$$Y_t := \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix},$$

$$\boldsymbol{\nu} := \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$U_t := \begin{bmatrix} u \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\boldsymbol{A} := \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_k & 0 & \cdots & 0 & 0 \\ 0 & I_k & \cdots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_k & 0 \end{bmatrix}.$$

En analogía con la discusión anterior, (Y_t) es estable si

$$det(I_{Kp} - \boldsymbol{A} z) \neq 0 \quad para \quad |z| \leq 1. \tag{1.1.10}$$

Su primer momento es

$$\boldsymbol{\mu} := E(Y_t) = (I_{kp} - \boldsymbol{A})^{-1} \boldsymbol{\nu}$$

y las autocovarianzas están dadas por

$$\Gamma_Y(h) = \sum_{i=0}^{\infty} \boldsymbol{A}^{h+i} \Sigma_U (\boldsymbol{A}^i)', \tag{1.1.11}$$

donde $\Sigma_U := E(U_t, U_t^{'})$.

Considerando a la matriz de dimensión $(K \times Kp)$,

$$J := [I_k : 0 : \ldots : 0], \tag{1.1.12}$$

el proceso puede ser escrito como $(y_t) = JY_t$. La media puede reescribirse como $E(y_t) = E(JY_t) = JE(Y_t) = J\boldsymbol{\mu}$, las autocovarianzas como

$$\begin{aligned} \Gamma_y(h) &= E[(y_t - J\boldsymbol{\mu})(y_{t-h} - J\boldsymbol{\mu})'] \\ &= E[(JY_t - J\boldsymbol{\mu})(JY_{t-h} - J\boldsymbol{\mu})'] \\ &= E[J(Y_t - \boldsymbol{\mu})(Y_{t-h} - \boldsymbol{\mu})J'] \\ &= JE[(Y_t - \boldsymbol{\mu})(Y_{t-h} - \boldsymbol{\mu})]J' \\ &= J\Gamma_Y(h)J' \end{aligned} \tag{1.1.13}$$

y ambas son invariantes bajo el tiempo.

El polinomio $\det(I_{Kp} - A_z) = det(I_k - A_1 z - \cdots - A_p z^p)$ se conoce como *el polinomio característico del proceso VAR(p)*. Por lo tanto, el proceso (1.1.1) es estable si su polinomio característico no tiene raíces en y sobre el círculo unitario complejo. Formalmente, (y_t) es estable si

$$det(I_k - A_1 z_1 - \ldots - A_p z^p) \neq 0 \quad para \quad |z| \leq 1. \tag{1.1.14}$$

Esta condición es llamada condición de estabilidad.

En resumen, decimos que y_t es un proceso VAR(p) estable si (1.1.14) se cumple y

$$y_t = JY_t = J\boldsymbol{\mu} + J\sum_{i=0}^{\infty} \boldsymbol{A}^i U_{t-i}. \tag{1.1.15}$$

Debido a que $U_t := (u_t^{'}, 0 \ldots 0)^{'}$ está determinado por el proceso de ruido blanco u_t, (y_t) también lo está. Un ejemplo importante es el caso en que u_t es un ruido blanco Gaussiano, esto es, $u_t \sim \mathcal{N}(0, \Sigma_u)$ para todo t y, u_t y u_s son independientes para $s \neq t$. En este caso, puede mostrarse que (y_t) es un proceso Gaussiano, esto es, las subcolecciones $y_t, \ldots, y_{t+h}$ tiene distribución normal multivariada para toda t y h.

1.1.2. La representación de la media móvil del proceso VAR

Previamente se ha considerado la representación VAR(1)

$$Y_t = \boldsymbol{\nu} + \boldsymbol{A}_{t-1} + U_t,$$

del proceso VAR(p) (1.1.1). Bajo la supocisión de estabilidad, el proceso Y_t tiene una representación

$$Y_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{A}^i U_{t-i}. \tag{1.1.16}$$

Esta forma del proceso es llamada representación media móvil o representación MA.

Luego, la representación MA de (y_t) puede ser encontrada multiplicando (1.1.16) por la matriz $J := [I_k : 0 : \ldots : 0]$ definida en (1.1.12), como sigue

$$\begin{aligned} y_t &= JY_t \\ &= J\left[\boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{A}^i U_{t-i}\right] \\ &= J\boldsymbol{\mu} + J\sum_{i=0}^{\infty} \boldsymbol{A}^i U_{t-i} \\ &= J\boldsymbol{\mu} + \sum_{i=0}^{\infty} J\boldsymbol{A}^i J' J U_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}. \end{aligned} \tag{1.1.17}$$

Aquí $\mu := J\boldsymbol{\mu}$, $\Phi_i := J\boldsymbol{A}^i J'$. Además $U_t = J'JU_t$ y $JU_t = u_t$.

Usando la Proposición C.2, la representación (1.1.17) brinda la posibi-

lidad de determinar la media y las autocovarianzas de (y_t) como sigue,

$$\begin{aligned} E(y_t) &= E\left(\mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}\right) \\ &= E(\mu) + E(\sum_{i=0}^{\infty} \Phi_i u_{t-i}) \\ &= \mu + \lim_{n\to\infty} \sum_{i=0}^{n}\sum_{j=0}^{n} \Phi_i E(u_{t-i}u_{t-j})\Phi_i{}' \\ &= \mu, \end{aligned}$$

$$\begin{aligned} \Gamma_y(h) &= E[(y_t-\mu)(y_{t-h}-\mu)^{'})] \\ &= E[(\sum_{i=0}^{\infty}\Phi_i u_{t-i})(\sum_{i=0}^{\infty}\Phi_i u_{t-h-i})^{'}] \\ &= E[(\sum_{i=0}^{h-1}\Phi_i u_{t-i} + \sum_{i=0}^{\infty}\Phi_{h+i} u_{t-h-i})(\sum_{i=0}^{\infty}\Phi_i u_{t-h-i})^{'}] \\ &= E[(\sum_{i=0}^{h-1}\Phi_i u_{t-i})(\sum_{i=0}^{\infty}\Phi_i u_{t-h-i})^{'} + (\sum_{i=0}^{\infty}\Phi_{h+i} u_{t-h-i})(\sum_{i=0}^{\infty}\Phi_i u_{t-h-i})^{'}] \\ &= E[(\sum_{i=0}^{h-1}\Phi_i u_{t-i})(\sum_{i=0}^{\infty}\Phi_i u_{t-h-i})^{'}] + E\left[(\sum_{i=0}^{\infty}\Phi_{h+i} u_{t-h-i})(\sum_{i=0}^{\infty}\Phi_i u_{t-h-i})^{'}\right] \\ &= \sum_{i=0}^{\infty}\Phi_{h+i}\Sigma_u\Phi^{'}{}_i. \end{aligned} \tag{1.1.18}$$

Pero no es necesario calcular las matrices de coeficientes Φ_i a través de la representación VAR(1), como antes. Una forma más directa de determinar estas matrices es escribir el proceso VAR(p) en notación de operadores de retraso. El operador de retraso, L, es definido de tal manera que $Ly_t = y_{t-1}$, esto es, retrocede el índice en un periodo. Usando este operador, (1.1.1) puede escribirse como

$$\begin{aligned} y_t &= \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t \\ &= \nu + A_1 L y_t + \ldots + A_p L^p y_t + u_t \\ &= \nu + (A_1 L + \ldots + A_p L^p) y_t + u_t. \end{aligned}$$

Así,

$$y_t - (A_1 L + \ldots + A_p L^p) y_t = \nu + u_t, \tag{1.1.19}$$

esto es,

$$(I_k - A_1 L - A_2 L^2 - \ldots - A_p L^p) y_t = \nu + u_t \tag{1.1.20}$$

o

$$A(L) y_t = \nu + u_t, \tag{1.1.21}$$

donde $A(L) = I_k - A_1 L - \ldots - A_p L^p$.

Sea $\Phi(L) := \sum_{i=0}^{\infty} \Phi_i L^i$, entonces

$$\Phi(L) A(L) = I_K. \tag{1.1.22}$$

Multiplicando (1.1.21) por $\Phi(L)$, se obtiene que

$$\begin{aligned} y_t &= \Phi(L)\nu + \Phi(L) u_t \\ &= (\sum_{i=0}^{\infty} \Phi_i)\nu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}. \end{aligned} \tag{1.1.23}$$

Se dice que el operador $A(L)$ es invertible si $|A(z)| \neq 0$, para $|z| \neq 1$; y que el operador $\Phi(L)$ es el inverso de $A(L)$. Si esta condición se cumple, la matriz de coeficientes $\Phi(L) = A(L)^{-1}$ es absolutamente sumable, y entonces $\Phi(L) u_t = A(L)^{-1} u_t$ (ver Apéndice B).

Las matrices de coeficientes Φ_i pueden ser obtenidas como en (1.1.22) usando las relaciones

$$
\begin{aligned}
I_k &= (\Phi(L)A(L) \\
&= (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \ldots)(I_k - A_1 L \ldots - A_p L^p) \\
&= \Phi_0 - \Phi_0 A_1 L - \Phi_0 A_2 L^2 - \ldots - \Phi_0 A_p L^p + \Phi_1 L A_1 L - \Phi_1 L A_2 L^2 - \\
&\Phi_1 L A_p L^p + \ldots - \Phi_2 L^2 A_p L^p, \ldots, \\
&= \Phi_0 + (\Phi_1 - \Phi_0 A_1)L + (\Phi_2 - \Phi_1 A_1 - \Phi_0 A_2)L^2 + .. + (\Phi_i - \sum_{j=1}^{i} \Phi_{i-j} A_j) \\
&L^i + \ldots,
\end{aligned}
$$

o bien

$$
\begin{aligned}
I_k &= \Phi_0 \\
\overline{0} &= \Phi_1 - \Phi_0 A_1 \\
\overline{0} &= \Phi_2 - \Phi_1 A_1 - \Phi_0 A_2 \\
&\vdots \\
\overline{0} &= \Phi_i - (\Phi_i - \sum_{j=1}^{i} \Phi_{i-j} A_j),
\end{aligned}
$$

donde $A_j = 0$ para $j > p$. Por lo tanto, Φ_i puede ser calculada recursivamente usando

$$
\begin{aligned}
\Phi_0 &= I_k, \\
\Phi_i &= \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \ldots
\end{aligned}
\tag{1.1.24}
$$

La media μ puede ser obtenida como sigue:

$$
\mu = (I_k - A_1 - \ldots)^{-1} \nu = A(1)^{-1} \nu = \Phi(1)\nu. \tag{1.1.25}
$$

Para el proceso VAR(1), la recursión (1.1.24) implica que

$$\begin{aligned}
\Phi_0 &= I_k \\
\Phi_1 &= \textstyle\sum_{j=1}^{1} \Phi_{i-j} A_j = \Phi_0 A_1 = I_k A_1 = A \\
\Phi_2 &= \textstyle\sum_{j=1}^{2} \Phi_{i-j} A_j = \Phi_1 A_1 + \Phi_0 A2 = \Phi_1 A_1 = A_1 A_1 = {A_1}^2 \\
&\vdots \\
\Phi_i &= \textstyle\sum_{j=1}^{i} \Phi_{i-j} = {A_1}^i \\
&\vdots
\end{aligned}$$

Para un proceso VAR(2), la recursión (1.1.24) resulta en

$$\begin{aligned}
\Phi_1 &= A_1 \\
\Phi_2 &= \Phi_1 A_1 + A_2 = {A_1}^2 + A_2 \\
\Phi_3 &= \Phi_2 A_1 + \Phi_1 A_2 = {A_1}^3 + A_2 A_1 + A_1 A_2 \\
&\vdots \\
\Phi_i &= \Phi_{i-1} A_1 + \Phi_{i-2} A_2 \\
&\vdots
\end{aligned}$$

Vale la pena señalar que la representación MA de un proceso estable VAR(p) no necesariamente es de orden infinito. Esto es, las Φ_i pueden ser todas cero para i mayor que algún entero finito q.

1.1.3. Procesos Estacionarios

Para cumplir el objetivo principal, pronosticar, los procesos de interés son los procesos estables, por lo que ahora es necesario estudiar lo siguiente.

A continuación se enuncia una definición que complementa a la idea de estabilidad y dará pie a que los estimadores considerados más adelante tengan propiedades deseables tales como: consistencia y distribución normal.

Definición 1.1 (Proceso estacionario)**.** *Un proceso estocástico es estacionario si su primero y segundo momentos son invariantes, esto es, si*

$$E(y_t) = \mu$$

y

$$E[(y_t - \mu)(y_{t-h} - \mu)^{'}] = \Gamma_y(h) = \Gamma_y(-h)^{'},$$

para todo t *y* $h = 0, 1, 2, \ldots$.

A menudo se usan otras definiciones de estacionariedad. Por ejemplo, la que asume que la distribución conjunta de n vectores consecutivos es invariante en el tiempo para todo n. Sin embargo, aquí se usará la definición anterior. Por la definición 1.1, el proceso de ruido blanco en (1.1.1) es un ejemplo de un proceso estacionario. Además, de (1.1.18) se tiene que un proceso VAR(p) estable es estacionario. Este hecho se enuncia en la siguiente proposición.

Proposición 1.2 (Condición de estacionariedad)**.** *Todo proceso VAR(p) estable* (y_t)*, es estacionario.*

Demostración. Sea (y_t) un proceso estable, por (1.1.15) (y_t) es de la forma

$$(y_t) = JY_t = J\boldsymbol{\mu} + J\sum_{i=0}^{\infty} \boldsymbol{A}^i U^{t-i} = \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \tag{1.1.26}$$

entonces

$$\Gamma_y(h) = \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_u \Phi^{'}{}_i \tag{1.1.27}$$

(por (1.1.18)), así

$$\begin{aligned}\Gamma_y(-h) &= \sum_{i=0}^{\infty} \Phi_{-h+i}\Sigma_u\Phi_i^{'} \\ &= \left[\sum_{i=0}^{\infty} \Phi_i\Sigma_u\Phi_{-h+i}^{'}\right]^{'} \\ &= \left[\sum_{i=h}^{\infty} \Phi_i\Sigma_u\Phi_{-h+i}^{'}\right]^{'} \\ &= \Gamma_y(h)^{'}.\end{aligned}$$

De lo que se concluye que $\Gamma_y(h) = \Gamma_y(-h)^{'}$.

□

Como la estabilidad implica estacionariedad, la condición de estabilidad es frecuentemente referida como *condición de estacionariedad* en la literatura de series de tiempo. El recíproco de esta proposición no siempre es cierta. En otras palabras, un proceso inestable no necesariamente es un proceso no estacionaro [5].

1.2. Cálculo de autocovarianza y autocorrelación de procesos estables VAR

Aunque las autocovarianzas de un proceso estable estacionario pueden darse en términos de sus matrices de coeficentes como en (1.1.18), la fórmula no es viable en la práctica porque involucra sumas infinitas. Para fines prácticos es más fácil calcular las autocovarianzas directamente de las matrices de coeficientes.

1.2.1. Autocovarianzas de un proceso estable VAR(p)

Para ilustrar el cálculo de las autocovarianzas cuando en el proceso los coeficientes están dados, suponemos que (y_t) es un proceso VAR(p) estable

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \tag{1.2.1}$$

con matriz de covarianza $E(u_t u'_t) = \Sigma_u$. Alternativamente, el proceso se puede escribir en su forma con media ajustada como

$$y_t - \mu = A_1(y_{t-1} - \mu) + ... + A_p(y_{t-p} - \mu) + u_t \tag{1.2.2}$$

donde $\mu = E(y_t)$. Multiplicando por $(y_{t-h} - \mu)'$ tenemos

$$\begin{aligned}(y_t - \mu)(y_{t-h} - \mu)' = A_1(y_{t-1} - \mu)(y_{t-h} - \mu)' + .. + A_p(y_{t-p} - \mu)(y_{t-h} - \mu)' \\ + u_t(y_{t-h} - \mu)',\end{aligned} \tag{1.2.3}$$

y tomando esperanzas

$$\begin{aligned}\Gamma_y(h) &= E[(y_t - \mu)(y_{t-h} - \mu)'] \\ &= A_1 E[(y_{t-1} - \mu)(y_{t-h} - \mu)] + .. + A_p[(y_{t-p} - \mu)(y_{t-h} - \mu)'] \\ &+ E[u_t(y_{t-h} - \mu)' \\ &= A_1\Gamma_y(h-1) + ... + A_p\Gamma_y(h-p) + E[u_t(y_{t-h} - \mu)'].\end{aligned}$$

Si $h = 0$, entonces

$$\begin{aligned}\Gamma_y(0) &= A_1\Gamma_y(-1) + ... + A_p\Gamma_y(-p) + E[u_t(y_t - \mu)'] \\ &= A_1\Gamma_y(-1) + ... + A_p\Gamma_y(-p) + \Sigma_u \\ &= A_1\Gamma_y(1)' + ... + A_p\Gamma_y(p)' + \Sigma_u,\end{aligned}$$

pues $\Gamma_y(i) = \Gamma_y(-i)'$, ya que el proceso es estable y $E[u_t(y_{t-h} - \mu)'] = \Sigma_u$. Si $h > 0$, entonces

$$\Gamma_y(h) = A_1\Gamma_y(h-1) + ... + A_p\Gamma_y(h-p). \tag{1.2.4}$$

Estas ecuaciones son conocidas como *ecuaciones de Yule-Walker* y pueden ser usadas para calcular $\Sigma_y(h)$ recursivamente para $h \geq p$, siempre que $A_1, A_2, ...A_p$ y $\Gamma_y(p-1), ..., \Gamma_y(0)$ son conocidas.

Las matrices de autocovarianzas iniciales para $|h| < p$ pueden ser determinadas usando el proceso VAR(1) correspondiente a (1.2.1)

$$Y_t - \boldsymbol{\mu} = \boldsymbol{A}(Y_{t-1} - \boldsymbol{\mu}) + U_t, \tag{1.2.5}$$

donde Y_t, $\mathbf{A}$ y U_t son como en (1.1.9) y $\boldsymbol{\mu} := (\mu', ..., \mu') = E(Y_t)$. Procediendo de manera análoga al caso anterior, se obtiene que para $h = 0$

$$\Gamma_Y(0) = \boldsymbol{A}\Gamma_Y(1)' + \Sigma_U \tag{1.2.6}$$

y para $h > 0$

$$\Gamma_Y(h) = \boldsymbol{A}\Gamma_Y(h-1). \tag{1.2.7}$$

Si $A_1, \ldots, A_p$ y $\Gamma_Y(0)$ son conocidas, entonces $\Gamma_Y(h)$ puede calcularse recursivamente usando (1.2.7). Si $A_1, \ldots, A_p$ y Σ_U son conocidas, $\Gamma_Y(0)$ puede ser calculada como sigue. Para $h = 1$, de (1.2.7) se tiene que $\Gamma_Y(1) = \boldsymbol{A}\Gamma_Y(0)$. Sustituyendo $\boldsymbol{A}\Gamma_Y(0)$ por $\Gamma_Y(1)$ en (1.2.6), tenemos

$$\Gamma_Y(0) = \boldsymbol{A'}\Gamma_Y(0)\boldsymbol{A'} + \Sigma_U, \tag{1.2.8}$$

o bien

$$\begin{aligned} vec\Gamma_Y(0) &= vec(\boldsymbol{A}\Gamma_Y(0)\boldsymbol{A'} + vec\Sigma_U \\ &= (\boldsymbol{A} \otimes \boldsymbol{A})vec\Gamma_Y(0) + vec\Sigma_U, \end{aligned}$$

(por la regla (1) y (2) del Apéndice A.5), donde *vec* es el operador de apilamiento de columnas como se define en el Apéndice A.5, $\Sigma_u = E(U_t U_t^{'})$ y

$$\begin{aligned} \Gamma_Y(0) &= E\left(\begin{bmatrix} y_t - \mu \\ \vdots \\ y_{t-p-1} - \mu \end{bmatrix}\right)[(y_t - \mu)^{'}, ..., (y_{t-p+1} - \mu)^{'}] \\ &= \begin{bmatrix} \Gamma_y(0) & \Gamma_y(1) & \cdots & \Gamma_y(p-1) \\ \Gamma_y(-1) & \Gamma_y(0) & \cdots & \Gamma_y(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_y(-p+1) & \Gamma_y(-p+2) & \cdots & \Gamma_y(0) \end{bmatrix}. \end{aligned} \tag{1.2.9}$$

Así, $\Gamma_y(h), h = -p+1, ..., p-1$, son obtenidos de

$$vec\Gamma_y(0) = (I_{(Kp)^2} - \boldsymbol{A} \otimes \boldsymbol{A})^{-1} vec\Sigma_U. \tag{1.2.10}$$

1.2.2. Autocorrelaciones de un proceso estable VAR(p)

A veces las autocovarianzas son difíciles de interpretar. Por lo tanto las autocorrelaciones

$$R_y = D^{-1}\Gamma_(h)D^{-1}, \tag{1.2.11}$$

son usualmente más convenientes para trabajar. Aquí, D es una matriz diagonal con las desviaciones estándar de las componentes de (y_t) en la diagonal principal. Esto es, los elementos de la diagonal de D son las raíces cuadradas

de los elementos de la diagonal de $\Gamma_y(0)$. Denotando la covarianza entre $y_{i,t}$ y $y_{i,t-h}$ por $\gamma_{ij}(h)$ los elementos de la diagonal, $\gamma_{11}(0), ..., \gamma_{KK}(0)$ son varianzas de $y_{1t}, ..., y_{Kt}$, respectivamente. Así

$$D^{-1} = \begin{bmatrix} 1/\sqrt{\gamma_{11}(0)} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\gamma_{KK}(0)} \end{bmatrix} \tag{1.2.12}$$

y la correlación entre $y_{i,t}$ y $y_{j,t-h}$ es

$$\rho_{ij}(h) = \frac{\gamma ii(h)}{\sqrt{\gamma jj(h)}}, \tag{1.2.13}$$

donde $\rho_{ij}(h)$ es el ij-ésimo elemento de $R_y(h)$.

Capítulo 2

Pronosticación

Como ya se ha mencionado, pronosticar es uno de los objetivos principales del análisis de series de tiempo múltiples. Por lo tanto, ahora se abordarán algunos predictores.

El investigador usualmente se encuentra en una situación en la tiene que hacer afirmaciones sobre los valores futuros de las variables $y_1, \ldots, y_K$, para un periodo en particular. Para ello tiene disponible un modelo para el proceso de generación de datos y un conjunto, digamos Ω_t , que contiene la información disponible hasta el periodo t. El proceso de generación de datos puede ser, por ejemplo un proceso VAR(p) y Ω_t puede contener las variables pasadas y presentes del sistema en consideración, esto es $\Omega_t = \{y_s | s \leq t\}$, donde $y_s = (y_{1s}, \ldots, y_{Ks})'$.

En el contexto de los modelos VAR, los predictores que minimizan el pronóstico del error cuadrático medio (MSE) son los más utilizados [5].

2.1. Pronosticación puntual

Supongamos $(y_t) = (y_{1t}, .., y_{Kt})'$ es un proceso VAR(p) estable de dimensión K. Entonces, el predictor de error cudrático medio mínimo (MSE) para el pronóstico de horizonte h, en el origen del pronóstico t, es el valor esperado condicional

$$E_t(y_{t+h}) := E(y_{t+h}|\Omega_t) = E(y_{t+h}|\{y_s|s \leq t\}). \tag{2.1.1}$$

Este estimador minimiza el error cuadrático medio para cada componente de (y_t). En otras palabras, si $(y_t)(h)$ es cualquier predictor de paso h en el

origen t,

$$\begin{aligned} MSE[\overline{y}_t(h)] &= E[(y_{t+h} - \overline{y}_t(h))(y_{t+h} - \overline{y}_t(h))'] \\ &\geq MSE[E_t(y_{t+h})] = E[(y_{t+h} - E_t(y_{t+h}))(y_{t+h} - E_t(y_{t+h}))'], \end{aligned} \tag{2.1.2}$$

donde el signo de desigualdad entre dos matrices significa que la diferencia entre la matriz izquierda y derecha es semidefinida positiva.

La optimización de la esperanza condicional se ve reflejada en lo siguiente,

$$\begin{aligned} MSE[\overline{y}_t(h)] &= E\{[y_{t+h} - E_t(y_{t+h}) + E_t(y_{t+h})\overline{y}_t(h)] \\ &\quad \times [y_{t+h} - E_t(y_{t+h}) + E_t(y_{t+h}) - \overline{y}_t(h)]'\} \\ &= MSE[E_t(y_{t+h})] + E\left\{[E_t(y_{t+h} - \overline{y}_t(h))][E_t(y_{t+h} - \overline{y}_t(h))]'\right\}, \end{aligned}$$

pues $E\left\{[y_{t+h} - E(y_{t+h})]\,[E_t(y_{t+h}) - \overline{y}_t(h)]'\right\} = 0$, ya que $[y_{t+h} - E_t\,(y_{t+h})]$ es una función de innovaciones después del periodo t que no están correlacionadas con los términos contenidos en $[E_t(y_{t+h}) - \overline{y}_t(h)]$, las cuales son funciones de $y_s, s \leq t$.

La optimización de la esperanza condicional implica que

$$E_t(y_{t+h}) = \nu + A_1 E_t(y_{t+h-1}) + ... + A_p E_t(y_{t+h-p}) \tag{2.1.3}$$

es el estimador óptimo de paso h de un proceso VAR(p), siempre que u_t sea ruido blanco independiente de modo que $s \neq t$, y por lo tanto, $E_t(u_{t+h}) = 0$ para $h > 0$.

La fórmula (2.1.3) puede ser usada para calcular recursivamente predictores de paso h, empezando con $h = 1$

$$\begin{aligned} E_t(y_{t+1}) &= \nu + A_1 y_t + ... + A_p y_{t-p+1}, \\ E_t(y_{t+2}) &= \nu + A_1 E_t(y_{t+1} + A_2 y_t) + ... + A_p y_{t-p+2}, \\ &\vdots \end{aligned}$$

Por esta recursión tenemos que para un proceso VAR(1)

$$\begin{aligned}
E_t(y_{t+1}) &= \nu + A_1 y_t, \\
E_t(y_{t+2}) &= \nu + A_1 E_t(y_{t+1}) = \nu + A_1(\nu + A_1 y_t) \\
&= \nu + A_1\nu + A_1^2 y_t = (I_k + A_1)\nu + A_1^2 y_t, \\
&\vdots \\
E_t(y_{t+h}) &= (I_k + A_1 + ... + A_1^{h-1})\nu + {A_1}^h y_t.
\end{aligned}$$

La esperanza condicional tiene las siguientes propiedades:

(1) Es un estimador insesgado, esto es, $E[y_{t+h} - E_t(y_{t+h})] = 0$.

(2) Si u_t es ruido blanco independiente,

$$MSE[E_t(y_{t+h})] = MSE[E_t(y_{t+h})|y_t, y_{t-1}, ...]. \tag{2.1.4}$$

Por otro lado, si u_t es ruido blanco no independiente, usualmente se requieren suposiciones adicionales para encontrar el predictor óptimo de un proceso VAR(p). Considerando un proceso VAR(1)

$$y_t = A_1 y_{t-1} + u_t. \tag{2.1.5}$$

Como en (1.1.3), se tiene que

$$\begin{aligned}
y_{t+1} &= A_1 y_t + u_{t+1}, \\
y_{t+2} &= A_1 y_{t+1} + u_{t+2} = A_1(A_1 y_t + u_{t+1}), \\
&= A_1^2 y_t + A_1 u_{t+1} + u_{t+2}, \\
y_{t+3} &= A_1 y_{t+2} + u_{t+3} = A_1(A_1^2 y_t + A_1 u_{t+1} + u_{t-2}) + u_{t+3}, \\
&= A_1^3 A_1^2 u_{t+1} + A_1 u_{t+2} + u_{t+3}, \\
&\vdots \\
y_{t+h} &= A_1^h + \sum_{i=0}^{h-1} A_1^i u_{t+h-i}.
\end{aligned}$$

Entonces, para un predictor

$$y_t(h) = B_0 y_t + B_1 y_{t-1} + \cdots,$$

donde cada B_i es una matriz cuadrada de dimensión K, se obtiene el error de pronóstico

$$\begin{aligned} y_{t+h} - y_t(h) &= A^h{}_1 y_t + \sum_{i=0}^{h-1} A^i{}_1 u_{t+h-i} - [B_0 y_t + B_1 y_{t-1} + \cdots] \\ &= \sum_{i=0}^{h-1} A^i{}_1 u_{t+h-i} + (A^h{}_1 - B_0) y_t - \sum_{i=1}^{\infty} B_i y_{t-i}. \end{aligned}$$

Suponiendo que u_{t+j}, para $j > 0$, no está correlacionado con y_{t-i}, para $i \geq 0$, se obtiene

$$\begin{aligned} & MSE[y_t(h)] \\ &= E\left[\left(\sum_{i=0}^{h-1} A^i{}_1 u_{t+h-i}\right)\left(\sum_{i=0}^{h-1} A^i{}_1 u_{t+h-i}\right)'\right] \\ &+ E\left\{\left[(A^h{}_1 - B_0) y_t - \sum_{i=1}^{\infty} B_{i1} y_{t-i}\right]\left[(A^h{}_1 - B_0) y_t - \sum_{i=1}^{\infty} B_{i1} y_{t-i}\right]'\right\}. \end{aligned}$$

Esta matriz se minimiza cuando $B_0 = A_1^h$ y $B_i = 0$. Entonces, el predictor óptimo para este caso especial es

$$y_t(h) = A_1^h y_t = A_1 y_t(h-1). \tag{2.1.6}$$

El error de pronóstico está dado por

$$\sum_{i=0}^{h-1} A^i{}_1 u_{t+h-i} \tag{2.1.7}$$

y la covarianza del error del pronóstico es

$$\begin{aligned} \Sigma_y(h): &= MSE[y_t(h)] = E\left[\left(\sum_{i=0}^{h-1} A^i{}_1 u_{t+h-i}\right)\left(\sum_{i=0}^{h-1} A^i{}_1 u_{t+h-i}\right)'\right] \\ &= \sum_{i=0}^{h-1} A^i{}_1 \Sigma_u A^i{}_1{}' = MSE[y_t(h-1)] + A_1{}^{h-1} \Sigma_u (A_1{}^{h-1})', \end{aligned}$$

pues $E[(\sum_{i=0}^{h-2} A_1^i u_{t+h-1-i})(A_1 u_{t+h-i})'] = 0 = E[(A_1 u_{t+h-i})(\sum_{i=0}^{h-2} A_1^i u_{t+h-1-i})']$, porque $\sum_{i=0}^{h-2} A_1^i u_{t+h-1-i}$ y $A_1 u_{t+h-i}$ no están correlacionados.

En general, como un proceso $VAR(p)$ con media cero

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \tag{2.1.8}$$

puede escribirse como

$$Y_t = \boldsymbol{A} Y_{t-1} + U_t, \tag{2.1.9}$$

donde (Y_t), $\boldsymbol{A}$ y U_t son como se definieron en (1.1.9). Usando argumentos análogos a los del caso $VAR(1)$, el predictor óptimo resultante para Y_{t+h} es

$$Y_t(h) = \boldsymbol{A}^h Y_t = \boldsymbol{A} Y_t(h-1). \tag{2.1.10}$$

Haciendo inducción con respecto a h, se tiene que

$$Y_t(h) = \begin{bmatrix} y_t(h) \\ y_t(h-1) \\ \vdots \\ y_t(h-p+1) \end{bmatrix},$$

donde $y_t(j) := y_{t+j}$ para $j \leq 0$. Considerando nuevamente la matriz $J := [I_K : 0 : \cdots : 0]$ de dimensión $(K \times K_p)$, se obtiene que el predictor óptimo de paso h del proceso y_t es

$$\begin{aligned} y_t(h) &= J\boldsymbol{A} Y_t(h-1) = [A_1, ..., A_p] Y_t(h-1) \\ &= A_1 y_t(h-1) + ... + A_p y_t(h-p). \end{aligned}$$

Esta fórmula puede ser usada para calcular los pronósticos recursivamente. Es importante notar que $y_t(h)$ es la esperanza condicional $E_t(y_{t+h})$ si u_t es ruido blanco independiente, porque la recursión en (2.1.3) es la misma que la obtenida aquí para un proceso de media cero con $\nu = 0$.

Si el proceso y_t tiene media distinta de cero, esto es,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

se define $x_t := y_t - \mu$, donde $\mu := E(y_t) = (I - A_1 - \cdots - A_p)^{-1}\nu$. El proceso x_t tiene media cero y el predictor óptimo de paso h es

$$x_t(h) = A_1 x_t(h-1) + \cdots + A_p x_t(h-p).$$

Sumando μ a ambos lados de la ecuación se tiene el predictor óptimo de y_t

$$\begin{aligned} y_t(h) &= x_t(h) + \mu = \mu + A_1(y_t(h-1) - \mu) + \cdots + A_p(y_t(h-p) - \mu) \\ &= \nu + A_1 y_y(h-1) + \cdots + A_p y_t(h-p). \end{aligned} \tag{2.1.11}$$

De ahora en adelante, se hará referencia a $y_t(h)$ como el *predictor óptimo* independiente de las propiedades del proceso de ruido blanco u_t, esto es, incluso si u_t no es independiente sino sólo ruido blanco no correlacionado.

Usando

$$Y_t = \boldsymbol{A}^h Y_t + \sum_{i=0}^{h-1} \boldsymbol{A}^i U_{t+h-i}$$

para un proceso de media cero, se obtiene el error de pronóstico

$$\begin{aligned} y_{t+h} - y_t(h) = J[Y_{t+h} - Y_(h)] &= J\left[\sum_{i=0}^{h-1} \boldsymbol{A}^i U_{t+h-i}\right] \\ &= \sum_{i=0}^{h-1} J\boldsymbol{A}^i J' J U_{t+h-i} = \sum_{i=0}^{h-1} \Phi_i u_{y+h-i}, \end{aligned} \tag{2.1.12}$$

donde las Φ_i son como en (1.1.17). El error de pronóstico no cambia si tiene una media distinta de cero porque el sumando asociado a la media se cancela. La representación del error de pronóstico (2.1.12) muestra que el predictor $y_t(h)$ puede ser expresado en términos de la representación MA (1.1.17) como

$$y_t(h) = \mu + \sum_{i=h}^{\infty} \Phi_i u_{t+h-i} = \mu + \sum_{i=0}^{\infty} \Phi_{h+i} u_{t-i}. \tag{2.1.13}$$

De (2.1.12) se obtiene que, la matriz de covarianza del error del pronóstico está dada por

$$\Sigma_y(h) := MSE[y_t(h)] = \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i' = \Sigma_y(h-1) + \Phi_{h-1} \Sigma_u \Phi_{h-1}'. \tag{2.1.14}$$

Entonces, las varianzas de los errores de predicción son monótonoamente crecientes, y cuando $h \to \infty$, se aproximan a la matriz de covarianza del proceso (y_t),

$$\Gamma_y(0) = \Sigma_y = \sum_{i=0}^{\infty} \Phi_i \Sigma_u \Phi_i', \tag{2.1.15}$$

ver (1.1.18). Esto es,

$$\Sigma_y(h) \longrightarrow \Sigma_y \tag{2.1.16}$$

cuando $h \to \infty$. Lo cual significa que $\Sigma_y(h)$ está cotada por Σ_y.

2.2. Pronósticos por intervalo y regiones de pronóstico

Para estudiar los pronósticos hacemos una suposición sobre las distribuciones de y_t o de u_t. Suponiendo que $u_t \sim N(\mu, \Sigma_u)$ y, que u_t y u_s son independientes para $s \neq t$, los errores son normalmente distribuidos

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \sim N(0, \Sigma_y(h)). \tag{2.2.1}$$

Este resultado implica que los errores de pronóstico de los componentes individuales son normales(por Proposición (C.4)), así que

$$\frac{y_{k,t+h} - y_{k,t}(h)}{\sigma_k(h)} \sim N(0,1), \tag{2.2.2}$$

donde $y_{k,t}(h)$ es la $k-$ésima componente de $y_t(h)$ y $\sigma_k(h)$ es la raíz cuadrada del $k-$ésimo elemento de la diagonal de $\Sigma_y(h)$. Seleccionando dos valores de las colas de esta distribución, $z_{\alpha/2}$ y $-z_{\alpha/2}$, tales que

$$\mathcal{P}\left\{-z_{\alpha/2} \leq \frac{y_{k,t+h} - y_{k,t}(h)}{\sigma_k(h)} \leq z_{\alpha/2}\right\} = 1 - \alpha,$$

y haciendo los cálculos necesarios tenemos que,

$$1 - \alpha = \mathcal{P}\left\{y_{k,t}(h) - z_{\alpha/2}\sigma_k(h) \leq y_{k,t+h} \leq y_{k,t}(h) + z_{\alpha/2}\sigma_k(h)\right\}$$

donde $1 - \alpha$ es el coeficiente de confianza. Por lo tanto, un pronóstico por intervalo de $1 - \alpha$, de paso h, para la $k-$ésima componente de (y_t) es

$$y_{k,t}(h) \pm z_{\alpha/2}\sigma_k(h) \tag{2.2.3}$$

o bien

$$\left[y_{k,t}(h) - z_{\alpha/2}\sigma_k(h), y_{k,t}(h) + z_{\alpha/2}\sigma_k(h)\right]. \tag{2.2.4}$$

La amplitud de los intervalos de predicción va creciendo conforme h crece. Si los intervalos de este tipo son calculados repetidamente a partir de un gran número de series de tiempo, entonces aproximadamente el $(1-\alpha)\,\%$ de los intervalos contendrán el valor real de la variable aleatoria $y_{k,t+h}$.

El resultado en (2.2.1) también se puede usar para establecer regiones de pronóstico conjuntas para dos o más variables. Por ejemplo, si se desea una región de pronóstico de N componentes, definimos la matriz $F := [I_N : 0]$ y con ayuda de (2) de la Proposición (C.4), tenemos que

$$[y_{t+h} - y_t(h)]' F' [F\Sigma_y(h)F']^{-1} F[y_{t+h} - y_t(h)] \sim \chi^2(N). \tag{2.2.5}$$

Por lo tanto, la distribución $\chi^2(N)$ se puede usar para determinar un pronóstico del $100\,\%(1-\alpha)$ para los primeros N componentes del proceso. Pero en la práctica la construcción es bastante exigente si N es mayor que dos o tres. Por lo que puede usarse un enfoque aproximado para construir regiones de confianza conjuntas. Para ello nos basamos en el hecho de que para los eventos $E_1, \ldots, E_n$ la siguiente desigualdad se cumple

$$\mathcal{P}(E_1 \cup \ldots E_n) \leq \mathcal{P}(E_1) + \ldots + \mathcal{P}E_n).$$

Por lo tanto

$$\mathcal{P}\left(\bigcap_{i=1}^{N} E_i\right) = 1 - \mathcal{P}\left(\bigcup_{i=1}^{N} \bar{E}_i\right) \geq 1 - \sum_{i=1}^{N} \mathcal{P}\left(\bar{E}_i\right)$$

Consecuentemente, si E_i es el evento en el que $y_{i,t+h}$ cae dentro de un intervalo H_i, se tiene que

$$\mathcal{P}\left(Fy_{t+h} \in H_1 \times \ldots H_N\right) \geq 1 - \sum_{i=1}^{N} \mathcal{P}(\bar{E}_i). \tag{2.2.6}$$

En otras palabras, si se quiere calcular un intervalo de predicción del $\left(1 - \frac{\alpha}{N}\right)\,\%$ para cada una de las N componentes, el resultado de la región de pronóstico conjunta tiene la probabilidad de $(1-\alpha)\,\%$ de contener a las N variables conjuntamente.

Capítulo 3

Estimación para un proceso autorregresivo múltidimesional

En este capítulo se asumirá que una serie múltiple $K-$dimensional $y_1, \ldots, y_T$, con $y_t = (y_{1t}, \ldots, y_{kt})'$ está disponible y se sabe que es generada por

$$y_t = \nu + A_1 y_{y-1} + \ldots + A_p y_{t-p} + u_t. \tag{3.0.1}$$

En contraste con las suposiciones que se hicieron en el capítulo anterior, ahora los coeficientes $\nu, A_1, \ldots, A_p$ y Σ_u se suponen desconocidos, motivo por el cual las secciones siguientes se discuten diferentes posibilidades para estimar un proceso VAR(p).

3.1. Estimación multivariada por mínimos cuadrados

En esta sección se discute la estimación múltiple por mínimos cuadrados (LS). Se considera el estimador obtenido para la forma estándar (3.0.1) de un proceso $VAR(p)$ y se enuncian algunas propiedades del estimador.

Suponga que una serie de tiempo $y_1, \ldots, y_T$ está disponible, es decir, se tiene una muestra de tamaño T para cada una de las K variables. Además, suponga que están disponibles p valores de muestra para cada variable, $y_{-p+1}, \ldots, y_0$. Es conveniente particionar una serie de tiempo en valores de muestra y pre-muestra para simplificar. Para esto, se considera la siguiente notación

$$\begin{aligned}
Y &:= (y_1, \ldots, y_T) \qquad (K \times T),\\
B &:= (\nu, A_1, \ldots, A_p) \qquad (K \times (Kp+1)),\\
Z &:= (Z_0, \ldots, Z_{T-1}) \qquad ((Kp+1) \times T),\\
U &:= (u_1, \ldots, u_T) \qquad (K \times T),\\
\boldsymbol{y} &:= vec(Y) \qquad (KT \times 1),\\
\beta &:= vec(B) \qquad ((K^2p+K) \times 1)\\
\boldsymbol{b} &:= vec(B') \qquad ((K^2p+K) \times 1)\\
\boldsymbol{u} &:= vec(U) \qquad (KT \times 1),
\end{aligned} \tag{3.1.1}$$

$$Z_t = \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} ((Kp+1) \times 1).$$

Aquí *vec* es el operador de apilamiento de columnas y los términos que siguen a cada una de las definiciones anteriores, representan las dimensiones. Entonces, el modelo $VAR(p)$ puede ser escrito de forma compacta como

$$Y = BZ + U \tag{3.1.2}$$

o

$$\begin{aligned}
vec(Y) &= vec(BZ) + vec(U)\\
&= (Z' \otimes I_K)vec(B) + vec(U)
\end{aligned}$$

o

$$y = (Z' \times I_K)\beta + u, \tag{3.1.3}$$

donde $\otimes$ señala el producto de Kronecker.

Note que la matriz de covarianza de $\boldsymbol{u}$ es

$$\Sigma_{\boldsymbol{u}} = I_T \otimes \Sigma_u. \tag{3.1.4}$$

Por lo tanto, la estimación LS multivariada de β se basa en elegir el estimador que minimiza

$$\begin{aligned}
S(\beta) &= \boldsymbol{u'}(I_T \otimes \Sigma_u)^{-1}\boldsymbol{u} = \boldsymbol{u'}(I_T \otimes \Sigma_u)\boldsymbol{u}\\
&= [\boldsymbol{y} - (Z' \otimes I_K)\boldsymbol{\beta}]'(I_K \otimes {\Sigma_u}^{-1}[\boldsymbol{y} - (Z' \otimes I_K)\boldsymbol{\beta}]\\
&= vec(Y - BZ)'(I_T \otimes \Sigma_u^{-1})vec(Y - BZ)\\
&= tr[(Y - BZ)'\Sigma_u^{-1}(Y - BZ)],
\end{aligned} \tag{3.1.5}$$

usando las Reglas 5 y 6 del Apéndice A.5 y la ecuación (3.1.3).

Para encontrar el mínimo de esta función, partimos del término de la segunda igualdad de (3.1.5), así que

$$\begin{aligned} S(\beta) &= \boldsymbol{y}'(I_t \otimes \Sigma_u^{-1})\boldsymbol{y} + \beta^{'}(Z \otimes I_K)(I_T\Sigma_u^{-1})(Z' \otimes I_K)\beta \\ &\quad - 2\beta^{'}(Z \otimes I_K)(I_T \otimes \otimes\Sigma_u^{-1})\boldsymbol{y} \\ &= \boldsymbol{y}'(I_T \otimes \Sigma_u^{-1})\boldsymbol{y} + \beta'(ZZ' \otimes \Sigma_u^{-1})\beta - 2\beta'(Z \otimes \Sigma_u^{-1})\boldsymbol{y}, \end{aligned} \tag{3.1.6}$$

usando la Regla 4 del Apéndice A.4.
Luego por las Reglas de derivación del Apéndice A.6, se tiene que

$$\frac{\partial S(\beta)}{\partial x} = 2(zz' \otimes \Sigma_u^{-1})\beta - 2(Z \otimes \Sigma_u^{-1})\boldsymbol{y}.$$

Igualando a cero las ecuaciones normales, se tiene que

$$(ZZ' \otimes \Sigma_u^{-1})\widehat{\beta} = (Z \oplus \Sigma_u^{-1})\boldsymbol{y} \tag{3.1.7}$$

y consecuentemente, el estimador LS es

$$\begin{aligned} \widehat{\beta} &= ((ZZ')^{-1} \otimes \Sigma_u^{-1}\boldsymbol{y}) \\ &= ((ZZ')^{-1}Z \otimes I_K)\boldsymbol{y}. \end{aligned} \tag{3.1.8}$$

La matriz Hessiana de $S(\beta)$ dada por

$$\frac{\partial^2 S}{\partial\beta\partial\beta'} = 2(ZZ' \otimes \Sigma_u^{-1})$$

es definida positiva lo que confirma que $\widehat{\beta}$ es un vector minimizador. Para que estos resultados se cumplan se debe suponer que ZZ' es no singular.

El estimador LS puede escribirse de formas diferentes que serán de gran utilidad más adelante:

$$\begin{aligned} \widehat{\beta} &= ((ZZ')^{-1}Z \otimes I_K)[(Z' \otimes I_K)\beta + \boldsymbol{u}] \\ &= \beta + ((ZZ')^{-1}Z \otimes I_K)\boldsymbol{u} \end{aligned} \tag{3.1.9}$$

o

$$\begin{aligned} vec(\widehat{B}) = \widehat{\beta} &= ((ZZ')^{-1}Z \otimes I_K)vec(Y) \\ &= vec(YZ'(ZZ')^{-1}). \end{aligned} \tag{3.1.10}$$

Por lo tanto,

$$\begin{aligned}\widehat{B} &= YZ'(ZZ')^{-1} \\ &= (BZ+U)Z'(ZZ')^{-1} \\ &= B + UZ'(ZZ')^{-1}.\end{aligned} \tag{3.1.11}$$

Otra posibilidad para hallar este estimador resulta de multiplicar $y_t = BZ_{t-1} + u_t$ por $Z_{t-1}^{'}$ y tomando esperanzas. Así, se tiene que:

$$E(y_t Z_{t-1}^{'}) = BE(Z_{t-1}Z_{t-1}^{'}.) \tag{3.1.12}$$

Estimando $E(y_t Z_{t-1}^{'})$ por

$$\frac{1}{T}\sum_{t=1}^{T} y_t Z_{t-1}^{'} = \frac{1}{T} YZ'$$

y $E(Z_{t-1}Z_{t-1}^{'}) = \frac{1}{T}$ por,

$$\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}Z'_{t-1} = \frac{1}{T} ZZ',$$

se obtienen las ecuaciones normales

$$\frac{1}{T}YZ' = \widehat{B}\frac{1}{T}, \tag{3.1.13}$$

y por lo tanto,

$$\widehat{B} = YZ'(ZZ')^{-1}. \tag{3.1.14}$$

Así, otra posibilidad de escribir el estimador LS es

$$\widehat{b} = vec(\widehat{B}') = (I_K \otimes (ZZ)^{-1}Z)vec(Y'). \tag{3.1.15}$$

3.2. Propiedades asintóticas del estimador de mínimos cuadrados

Como las propiedades de muestras pequeñas del estimador por mínimos cuadrados (LS) son difíciles de derivar analíticamente, se analizarán las

propiedades asintóticas. La consistencia y la normalidad asintótica del estimador por mínimos cuadrados (LS) se cumplen fácilmente si se satisafacen los siguientes resultados:

$$\Gamma := plim ZZ'/T \tag{3.2.1}$$

existe y es no singular; donde *plim* denota convergencia en probabilidad. Y

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} vec(u_t Z_{t-1}') = \frac{1}{\sqrt{T}} vec(UZ') = \frac{1}{\sqrt{Z \otimes I_K}U} \xrightarrow{d} \mathcal{N}(0, \Gamma \otimes \Sigma_u) \tag{3.2.2}$$

cuando $T \to \infty$ y $\xrightarrow{d}$ denota la convergencia en distribución. Las condiciones establecidas en la siguiente definición son suficientes para que las condiciones anteriores se cumplan.

Definición 3.1 (Ruido blanco estándar)**.** *Un proceso de ruido blanco* $u_t = (u_{1t}, \cdots, u_{Kt})'$ *es llamado ruido blanco estándar si los vectores* u_t *aleatorios continuos satisfacen que* $E(u_t) = 0$, $\Sigma_u = E(u_t u_t')$ *es no singular,* u_t *y* u_s *son independientes para* $s \neq t$ *y para alguna constante finita* c*, se satisface que* $E|u_{it}u_{jt}u_{kt}u_{mt}| \leqslant c$*, para* $i, j, k, m = 1, \ldots, K$ *y toda* t*.*

La última condición implica que todos los cuartos momentos existen y están acotados, lo que servirá para la demostración del Corolario 3.5.

Con esta definición es fácil establecer condiciones de consistencia y normalidad asintótica del estimador por mínimos cuadrados (LS). El siguiente lema será útil para probar estas condiciones.

Lema 3.2. *Si* y_t *es un proceso* $VAR(p)$ *estable de dimensión* K*, como en* (3.0.1)*, con error de ruido blanco estandar* u_t*, entonces* (3.2.1) *y* (3.2.2) *se cumplen.*

Proposición 3.3 (Propiedades asintóticas del estimador LS)**.** *Sea* y_t *un proceso* $VAR(p)$ *estable de dimensión* K*, como en* (3.0.1)*, con error de ruido blanco estándar y* $\widehat{B} = YZ'(ZZ')^{-1}$ *el estimador LS de* B*. Entonces*

$$plim \widehat{B} = B \tag{3.2.3}$$

y

$$\sqrt{T}(\widehat{B} - B) = \sqrt{T} vec(\widehat{B} - B) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_u) \tag{3.2.4}$$

o equivalentemente

$$\sqrt{T}(\widehat{b}-b)=\sqrt{T}vec(\widehat{B'}-B')\xrightarrow{d}\mathcal{N}(0,\Sigma_u\otimes\Gamma^{-1}), \tag{3.2.5}$$

donde $\Gamma = plim ZZ'/T$, *plim denota convergencia en probabilidad y* $\xrightarrow{d}$ *denota convergencia en distribución.*

Demostración. Usando (4.2.3), se tiene que

$$plim\widehat{B}-B=plim\left(\frac{UZ'}{T}plim(\frac{ZZ'}{T})^{-1}\right)=0,$$

por el Lema 3.2, pues (3.2.2) implica que $plim(\frac{ZZ'}{T})=0$.

Usando (3.1.9), se tiene que

$$\begin{aligned}\sqrt{T}(\widehat{\beta}-\beta)&=\sqrt{T}((ZZ')^{-1}Z\otimes I_K)\boldsymbol{u}\\&=\left((\frac{1}{T}ZZ')^{-1}\otimes I_K\right)\frac{1}{\sqrt{T}}(Z\otimes I_K)\boldsymbol{u}.\end{aligned} \tag{3.2.6}$$

Por lo tanto, por la regla (4) de la Proposición B.1 , se tiene que $\sqrt{T}(\widehat{\beta}-\beta)$ tiene la misma distribución asintótica que

$$\left[plim(\frac{1}{T})^{-1}\otimes I_K\right]\frac{1}{\sqrt{T}}(Z\otimes I_K)\boldsymbol{u}=(\Gamma^{-1}\otimes I_K)\frac{1}{\sqrt{T}}(Z\otimes I_K)\boldsymbol{u}.$$

Entonces, por el Lema 3.2, la distribución asintótica de $\sqrt{T}(\widehat{\beta}-\beta)$ es normal y la matriz de covarianza es

$$(\Gamma^{-1}\otimes I_K)(\Gamma\otimes\Gamma_u)(\Gamma^{-1}\otimes I_K)=\Gamma^{-1}\otimes\Sigma_u.$$

El resultado (3.2.5) se demuestra análogamente. □

Como se mencionó anteriormente, si u_t es un ruido blanco Gaussiano, satisface las condiciones de la Proposición 3.3, de modo que la consistencia y la normalidad asintótica del estimador LS se cumplen para procesos $VAR(p)$ estables Gaussianos.

Para evaluar la dispersión asintótica del estimador LS se necesita conocer (o al menos estimar) las matrices Γ y Σ_u. De (3.2.1), un estimador consistente de Γ es [5]

$$\widehat{\Gamma}=ZZ'/T.$$

Como $\Sigma_u = E(u_t u'_t)$, un estimador plausible para esta matriz es

$$\begin{aligned}\widetilde{\Sigma}_u &= \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}'_t = \frac{1}{T}\widehat{U}\widehat{U}' = \frac{1}{T}(Y-\widehat{B}Z)(Y-\widehat{B}Z)' \\ &= \frac{1}{T}\left[Y - YZ'(ZZ')^{-1}Z\right]\left[Y - YZ'(ZZ')^{-1}Z\right]' \\ &= \frac{1}{T}Y\left[I_T - Z'(ZZ')^{-1}Z\right]\left[I_T - Z'(ZZ')^{-1}Z\right]'Y' \\ &= \frac{1}{T}Y\left(I_T - Z'(ZZ')^{-1}Z\right)Y'.\end{aligned} \tag{3.2.7}$$

Frecuentemente se requiere un ajuste para los grados de libertad Entonces, se puede considerar el estimador [5]

$$\widehat{\Sigma}_u = \frac{T}{T-Kp-1}\widetilde{\Sigma}_u. \tag{3.2.8}$$

Note que hay $Kp+1$ parámetros en cada una de las K ecuaciones de (3.0.1) y, por lo tanto, hay $Kp+1$ parámetros en cada ecuación del sistema (3.1.2). Por supuesto, $\widehat{\Sigma}_u$ y $\widetilde{\Sigma}_u$ son asintóticamente equivalentes. Son estimadores consistentes de Σ_u si las condiciones de la Proposición 3.3 se cumplen.

Proposición 3.4 (Propiedades asintóticas de los estimadores de la matriz de covarianza de ruido blanco)**.** *Sea (y_t) un proceso $VAR(p)$ estable $K-$dimensional con innovaciones de ruido blanco estándar y $\bar{B}$ un estimador de B (coeficientes asociados con el proceso VAR) tal que $\sqrt{T}vec(\overline{B}-B)$converge en distibución. Supongamos que*

$$\overline{\Sigma}_u = (Y-\overline{B}Z)(Y-\overline{B}Z)'/(T-c),$$

donde c es una constante fija. Entonces

$$plim\sqrt{T}(\overline{\Sigma}_u - UU'/T) = 0.$$

Demostración.

$$\begin{aligned}\frac{1}{T}(Y-\overline{B}Z)(Y-\overline{B}Z)' =& (B-\overline{B})\left(\frac{ZZ'}{T}\right)(B-\overline{B})' + (B-\overline{B})\frac{ZU'}{T} \\ &+ \frac{UZ'}{T}(B-\overline{B})' + \frac{UU'}{T}.\end{aligned}$$

Bajo las condiciones de la proposición, $plim(B - \overline{B}) = 0$. Por lo tanto , por el Lema (3.2)

$$plim(B - \overline{B})ZU'/\sqrt{T} = 0$$

y

$$plim\left[(B - \overline{B})\frac{ZZ'}{T}\sqrt{T}(B - \overline{B})'\right] = 0.$$

Así, $\sqrt{T}\left[(Y - \overline{B}Z)(Y - \overline{B}Z)'/T - UU'/T\right] = 0$.
Por lo tanto, la Proposición anterior se cumple, considerando que $T/(T-c) \to 1$ cuando $T \to \infty$. □

La Proposición anterior se cumple para ambos estimadores $\widehat{\Sigma}_u$ y $\widetilde{\Sigma}_u$. Esto implica que los estimadores $\widehat{\Sigma}_u$ y $\widetilde{\Sigma}_u$ tienen las mismas propiedades asintóticas que el estimador

$$\frac{UU'}{T} = \frac{1}{T}\sum_{t=1}^{T} u_t u'_t,$$

el cuál se basa en los residuos desconocidos y, por lo tanto, no es factible en la práctica. Un resultado inmediato de la Proposición 3.4 es que $\widehat{\Sigma}_u$ y $\widetilde{\Sigma}_u$ son estimadores consistentes de Σ_u, lo cual se establece en el siguiente corolario:

Corolario 3.5. *Bajo las condiciones de la Proposición 3.4, se tiene que*

$$plim\widetilde{\Sigma}_u = plim\widehat{\Sigma}_u = plimUU'/T = \Sigma_u.$$

Demostración. Por la Proposición 3.4, es suficiente mostrar que $plimUU'{=}\Sigma_u$, lo cual se sigue de B.4(4), pues

$$E\left(\frac{1}{T}UU'\right) = \frac{1}{T}\sum_{t=1}^{T} E(u_t u'_t) = \Sigma_u$$

y

$$Var\left(\frac{1}{T}vec(UU')\right) = \frac{1}{T^2}\sum_{t=1}^{T} Var[(u_t u'_t)] \leq \frac{T}{T^2}g \longrightarrow 0 \qquad (3.2.9)$$

cuando $T \longrightarrow \infty$, donde g es una cota superior constante para $Var[(u_t u'_t)]$. Esta cota existe porque los cuartos momentos de u_t son acotados, por la definición 3.1. □

3.3. Estimación por mínimos cuadrados con datos ajustados a la media y estimación de Yule-Walker

3.3.1. Estimación cuando se conoce la media del proceso

A veces un modelo VAR(p) es dado en la forma media ajustada

$$(y_t - \mu) = A_1(y_{t-1} - \mu) + \cdots + A_p(y_{t-p} - \mu) + u_t. \tag{3.3.1}$$

La estimación multivariada por mínimos cuadrados de esta forma del modelo se puede llevar a cabo si el vector μ es conocido.

Para esto, consideramos la siguiente notación

$$\begin{array}{ll} Y^0 = (y_1 - \mu, \cdots, y_T - \mu) & (K \times T), \\ A = (A_1, \cdots, A_p) & (K \times Kp), \\ X = (Y_0^0, \cdots, Y_{T-1}^0) & (KT \times 1), \\ y^0 = Vec(Y^0) & (K^2p \times 1) \\ \alpha = Vec(A) & (K^2p \times 1) \end{array} \tag{3.3.2}$$

$$Y_t^0 := \begin{bmatrix} y_t - \mu \\ \vdots \\ y_{t-p+1} \end{bmatrix} (Kp \times 1),$$

donde los términos entre paréntesis denotan las dimensiones de cada vector y matriz, respectivamente. Podemos escribir (3.3.1), para $t = 1, \cdots, T$; compactamente como

$$Y^0 = AX + U \tag{3.3.3}$$

o

$$y^0 = (X \otimes I_K)\alpha + \boldsymbol{u}. \tag{3.3.4}$$

El estimador por mínimos cuadrados (LS) es visto como

$$\widehat{\alpha} = ((XX')^{-1}X \otimes I_K)y^0 \tag{3.3.5}$$

o bien

$$\widehat{A} = Y^0X'(XX')^{-1}. \tag{3.3.6}$$

Si (y_t) es estable y u_t es ruido blanco estándar, se puede mostrar que [5]

$$\sqrt{T}(\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\alpha}}), \tag{3.3.7}$$

donde

$$\Sigma_{\widehat{\alpha}} = \Gamma_Y(0)^{-1} \otimes \Sigma_u \tag{3.3.8}$$

y

$$\Gamma_Y(0) := E(Y_t^0 Y_t^o). \tag{3.3.9}$$

3.4. Estimación del proceso media

Frecuentemente μ es desconocido, en este caso puede ser estimada por el vector de medias muestrales:

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t. \tag{3.4.1}$$

Usando (3.3.1), $\overline{y}$ puede escribirse como

$$\begin{aligned}\overline{y} =& \mu + A_1 + \left[\overline{y} + \frac{1}{T}(y_0 - y_T) - \mu\right] + \cdots \\ &+ A_p\left[\overline{y} + \frac{1}{T}(y_{-p+1} + \cdots + y_0 - y_{T-p+1} - \cdots - y_T) - \mu\right] \\ &+ \frac{1}{T}\sum_{t=1}^{T} u_t.\end{aligned} \tag{3.4.2}$$

Entonces

$$(I_k - A_1 - \cdots - A_p)(\overline{y} - \mu) = \frac{1}{T} z_T + \frac{1}{T}\sum_t u_t, \tag{3.4.3}$$

donde

$$z_T = \sum_{i=1}^{p} A_i \left[\sum_{j=0}^{i-1}(y_{0-j} - y_{T-j})\right]. \tag{3.4.4}$$

Además,

$$E(z_T/\sqrt{T}) = \frac{1}{\sqrt{T}} E(z_T) = 0 \tag{3.4.5}$$

y

$$Var(z_T/\sqrt{T}) = \frac{1}{T}Var(z_T) \quad \longrightarrow \quad 0 \tag{3.4.6}$$

cuando $T \to \infty$, porque (y_t) es estable. En otras palabras, $z_T/\sqrt{T}$ converge a cero en media. Se sigue que $\sqrt{T}\,(I_K - A_1 - \cdots - A_p)\,(\overline{y} - \mu)$ tiene la misma distribución asintótica que $\sum u_t/\sqrt{T}$ (ver Proposición B.1). Entonces, por el Lema Central del Límite (Proposición B.5)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \xrightarrow{d} \mathcal{N}(0, \Sigma_u), \tag{3.4.7}$$

si u_t es ruido blanco estándar, se obtiene el siguiente resultado

Proposición 3.6 (Propiedades asintóticas de la media muestral)**.** *Si el proceso $VAR(p)$ (y_t) dado en* (3.3.1) *es estable y u_t es ruido blanco estándar, entonces*

$$\sqrt{T}(\overline{y} - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\overline{y}}), \tag{3.4.8}$$

donde

$$\Sigma_{\overline{y}} = (I_K - A_1 - \cdots - A_p)^{-1}\,\Sigma_u\,(I_K - A_1 - \cdots - A_p)'^{-1}\,.$$

En particular, plim $\overline{y} = \mu$.

La proposición sigue de (3.4.3), (3.4.7) y (B.6).

Como $\mu = (I_k - A_1 - \cdots - A_p)^{-1}\nu$, un estimador alternativo para el proceso media es obtenido del estimador LS:

$$\widehat{\mu} = (I_k - \widehat{A}_1 - \cdots - \widehat{A}_p)^{-1}\nu \tag{3.4.9}$$

Usando otra vez la Proposición B.6, este estimador es consistente y tiene una distribución asintótica normal

$$\sqrt{T}(\widehat{\mu} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial \mu}{\partial \beta'}(\Gamma^{-1} \otimes \Sigma_u)\frac{\partial \mu'}{\partial \beta}\right), \tag{3.4.10}$$

siempre que se cumplan las condiciones de la Proposición 3.3. Se puede demostrar que

$$\frac{\partial \mu}{\partial \beta'}(\Gamma^{-1} \otimes \Sigma_u)\frac{\partial \mu'}{\partial \beta} = \Sigma_{\overline{y}} \tag{3.4.11}$$

y por lo tanto, los estimadores $\overline{\mu}$ y $\overline{y}$ para μ son asintóticamente equivalentes. Este resultado indica que no importa asintóticamente si la media se estima por separado o conjuntamente con los otros coeficientes VAR.

3.5. El estimador de Yule-Walker

El estimador LS también puede ser derivado de las ecuaciones de $Yule-Walker$ dadas en (1.2.4). Así, se tiene que

$$\Gamma_y(h) = [A_1, \cdots, A_p] \begin{bmatrix} \Gamma_y(h-1) \\ \vdots \\ \Gamma_y(h-p) \end{bmatrix}, h > 0,$$

o bien,

$$\begin{aligned} [\Gamma_y(1), \cdots, \Gamma_y(p)] &= [A_1, \cdots, A_p] \begin{bmatrix} \Gamma_y(0) & \cdots & \Gamma_y(p-1) \\ \vdots & \ddots & \vdots \\ \Gamma_y(-p+1) & \cdots & \Gamma_y(0) \end{bmatrix} \\ &= A\Gamma_Y(0) \end{aligned} \tag{3.5.1}$$

y, por lo tanto

$$A = [\Gamma_y(1), \cdots, \Gamma_y(p)]\, \Gamma_Y(0)^{-1}.$$

Estimando $\Gamma_Y(0)$ por $\widehat{X}\widehat{X}'/T$ y $[\Gamma_y(1), \cdots, \Gamma_y(p)]$ por $\widehat{Y}^0\widehat{X}'/T$, el estimador resultante es el estimador LS,

$$\widehat{\widehat{A}} = \widehat{Y}^0\widehat{X}'(\widehat{X}\widehat{X}')^{-1}. \tag{3.5.2}$$

Alternativamente, las matrices de momentos $\Gamma_y(y)$ pueden estimarse utilizando tantos datos como estén disponibles, incluidos los valores de premuestra. Por lo tanto, si una muestra $y_1, \ldots, y_T$ y p observaciones de premuestra $y_{-p+1}, \ldots, y_0$ están disponibles, μ puede estimarse como

$$\overline{y}^* = \frac{1}{T+p} \sum_{t=-p+1}^{T} y_t \tag{3.5.3}$$

y $\Gamma_y(h)$ puede ser estimado como

$$\widehat{\Gamma}_y(h) = \frac{1}{T+p-h} \sum_{-p+h+1}^{T} (y_t - \overline{y}^*)(y_{t-h} - \overline{y}^*)'. \tag{3.5.4}$$

Usando estos estimadores en (3.5.1), se obtiene el estimador de $Yule-Walker$, para A. Para procesos estables, este estimador tiene las mismas propiedades asintóticas que estimador LS. Sin embargo, puede tener propiedades menos bondadosas para muestras pequeñas [5].

3.6. Estimación por máxima verosimilitud

3.6.1. Función de Verosimilitud

Suponiendo que la distribución del proceso es conocida, la estimación por máxima verosimilitud (ML) es una alternativa a la estimación LS. Ahora se considerará que el proceso (y_t) es Gaussiano. Es decir,

$$\boldsymbol{u} = vec(U) \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \sim \mathcal{N}(0, I_T \otimes \Sigma_u). \tag{3.6.1}$$

En otras palabras, la función de densidad de probabilidad de $\boldsymbol{u}$ es

$$f_{\boldsymbol{u}}(\boldsymbol{u}) = \frac{1}{(2\pi)^{KT/2}} |I_T \otimes \Sigma_u|^{1/2} exp \left[-\frac{1}{2} \boldsymbol{u}'(I_T \otimes \Sigma_u^{-1}) \boldsymbol{u} \right]. \tag{3.6.2}$$

Además

$$\begin{aligned} \boldsymbol{u} = & \begin{bmatrix} I_K & 0 & \cdots & 0 & \cdots & \cdots & 0 \\ -A_1 & I_K & & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & & & 0 \\ -A_p & -A_{p-1} & & I_K & & & 0 \\ 0 & -A_p & & & \ddots & & 0 \\ \vdots & & \ddots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -A_p & \cdots & \cdots & I_K \end{bmatrix} (\boldsymbol{y} - \boldsymbol{\mu}^*) \\ & + \begin{bmatrix} -A_1 & -A_2 & \cdots & -A_p \\ -A_2 & -A_3 & \cdots & 0 \\ \vdots & & & \vdots \\ -A_p & 0 & \cdots & 0 \\ & \cdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} (Y_0 - \boldsymbol{\mu}) \end{aligned} \tag{3.6.3}$$

donde $\boldsymbol{y} = vec(Y)$ y $\boldsymbol{\mu}^* = (\mu', \ldots, \mu')$ son vectores de dimensión $(TK \times 1)$; y $Y_0 := (y_0', \cdots, y_{-p+1}')$ y $\boldsymbol{\mu} = (\mu', \ldots, \mu')$ son vectores de dimensión $(Kp \times 1)$. Consecuentemente, $\partial \boldsymbol{u} / \partial \boldsymbol{y}'$ es una matriz triangular inferior con unos en la diagonal principal y determinante igual a uno.

Por lo tanto, usando que $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\alpha$,

$$\begin{aligned} f_{\boldsymbol{y}}(\boldsymbol{y}) &= \left| \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{y}'} \right| \\ &= \frac{1}{(2\pi)^{KT/2}} \left| I_T \otimes \Sigma_u \right|^{-1/2} \\ &\times exp\left[\frac{1}{2} (\boldsymbol{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\alpha)' (I_T \otimes \Sigma_u^{-1}) \times (\boldsymbol{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\alpha) \right] \end{aligned} \tag{3.6.4}$$

donde X y α son definidos como en (3.3.2). Por simplicidad, los valores iniciales de Y_0 son supuestos como números fijos. Así, se obtiene la función log-verosimilitud

$$\begin{aligned} & lnL(\mu, \boldsymbol{\alpha}, \Sigma_u) \\ &= -\frac{KT}{2} ln 2\pi - \frac{T}{2} ln \left| \Sigma_u \right| \\ &- \frac{1}{2} \left[\boldsymbol{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\alpha \right]' \left(I_T \otimes \Sigma_u^{-1} \right) \left[\boldsymbol{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\alpha \right] \\ &= -\frac{KT}{2} ln 2\pi - \frac{T}{2} ln \left| \Sigma_u \right| - \frac{1}{2} \sum_{t=1}^{T} \left[(y_t - \mu) - \sum_{i=1}^{p} A_i (y_{t-i} - \mu) \right]' \\ &\times \Sigma_u^{-1} \left[(y_t - \mu) - \sum_{i=1}^{p} A_i (y_{t-i} - \mu) \right] \\ &= -\frac{KT}{2} ln 2\pi - \frac{T}{2} ln \left| \Sigma_u \right| \\ &- \frac{1}{2} \sum_t \left(y_t - \sum_i A_i y_{t-i} \right)' \Sigma_u^{-1} \left(y_t - \sum_i A_i y_{t-i} \right) \\ &+ \mu' \left(I_K - \sum A_i \right)' \Sigma_u^{-1} \sum_t \left(y_t - \sum_i A_i y_{t-i} \right) \\ &- \frac{T}{2} \mu' \left(I_K - \sum A_i \right)' \Sigma_u^{-1} \left(I_K - \sum A_i \right) \mu \\ &= -\frac{KT}{2} ln 2\pi - \frac{1}{2} ln \left| \Sigma_u \right| - \frac{1}{2} tr \left[(Y^0 - AX)' \Sigma_u^{-1} (Y^0 - AX) \right], \end{aligned} \tag{3.6.5}$$

donde $Y^0 := (y_1 - \mu, \cdots, y_T - \mu)$ y $A := (A_1, \cdots, A_p)$ son como en (3.3.2). Estas diferentes expresiones de la función log-verosimilitud son utilizadas más adelante.

3.6.2. Los estimadores ML

Para hallar los estimadores ML de $\boldsymbol{\mu}$, α y Σ_u, se necesita el sistema de derivadas parciales de primer orden:

$$\begin{aligned}\frac{\partial lnL}{\partial \mu} &= \left(I_K - \sum_i A_i\right)' \Sigma_u^{-1} \sum_t \left(y_t - \sum_i A_i y_{t-i}\right) \\ &\quad - T\left(I_K - \sum_i A_i\right)' \Sigma_u^{-1} \sum_t \left(y_t - \sum_i A_i\right)\mu \\ &= \left[I_K - A(j \otimes I_K)\right]' \Sigma_u^{-1} \left[\sum_t (y_t - \mu - AY_{t-1}^0)\right],\end{aligned} \tag{3.6.6}$$

donde Y_t^0 es definido como en (3.3.2) y $\boldsymbol{j} := (1, \cdots, 1)'$ es un vector de dimensión $(p \times 1)$,

$$\begin{aligned}\frac{\partial lnL}{\partial \alpha} &= (X \otimes I_K)(I_T \otimes \Sigma_u^{-1})\left[\boldsymbol{y} - \boldsymbol{\mu} * - (X' \otimes I_K)\boldsymbol{\alpha}\right] \\ &= (X \otimes \Sigma_u^{-1})(\boldsymbol{y} - \boldsymbol{\mu}*) - (XX' \otimes \otimes \Sigma_u^{-1})\boldsymbol{\alpha},\end{aligned} \tag{3.6.7}$$

$$\frac{\partial lnL}{\partial \Sigma_u} = -\frac{T}{2}\Sigma_u^{-1} + \frac{1}{2}\Sigma_u^{-1}(Y^0 - AX)(Y^0 - AX)'\Sigma_u^{-1}. \tag{3.6.8}$$

Igualando a cero y resolviendo el sistema de ecuaciones para los estimadores, obtenemos

$$\widetilde{\mu} = \frac{1}{T}\left(I_K - \sum_i \widetilde{A}_i\right)^{-1} \sum_t \left(y_t - \sum_i \widetilde{A}_i y_{t-i}\right), \tag{3.6.9}$$

$$\widetilde{\alpha} = \left(\left(\widetilde{X}\widetilde{X}'\right)^{-1} \widetilde{X} \otimes I_K\right)(\boldsymbol{y} - \boldsymbol{\mu}^*) \tag{3.6.10}$$

$$\widetilde{\Sigma}_u = \frac{1}{T}\left(\widetilde{Y}^0 - \widehat{A}\widetilde{X}\right)\left(\widetilde{Y}^0 - \widehat{A}\widetilde{X}\right)', \tag{3.6.11}$$

donde $\widetilde{X}$ y $\widetilde{Y}^0$ son obtenidos de X y Y^0, respectivamente, reemplazando μ por $\widetilde{\mu}$.

3.7. Propiedades de los estimadores por mínimos cuadrados

Comparando los resultados con los estimadores por mínimos cuadrados obtenidos, resulta que los estimadores por máxima verosimilitud (ML) de μ y $\boldsymbol{\alpha}$ son semejantes a los estimadores por mínimos cuadrados (LS). Por lo tanto, $\widetilde{\mu}$ y $\widetilde{\boldsymbol{\alpha}}$ son estimadores consistentes si (y_t) es un proceso $VAR(p)$ Gaussiano estable y $\sqrt{T}(\widetilde{\mu}-\mu)$ y $\sqrt{T}(\widetilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$ tienen distribución asintótica normal. Además, la matriz de covarianza de la distribución asintótica de los estimadores de ML es el límite de T multiplicado por la matriz de información inversa. La matriz de información es

$$I(\boldsymbol{\delta}) = -E\left[\frac{\partial^2 lnL}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right] \tag{3.7.1}$$

donde $\boldsymbol{\delta}' := (\mu', \boldsymbol{\alpha}', \boldsymbol{\sigma}')$ con $\boldsymbol{\sigma} := vech(\Sigma_u)$. Note que este operador está relacionado con el operador vec por la matriz de eliminación $\boldsymbol{L}_K$, esto es, $vech(\Sigma_u) = \boldsymbol{L}_K vec(\Sigma_u)$ o, definiendo $\boldsymbol{\omega} := vec(\Sigma_u)$, $\boldsymbol{\sigma} = \boldsymbol{L}_K\boldsymbol{\omega}$. Se sabe que la matriz de covarianza asintótica del estimador ML de $\widetilde{\boldsymbol{\delta}}$ es

$$\lim_{T\to\infty}\left[I(\delta)/T\right]^{-1}. \tag{3.7.2}$$

Para determinar esta matriz, necesitamos las derivadas de segundo orden de la función $log-verosimilitud$. De (3.6.6) a (3.6.8), obtenemos

$$\frac{\partial^2 lnL}{\partial\mu\partial\mu'} = -T\left(I_K - \sum_i A_i\right)'\Sigma_u^{-1}\left(I_K - \sum_i A_i\right), \tag{3.7.3}$$

$$\frac{\partial^2 lnL}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'} = -(XX'\otimes\Sigma_u^{-1}), \tag{3.7.4}$$

$$\begin{aligned}\frac{\partial^2 lnL}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}'} =& \frac{T}{2}\left(\Sigma_u^{-1}\otimes\Sigma_u^{-1}\right) - \frac{1}{2}\left(\Sigma_u^{-1}\otimes\Sigma_u^{-1}UU'\Sigma_u^{-1}\right)\\ &- \frac{1}{2}\left(\Sigma_u^{-1}UU'\Sigma_u^{-1}\otimes\Sigma_u^{-1}\right)\end{aligned}, \tag{3.7.5}$$

$$\begin{aligned}\frac{\partial^2 lnL}{\partial\mu\partial\alpha'} =& -\left[I_K - (j'\otimes I_K)A'\right]\Sigma_u^{-1}, \sum_t Y_{t-1}^{0'}\otimes I_K\\ &- \left(\sum_t u'_t\Sigma_u^{-1}I_K\right)(I_K\otimes j'\otimes I_K)\frac{\partial vec(A')}{\partial\alpha'}\end{aligned}, \tag{3.7.6}$$

$$\frac{\partial^2 lnL}{\partial\omega\partial\mu'} = \frac{1}{2}(\Sigma_u^{-1}\otimes\Sigma_u^{-1})\left[(I_K\otimes U)\frac{\partial vec(U)}{\partial\mu'} + (U\otimes I_K)\frac{\partial vec(U)}{\partial\mu}\right] \quad (3.7.7)$$

y

$$\frac{\partial^2 lnL}{\partial\omega\partial\alpha'} = \frac{1}{2}(\Sigma_u^{-1}\otimes\Sigma_u^{-1})\left[(I_K\otimes UX')\frac{\partial vec(A')}{\partial\alpha'} + (UK'\otimes I_K)\right]. \quad (3.7.8)$$

De (3.7.6) tenemos que

$$\lim_{T\to\infty}\ T^{-1}E\left(\frac{\partial^2 lnL'}{\partial\mu\partial\boldsymbol{\alpha}}\right) = 0, \quad (3.7.9)$$

pues $E\left(\sum_t Y_{t-1}^0/T\right) \longrightarrow 0$. Además, por (3.7.7), tenemos que

$$E\left(\frac{\partial^2 lnL}{\partial\omega\partial\mu'}\right) = 0 \quad (3.7.10)$$

porque $E(U) = 0$ y $\partial vec(U')/\partial\mu'$ es una matriz de constante. También, de (3.7.8), tenemos que

$$\lim_{T\to\infty}\ T^{-1}E\left(\frac{\partial^2 lnL}{\partial\omega\partial\alpha'}\right) \quad (3.7.11)$$

pues $E(UX'/T) \longrightarrow 0$. Así, $\lim_{T\to\infty} I(\delta/T)$ es la matriz diagonal por bloques, pues $\frac{\partial^2 lnL}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}$ es una matriz diagonal. Y obtenemos las distribuciones asintóticas de $\mu, \boldsymbol{\alpha}$ y $\boldsymbol{\sigma}$, como sigue.

Multiplicar menos el inverso de (3.7.3) por T da la matriz de covarianza asintótica del estimador ML para el vector media μ, esto es,

$$\sqrt{T}(\widetilde{\mu}-\mu) \xrightarrow{d} \mathcal{N}\left(0, \left(I_K - \sum_{i=1}^{p} A_i\right)^{-1}\Sigma_u\left(I_K - \sum_{i=1}^{p} A'_i\right)^{-1}\right). \quad (3.7.12)$$

Por lo tanto, $\widetilde{\mu}$ tiene la misma distribución asintótica que $\overline{y}$ (ver Proposición 3.6). En otras palabras, los dos estimadores para μ son asintóticamente equivalentes. La equivalencia asintótica de $\widetilde{\mu}$ y $\overline{y}$ pueden ser vistos de (3.6.9) (ver argumento antes de la Proposición 3.6).

Tomando el límite de T^{-1} multiplicado por la esperanza de menos (3.7.4), resulta $\Gamma_Y(0)\otimes\Sigma_u^{-1}$. Note que $E(XX'/T)$ no es estrictamente igual a $\Gamma_Y(0)$ pues se asumen valores iniciales fijos $y_{-p+1},\ldots,y_0$. Sin embargo,

asintóticamente, a medida que T va al infinito, el impacto de los valores iniciales se desvanece. Por lo tanto, se obtiene

$$\sqrt{T}(\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Gamma_Y(0)^{-1} \otimes \Sigma_u). \tag{3.7.13}$$

Este resultado también sigue de la equivalencia de los estimadores ML y LS.

Como $E(UU') = T\Sigma_u$, se sigue de (3.7.5) que

$$E\left(\frac{\partial^2 ln\ L}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}'}\right) = -\frac{T}{2}\left(\Sigma_u^{-1} \otimes \Sigma_u^{-1}\right). \tag{3.7.14}$$

Denotando por $\boldsymbol{D}_K$ a la matriz de duplicación (ver Apéndice A.5), de modo que $\boldsymbol{\omega} = \boldsymbol{D}_K \boldsymbol{\sigma}$, se obtiene

$$\frac{\partial^2 ln\ l}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'} = \frac{\partial \boldsymbol{\omega}'}{\partial \sigma} \frac{\partial^2 ln\ l}{\partial \omega \partial \omega'} \frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\sigma}'} = \boldsymbol{D}_K' \frac{\partial^2 ln\ l}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}} \boldsymbol{D}_K, \tag{3.7.15}$$

y por lo tanto

$$\sqrt{T}\left(\widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widetilde{\boldsymbol{\sigma}}}), \tag{3.7.16}$$

con

$$\begin{aligned} \Sigma_{\widetilde{\boldsymbol{\sigma}}} &= -TE\left(\frac{\partial^2 ln\ l}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'}\right)^{-1} = 2\left[\boldsymbol{D}_K'(\Sigma_u^{-1} \otimes \Sigma_u^{-1} \boldsymbol{D}_K)\right]^{-1} \\ &= 2\boldsymbol{D}_K^{+}(\Sigma_u \otimes \Sigma_u)\boldsymbol{D}_K^{+'}, \end{aligned} \tag{3.7.17}$$

donde $\boldsymbol{D}_K^{+} = (\boldsymbol{D}_K' \boldsymbol{D}_K)^{-1} \boldsymbol{D}_K'$.

Las propiedades de los estimadores descritos anteriormente se resumen en la siguiente proposición

Proposición 3.7. *Sea y_t un proceso $VAR(p)$ Gaussiano estable, entonces los estimadores $\widetilde{\mu}$, $\widetilde{\boldsymbol{\alpha}}$ y $\widetilde{\boldsymbol{\sigma}} = vec(\widetilde{\Sigma}_u)$ ML son consistentes y*

$$\sqrt{T}\begin{bmatrix} \widetilde{\mu} - \mu \\ \widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \Sigma_{\widetilde{\mu}} & 0 & 0 \\ 0 & \Sigma_{\widetilde{\boldsymbol{\alpha}}} & 0 \\ 0 & 0 & \Sigma_{\widetilde{\boldsymbol{\sigma}}} \end{bmatrix}\right), \tag{3.7.18}$$

así que $\widetilde{\mu}$ es asintóticamente independiente de $\widetilde{\boldsymbol{\alpha}}$ y $\widetilde{\Sigma}_u$ y $\widetilde{\boldsymbol{\alpha}}$ es asintóticamente independiente de $\widetilde{\mu}$ y $\widetilde{\Sigma}_u$. Las matrices de covarianza dadas por

$$\Sigma_{\widetilde{\mu}} = \left(I_K - \sum_i A_i\right)^{-1} \Sigma_u \left(I_K - \sum_i A_i'\right)^{-1}, \tag{3.7.19}$$

$$\Sigma_{\widetilde{\boldsymbol{\alpha}}} = \Gamma_Y(0)^{-1} \otimes \Sigma_u, \tag{3.7.20}$$

$$\Sigma_{\widetilde{\boldsymbol{\sigma}}} = 2D_K^+(\Sigma_u \otimes \Sigma_u)D_K^{+'}. \tag{3.7.21}$$

se pueden estimar consistentemente reemplazando las cantidades desconocidas por sus estimadores de ML y estimando $\Gamma_Y(0)$ *por* $\widetilde{X}\widetilde{X}'/T$.

Capítulo 4

Selección del orden de un proceso VAR y verificación de la adecuación del modelo

4.1. Seleccción del orden de un proceso VAR

En la práctica, el orden generalmente es desconocido. Para la precisión del pronóstico es útil contar con procedimientos o criterios para elegir el orden del proceso VAR adecuado.

4.1.1. Un esquema de prueba para la determinación del orden VAR

Asumiendo que se sabe que M es un límite superior para el orden VAR (máximo número de parámetros distintos de cero), la siguiente secuencia de hipótesis nula y alternativa se pueden probar.

$$\begin{aligned}
H_0^1 &: A_M = 0 \qquad \textbf{\textit{vs}} \quad H_1^1 : A_M \neq 0 \\
H_0^2 &: A_{M-1} = 0 \qquad \textbf{\textit{vs}} \quad H_1^2 : A_{M-1} \neq 0 | A_M = 0 \\
&\vdots \\
H_0^i &: A_{M-i+1} = 0 \quad \textbf{\textit{vs}} \quad H_1^1 : A_{M-i+1} \neq 0 | A_M = \cdots = A_{M-i+2} \\
&\vdots \\
H_0^M &: A_1 = 0 \qquad \textbf{\textit{vs}} \quad H_1^M : A_1 = 0 | A_M = \cdots = A_2,
\end{aligned} \tag{4.1.1}$$

con $i = 1, 2, \cdots, M$.

En este esquema, cada hipótesis nula se prueba condicionando que las anteriores sean verdaderas. El procedimiento se termina y el orden VAR es elegido de acuerdo a si se rechaza una de las hipótesis nulas. Esto es, si H_0^i es rechazada, $\widehat{p} = M - i + 1$ será elegido como estimador del orden.

El estadístico para probar la $i-$ésima hipótesis (4.1.1), es

$$\lambda_{LR}(i) = T[ln|\widetilde{\Sigma}_u(M-i)| - ln|\widetilde{\Sigma}_u(M-i+1)|], \tag{4.1.2}$$

donde $i = 1, \cdots, M$, $\widetilde{\Sigma}_u(m)$ denota el estimador ML de Σ_u y el modelo $VAR(m)$ es ajustado a una serie de tiempo de longitud T. Por la Proposición 4.1 de [5], el estadístico de prueba tiene una distribución asintótica $\chi^2(K^2)$ si H_0^i y la hipótesis nula previa a H_0^i son verdaderas. Tenga en cuenta que los parámetros K^2 se establecen en H_0^i.

4.2. Criterios de selección del orden VAR

Como el objetivo es pronosticar, tiene sentido elegir el orden de manera que se minimice el error del pronóstico. El criterio que se considerará se llama *criterio de error de predicción final (FPE)*

$$\begin{aligned}
FPE(m) &= det\left[\frac{T+Km+1}{T}\frac{T}{T-Km-1}\widetilde{\Sigma}_u(m)\right] \\
&= \left[\frac{T+K_m+1}{T-K_m-1}\right]^K det\widetilde{\Sigma}_u(m),
\end{aligned} \tag{4.2.1}$$

donde m es el orden del modelo y $\widetilde{\Sigma}_u$ es el estimador ML de Σ_u.

Basado en el criterio FPE, la estimación $\widehat{p}(FPE)$ de p se elige de tal forma que
$FPE[\widehat{p}(FPE)] = min\left\{FPE(m)|m = 0, 1, ...M\right\}$.

El orden que minimiza los valores de FPE se elige como estimación para p.

Akaike, basado en un razonamiento bastante diferente, derivó un criterio muy similar, abreviado usualmente por AIC [1]. Para un proceso $VAR(m)$ el criterio es definido como

$$AIC(m) = ln|\widetilde{\Sigma}_u(m)| + \frac{2mK^2}{T}. \tag{4.2.2}$$

La estimación $\widehat{p}(AIC)$ para p se elige de modo que este criterio se minimice.

La similitud de los criterios AIC y FPE se puede ver al notar que

$$ln\ FPE(m) = AIC(m) + 2K/T + O(T^{-2}). \tag{4.2.3}$$

Si el interés se centra en el orden correcto, tiene sentido elegir un estimador que tenga propiedades asintóticas deseables, como la consitencia.

Se dice que un estimador es consistente si $p\lim_{T\to\infty}\widehat{p} = p$ o equivalentemente $\lim_{T\to\infty}\mathcal{P}(\widehat{p} = p) = 1$. y fuertemente consistente si $\mathcal{P}(lim\ \widehat{p} = p\}) = 1$.

Corolario 4.1. *Sea y_t un proceso $K-$dimensional estable asociado con ruido blanco estándar. Suponga que el orden máximo $M \geq p$ y $\hat{p}$ son elegidos para que minimicen a los criterios $FPE(m)$ y $AIC(m)$, para $m = 0, 1, \ldots, M$. Entonces $\widehat{p}(FPE)$ y $\widehat{p}(AIC)$ son no consistentes.*

Dos criterios consistentes que han sido bastante populares en trabajos aplicados recientemente son HQ y SC. El primero en [3]

$$HQ(m) = ln|\widetilde{\Sigma}_u(m)| + \frac{2\ ln\ lnT}{T}mK^2. \tag{4.2.4}$$

El estimador $\widehat{p}(HQ)$ es el orden que minimiza $HQ(m)$ para $m = 1, \cdots, M$ [7].

Utilizando argumentos bayesianos, Schwarz [10] dedujo el siguiente criterio:

$$SC(m) = ln|\widetilde{\Sigma}_u(m)| + \frac{lnT}{T}mK^2. \tag{4.2.5}$$

De igual forma $\widehat{p}(SC)$ es elegido de tal forma que minimice el criterio.

Corolario 4.2. *Sea y_t un proceso $K-$dimensional estable asociado con ruido blanco estándar. Suponga que el orden máximo $M \geq p$ y $\hat{p}$ son elegidos para que minimicen a los criterios SC y HQ, para $m = 0, 1, \ldots, M$. Entonces SC es fuertemente consistente y HQ es consistente. Si la dimensión K del proceso es mayor que uno, ambos criterios son fuertemente consistentes.*

Proposición 4.3. *Sea $y_{-M+1}, \ldots, y_0, y_1, \ldots, y_T$ una serie de tiempo múltiple $K-$dimensional y suponga los modelos $VAR(m)$, $m = 0, 1, \ldots, M$, se ajustan a $y_1, \ldots, y_T$. Entonces se cumplen las siguientes relaciones:*

$$\hat{p}(SC) \leq (\hat{AIC}) \quad si \quad T \geq 8, \tag{4.2.6}$$

$$\hat{p}(SC) \leq (\hat{HQ}) \quad para \quad todo \quad T, \tag{4.2.7}$$

$$\hat{p}(HQ) \leq (\hat{AIC}) \quad si \quad T \geq 16. \tag{4.2.8}$$

4.3. Las distribuciones asintóticas de las autocovarianzas residuales y las autocorrelaciones de un proceso VAR

4.3.1. Blancura de los residuos

Asumiendo que el modelo ha sido estimado por mínimos cuadrados (LS) y usando la notación de la Sección 3.1, el estimador de coeficientes se denota por $\widehat{B}$ y los residuos correspondientes son $U = (\widehat{u}_1, \cdots, \widehat{u}_T) := Y - \widehat{B}Z$. Además, sean

$$\begin{aligned} \widehat{C}_i &:= \frac{1}{T} F_i \widehat{U}', \quad i = 0, 1, \cdots, h \\ \widehat{\boldsymbol{C}}_h &:= (\widehat{C}_1, \cdots, \widehat{C}_h) = \frac{1}{T} \widehat{U} F (I_h \otimes \widehat{U}') \\ \widehat{\boldsymbol{c}}_h &:= vec(\widehat{\boldsymbol{C}}_h) \end{aligned} \tag{4.3.1}$$

y, correspondientemente,

$$\widehat{R} := \widehat{D}^{-1} \widehat{C}_i \widehat{D}^{-1}, \quad \widehat{\boldsymbol{R}}_h := (\widehat{R}_1, \cdots, \widehat{R}_h), \quad \widehat{\boldsymbol{r}}_h := vec(\widehat{\boldsymbol{R}}_h) \tag{4.3.2}$$

donde $\widehat{D}$ es la matriz diagonal, con las raíces cuadradas de los elementos de la diagonal de $\widehat{C_0}$ en la diagonal principal y F_i es de dimensión $(T \times T)$. Por ejemplo, para $i = 2$,

$$F := \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}',$$

y para $i = 0$, $F_0 = I_T$, y $F := (F_1, \dots, F_h)$ es una matriz de dimensión $(T \times hT)$.

Proposición 4.4 (Distribuciones asintóticas)**.** *Sea y_t un proceso $VAR(p)$ estable de dimensión K, con proceso de ruido blanco estándar u_t, idénticamente distribuido y considere la estimación por LS. Entonces*

$$\sqrt{T}\widehat{\boldsymbol{C}}_h \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\boldsymbol{c}}(h)\right), \tag{4.3.3}$$

con

$$\begin{aligned} \Sigma_{\boldsymbol{c}}(h) &= \left(I_h \otimes \Sigma_u \widetilde{G}' \Gamma^{-1} \widetilde{G}\right) \otimes \Sigma_u. \\ &= (I_h \otimes \Sigma_u \Sigma_u) - \overline{G}\left[\Gamma_Y(0)^{-1} \otimes \Sigma_u\right] \overline{G}', \end{aligned} \tag{4.3.4}$$

donde $\Gamma := ZZ'/T$, $\Gamma_Y(0)$ es la matriz de covarianza de $Y_t := (y_t', \cdots, y_{t-p+1}')$, $\overline{G} := \breve{G}' \otimes I_K$, $\widetilde{G} := \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \Sigma_u & \Phi_1\Sigma_u & \cdots & \Phi_{h-1}\Sigma_h \\ 0 & \Sigma_u & \cdots & \Phi_{h-2}\Sigma_u \\ & \vdots & & \vdots \\ 0 & 0 & & \Phi_{h-p}\Sigma_u \end{bmatrix}$ y $\breve{G}$ es una matriz de dimensión $(K_p \times K_h)$ que tiene la misma forma que $\widetilde{G}$, excepto que la primera fila de ceros es eliminada.

Ahora, la distribución asintótica para las autocorrelaciones estimadas se obtiene de una forma similar.

Proposición 4.5 (Distribuciones asintóticas de autocorrelaciones residuales)**.** *Sea D una matriz diagonal $(K \times K)$ con raíces cuadradas de Σ_u en la*

diagonal y $G_0 := \widetilde{G}(I_h \otimes D^{-1})$. *Entonces, bajo las condiciones de la proposición anterior, se tiene que*

$$\sqrt{T}\widehat{\boldsymbol{r}}_h \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\boldsymbol{r}}(h)\right),$$

donde

$\Sigma_{\boldsymbol{r}}(h) = [(I_h \otimes R_u) - G_0'\Gamma^{-1}G_0] \otimes R_u$.

Especificamente,

$\sqrt{T}vec(\widehat{R}_j) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_R(j)\right), \quad j = 1, 2, \ldots,$

donde

$$\Sigma_R(j) = \left(R_u - D^{-1}\Sigma_u \left[0 : \Phi_{j-1}' : \cdots : \Phi_{j-p}'\right] \Gamma^{-1} \begin{bmatrix} 0 \\ \Phi_{j-1} \\ \vdots \\ \Phi_{j-p} \end{bmatrix} \Sigma_1 D^{-1} \right) \otimes R_u \tag{4.3.5}$$

con $\Phi_i = 0$ *para* $i < 0$.

4.3.2. Prueba de Portmanteau

Los resultados anteriores también se pueden usar para construir una prueba para la significancia de las autocorrelaciones residuales hasta el retraso h. Esta prueba se denomina comúnmente prueba de Portmanteau. Está diseñada para probar

$$H_0 : \boldsymbol{R}_h = (R_1, \cdots, R_h) = 0 \quad contra \quad H_1 : \boldsymbol{R}_h \neq 0. \tag{4.3.6}$$

El estadístico de prueba es

$$\begin{aligned} Q_h &:= T\sum_{i=1}^{h} tr(\widehat{R}_i'\widehat{R}_u^{-1}\widehat{R}_i\widehat{R}_u^{-1}) \\ &= T\sum_{i=1}^{h} tr(\widehat{R}_i'\widehat{R}_u^{-1}\widehat{R}_i\widehat{R}_u^{-1}\widehat{D}^{-1}\widehat{D}) \\ &= T\sum_{i=1}^{h} tr(\widehat{D}\widehat{R}_i'\widehat{D}\widehat{D}^{-1}\widehat{R}_u^{-1}\widehat{D}^{-1}\widehat{D}\widehat{R}_i\widehat{D}\widehat{D}^{-1}\widehat{R}_u^{-1}\widehat{D}^{-1}) \\ &= T\sum_{i=1}^{h} tr(\widehat{C}_i'\widehat{C}_0^{-1}\widehat{C}_i\widehat{C}_0^{-1}). \end{aligned} \tag{4.3.7}$$

Por la Proposición 4.4, tiene un distribución asintótica aproximada χ^2, como se muestra en la siguiente Proposición.

Proposición 4.6 (Distribución asintótica del estadístico de Portmanteu)**.** *Bajo las condiciones de la Proposición 4.4, se tiene que, aproximadamente para T y h grandes*

$$\begin{aligned} Q_h &= T\sum_{i=1}^{h} tr(\widehat{C}_i'\widehat{C}_0^{-1}\widehat{C}_i\widehat{C}_0^{-1}) \\ &= Tvec(\widehat{\boldsymbol{C}}_h)'(I_h \otimes \widehat{C}_0^{-1} \otimes \widehat{C}_0^{-1})vec(\widehat{\boldsymbol{C}}_h) \approx \chi^2(K^2(h-p)). \end{aligned} \tag{4.3.8}$$

Para fines prácticos, es importante recordar que la distribución aproximada χ^2 del estadístico de prueba puede ser engañosa para valores pequeños de h.

4.3.3. Pruebas del multiplicador de Lagrange

Otra forma de probar un modelo VAR para autocorrelación residual es asumir un modelo VAR para el vector de error, $u_t = D_1u_{t-1} + \cdots + D_hu_{t-h} + v_t$, donde v_t es ruido blanco. Esta es igual para u_t si no hay autocorrelación residual. Por lo tanto, se desea contrastar el par de hipótesis

$$\begin{aligned} &H_0 : D_1 = \cdots = D_h = 0 \qquad \textbf{\textit{vs}} \\ &H_1 : D_j \neq 0, \end{aligned} \tag{4.3.9}$$

para al menos un $j \in \{1, \cdots, h\}$. La estadística estándar χ^2 para probar $D = 0$, es

$$\lambda_{LM}(h) = vec(\widehat{U}\widehat{\mathcal{U}}')' \left(\left[\widehat{\mathcal{U}}\widehat{\mathcal{U}}' - \widehat{\mathcal{U}}Z'(ZZ')^{-1}Z\widehat{\mathcal{U}}' \right]^{-1} \otimes \widehat{\Sigma}_u^{-1} \right) vec(\widehat{U}\widehat{\mathcal{U}}').$$

Proposición 4.7 (Distribución asintótica de la estadística λ_{LM} para la autocorrelación residual)**.** *Bajo las condiciones de la Proposición 4.4*

$$\lambda_{LM} \xrightarrow{d} \chi^2(hK^2).$$

A diferencia de la prueba de Portmanteau que debería usarse solo para h razonablemente grande, las pruebas de multiplicadores de Lagrange son más adecuadas para valores pequeños de h.

4.4. Pruebas de anormalidad

La no normalidad de los residuos puede indicar, de forma general, que el modelo no es una buena representación del proceso de generación de datos. Por lo que es deseable probar este supuesto.

4.4.1. Prueba de anormalidad para proceso VAR

Un proceso VAR(p) estacionario, digamos

$$y_t - \mu = A_1(y_t - \mu) + \ldots + A_p(y_{y-p} - \mu) + u_t, \tag{4.4.1}$$

es Gaussiano si y sólo si el proceso de ruido blanco u_t es Gaussiano. Por lo tanto, la normalidad de las y_t se puede comprobar mediante las u_t. En la práctica, los terminos u_t son remplazados por las estimaciones residuales.

La razón por la que las pruebas de normalidad se basan en los errores y no en las observaciones originales y_t, es que las pruebas basadas en estos últimos pueden ser menos poderosas que las basadas en los residuos de estimación.

Las pruebas desarrolladas a continuación se basan en el tercer y cuarto momento central, que brindan información acerca de la asimetría y la curtosis, de la ditribución normal, respectivamente.

Proposición 4.8. *Sea y_t un proceso VAR(p) Gaussiano de dimensión K, como en 4.4.1, donde u_t es ruido blanco de media cero y matriz de covarianza no singular Σ_u y sean $\hat{A}_1, ..., \hat{A}_p$ estimadores consistentes con distribución asintótica normal, basados en una muestra $y_1, ..., y_T$. Defina*

$$\hat{u}_t := (y_t - \bar{y}) - \hat{A}_1(y_{t-1} - \bar{y}) - \ldots - \hat{A}_p(y_{t-1} - \bar{y}), \qquad t = 1, \ldots, T.$$

$$\hat{\Sigma}_u := \frac{1}{T - Kp - 1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$$

y sea $\hat{P}$ una matriz que satisface $\hat{P}\hat{P}' = \hat{\Sigma}_u$, tal que $plim(\hat{P} - P) = 0$. Además, sean

$$\hat{w}_t = (\hat{w}_{1t} \ldots, \hat{w}_{Kt})' := \hat{P}^{-1}\hat{u}_t,$$
$$\hat{b}_1 = (\hat{b}_{11}, \ldots, \hat{b}_{K1})'$$

con

$$\hat{b}_{k1} := \frac{1}{T}\sum_{t=1}^{T} \hat{w}_{kt}^3, \qquad k = 1, \ldots, K,$$

y

$$\hat{b}_2 = (\hat{b}_{12}, \ldots, \hat{b}_{K2})',$$

$$\hat{b}_{k2} := \frac{1}{T}\sum_{t=1}^{T} \hat{w}_{kt}^4, \qquad k = 1, \ldots, K.$$

Entonces

$$\sqrt{T}\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 - 3_K \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} 6I_K & 0 \\ 0 & 24I_K \end{bmatrix}\right).$$

La Proposición 4.8 implica que

$$\hat{\lambda_s} := \hat{T}\hat{b'}_1\hat{b}_1/6 \xrightarrow{d} \chi^2(K), \tag{4.4.2}$$

$$\hat{\lambda_k} := T(\hat{b'}_2 - 3_K)'(\hat{b}_2 - 3_K)/24 \xrightarrow{d} \chi^2(K), \tag{4.4.3}$$

$$\hat{\lambda}_{sk} := \hat{\lambda}_s + \hat{\lambda}_k \xrightarrow{d} \chi^2(2K). \tag{4.4.4}$$

Por lo tanto, las tres estadísticas se pueden usar para probar la anormalidad.

La primera estadística se puede usar para probar

$$H_0 : E\begin{bmatrix} \omega_{1t}^3 \\ \vdots \\ \omega_{Kt}^3 \end{bmatrix} = 0 \qquad \boldsymbol{vs} \qquad H_1 : E\begin{bmatrix} \omega_{1t}^3 \\ \vdots \\ \omega_{Kt}^3 \end{bmatrix} \neq 0 \tag{4.4.5}$$

y λ_K se puede utilizar para probar

$$H_0 : E\begin{bmatrix} \omega_{1t}^4 \\ \vdots \\ \omega_{Kt}^4 \end{bmatrix} = 3_K \qquad \boldsymbol{vs} \qquad H_1 : E\begin{bmatrix} \omega_{1t}^4 \\ \vdots \\ \omega_{Kt}^4 \end{bmatrix} \neq 3_K \tag{4.4.6}$$

Más aún,

$$\hat{\lambda}_{sk} := \hat{\lambda}_s + \hat{\lambda}_k \xrightarrow{d} \chi^2(2K) \tag{4.4.7}$$

se puede utilizar para la prueba conjunta de las hipótesis nulas en 4.4.5 y 4.4.6.

4.4.2. Prueba de Shapiro-Wilks

Esta prueba fue desarrollada por Shapiro y Wilks. Fue la primera capaz de detectar desviaciones de la normalidad generadas por la asimetría o la curtosis. En los últimos años esta prueba se ha convertido en la prueba más auxiliada.

Cabe destacar que esta prueba requiere que la muestra esté ordenada de forma ascendente, pues está basada en regresiones y correlaciones que han sido empleadas en muestras completas de estadísticos de orden.

Las hipótesis de contraste y la estadística de prueba están definidas como sigue:
H_0 : La muestra sigue una distribución normal **vs**
H_1 : La muestra no sigue una distribución normal.

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{4.4.8}$$

donde $x_{(i)}$ es el número que ocupa la i-ésima posición en la muestra, es decir, la i-ésima estadística de orden; las a_i son constantes generadas a partir de las medias, varianzas y covarianzas de las estadísticas de orden de una muestra de tamaño n.

Capítulo 5

Una aplicación del Análisis de Series de Tiempo

Se denomina Producto Interno Bruto o PIB al valor total de todos los bienes y servicios finales producidos en una economía en un periodo de tiempo definido y estimado en unidades monetarias [8]. Cuando se habla de bienes y servicios finales, se hace referencia a aquellos que compra el consumidor final.

El PIB es el resultado de la suma del consumo de las familias, inversiones de empresas, gastos del gobierno y saldo de la balanza comercial.

El crecimiento del PIB revela un rasgo característico de la evolución y buena salud de la economía de un país, a su vez, es una buena señal que indica que hay mayor probabilidad de "encontrar un buen trabajo, o que nos aumente el sueldo y que cosas que hasta hoy no estaban vedadas se hallen a nuestro alcance" [6]. Por otro lado, con una recesión es probable que haya más desempleo y que esto afecte seriamente a muchas familias.

Actualmente es muy fácil obtener miles y miles de datos o valores numéricos de un tema en particular, que representan una realidad abstracta. Pero, " Aunque los datos pueden ser considerados como la base de la creación de información, lo cierto es que se convierten en información cuando se les añade un significado "[4]. En el caso del PIB, es necesario concentrarse en las predicciones de su crecimiento o decrecimiento para valorar el estado o situación de una economía y sus perspectivas de evolución a futuro.

Por lo general, la estimación del PIB es anual o trimestral, pero puede calcularse para períodos más cortos o largos de tiempo dependiendo de la disponibilidad de información estadística y de los intereses del investigador.

En México el PIB se da a conocer aproximadamente dos meses después de que termina el periodo en cuestión, lo que hace que otras formas de toma de decisiones tomen un lugar importante. Es por ello que en ocasiones se recurre a hacer predicciones sobre el comportamiento de la economía basado en el comportamiento más actual que se conozca.

Cabe recalcar que al recurrir a los pronósticos, esperar como resultado "precisión", es una idea errónea. Los pronósticos sólo sirven para tratar de tomar las mejores decisiones. Así, los pronósticos del PIB dan una idea a las empresas acerca de sus inversiones; a las familias de sus gastos; y al gobierno de su política económica.

En este trabajo se hace un análisis de la serie de tiempo múltiple que involucra al PIB trimestral de las actividades primarias (AP), secundarias (AS) y terciarias (AT), registradas entre 1993 y 2019; con el objetivo de elegir un modelo estadístico para la predicción de datos de las mismas, usando el método de Box-Jenkins, con ayuda de Matlab.

Es importante tener presente que dentro de las actividades primarias están aquellas en las que se utilizan los recursos tal como los proporciona la naturaleza, entre las que se encuentran la pesca, caza, agricultura y ganadería. En las actividades terciarias se consideran comercios, transportes, servicios profesionales, etc. Finalmente, en las actividades secundarias se encuentran la minería, construcción, suministro de energía eléctrica, etc.

La Figura 5.1 muestra el gráfico de las tres variables con las que se trabaja, es posible observar que las tres series presentan tendencia y no oscilan alrededor de un valor constante, en consecuencia no presentan estacionariedad. No obstante, con ayuda de Matlab se hizo la prueba de Dickey-Fuller aumentada que comprobó que en efecto la series son no estacionarias, por lo que se procedió a aplicar logaritmos a las series, como se puede ver en la Figura 5.2.

Gráficamente se puede observar que la varianza se ha estabilizado; pero las tres series siguen mostrando una tendencia creciente y además, la serie no oscila alrededor de un valor constante. Es decir, no son estacionarias; lo que se confirma al emplear nuevamente la prueba de Dickey-Fuller aumentada. Para eliminar la no estacionariedad se hizo una segunda transformación aplicando primeras diferencias a los datos que ya habían sido transformados y como resultado se obtuvieron los datos observados en la Figura 5.3.

Notamos ahora que las series muestran un mejor comportamiento, a partir de la Figura 5.3 se puede apreciar los valores de las series tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa

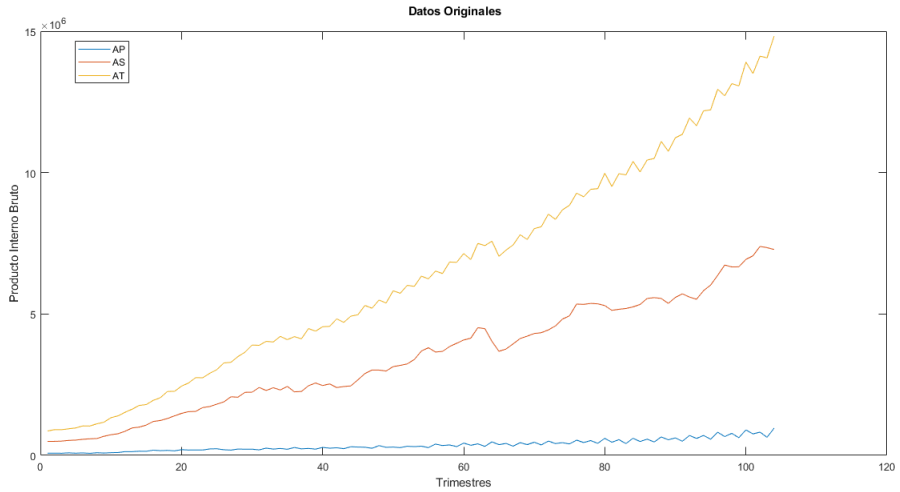

Figura 5.1: Gráfica del PIB trimestral de las actividades primarias, secundarias y terciarias, para el periodo de enero de 1993 a diciembre del 2019.

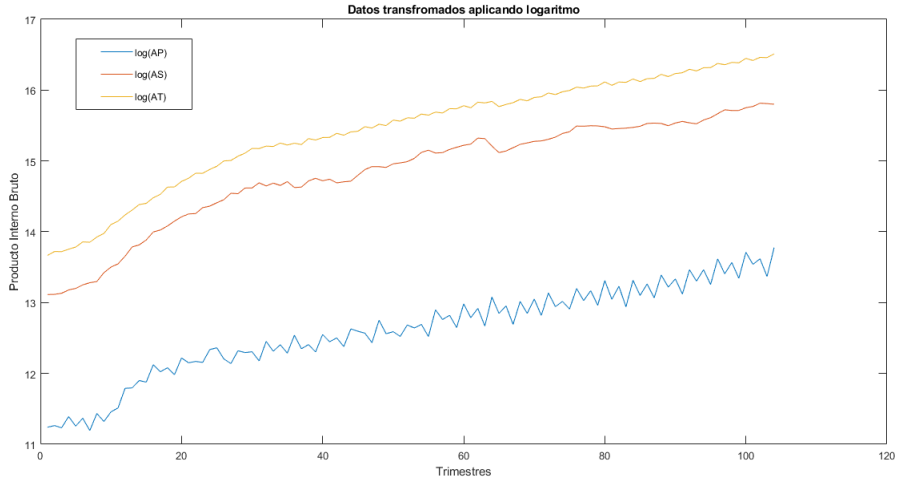

Figura 5.2: Gráfica del logaritmo del PIB trimestral de las actividades primarias, secundarias y terciarias, para el periodo de enero de 1993 a diciembre del 2019.

media también permanece constante en el tiempo, es decir, las tres series son ahora estacionarias. Para ratificar esto se recurrió nuevamente a la prueba de

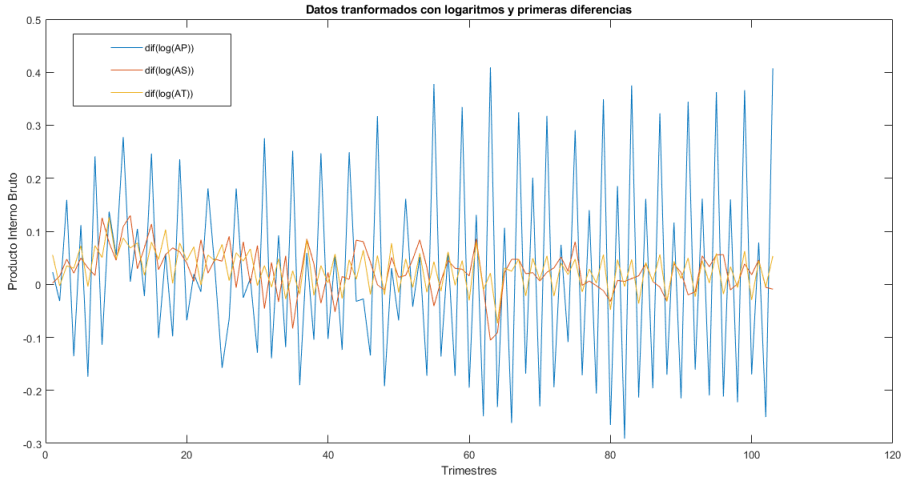

Figura 5.3: Gráfica de las primeras diferencias del logaritmo del PIB trimestral de las actividades primarias, secundarias y terciarias, para el periodo de enero de 1993 a diciembre del 2019.

Dickey-Fuller aumentada, la cual nos constata que las tres series satisfacen la condición de estacionariedad requerida.

Se propuso primero un modelo de orden p=1. Los criterios de selección de orden indican los siguientes resultados, AIC=-1000.6, BIC=–969.105. Para la diagnosis del modelo se hizo primero la prueba de Portmanteau, la cual indicó que los residuos no están autocorrelacionados. Luego se hizo uso de la herramienta Q-Q plot (Figura 5.4) para la comprobación de la normalidad de los residuos, donde gráficamente se observa que el conjunto de datos está distribuido sobre la línea recta, lo que sugiere que los datos se distribuyen normalmente; por otro lado, al realizar la prueba de Shapiro-Wilks, se obtiene que los residuos no tiene una distribución normal.

A partir de las estimaciones, este modelo se puede escribir como:

$$y_t = \begin{pmatrix} 0.05607 \\ 0.013184 \\ 0.02532 \end{pmatrix} + \begin{pmatrix} -0.68437 & -0.056522 & -0.10646 \\ 0.42613 & 0.075018 & 0.3197 \\ -1.0385 & 0.45033 & -0.1532 \end{pmatrix} y_{t-1}.$$

Por lo anterior se recurrió a un modelo de orden p=2, los resultados obtenidos se muestran en la figura 5.5, la cual incluye las estimaciones de los

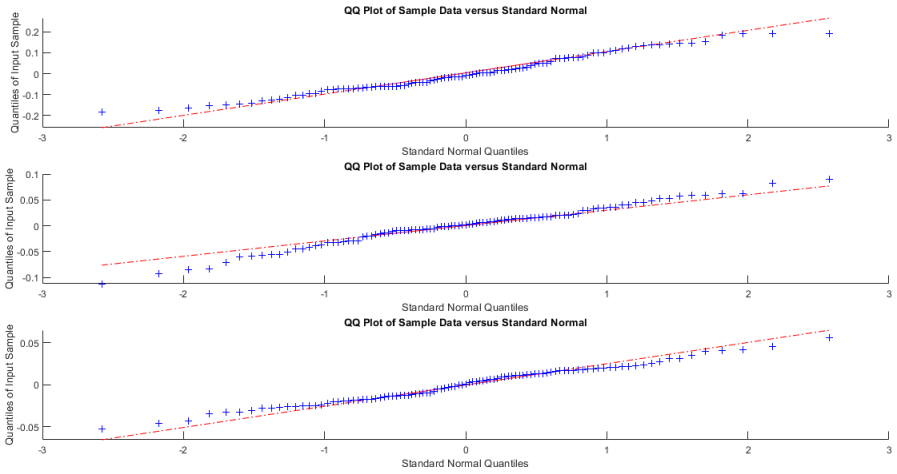

Figura 5.4: Q-Q plot de los residuos para el modelo VAR(1).

parámetros, con los correspondientes errores estándar y p-valores.

Los criterios de selección de orden devuelven los siguientes resultados, AIC=-1052.59 y BIC=-997.674; en este caso los valores obtenidos son menores en el modelo VAR(2) respecto al modelo VAR(1). Al estudiar los residuos, la prueba de Pormanteau indica que no hay autocorrelación entre ellos.

De forma similar al procedimiento anterior, se recurrió a la herramienta Q-Q plot (Figura 5.6), se observa que los datos se ajustan a una distribución normal pues los datos se distribuyen sobre la línea recta, por otro lado, de la prueba de Shapiro-Wilks se obtiene que en efecto los residuos tienen una distribución normal. El modelo resultante queda escrito como:

$$y_t = \begin{pmatrix} 0.010393 \\ 0.0049179 \\ 0.013732 \end{pmatrix} + \begin{pmatrix} -0.61416 & 0.020022 & -0.02934 \\ -0.031053 & 0.025708 & 0.23597 \\ -1.4408 & 0.67415 & 0.016638 \end{pmatrix} y_{t-1}$$

$$+ \begin{pmatrix} -0.27151 & 0.12873 & 0.095684 \\ 0.63235 & -0.058705 & 0.00075403 \\ 2.0562 & 0.024355 & 0.19916 \end{pmatrix} y_{t-2}$$

A partir de los resultados obtenidos del análisis de los modelos, se puede decir que los resultados del modelo VAR(2) mejoran respecto al modelo

AR-Stationary 3-Dimensional VAR(2) Model

Effective Sample Size: 101
Number of Estimated Parameters: 21
LogLikelihood: 547.296
AIC: -1052.59
BIC: -997.674

	Value	StandardError	TStatistic	PValue
Constant(1)	0.010393	0.015304	0.67911	0.49707
Constant(2)	0.0049179	0.0062645	0.78505	0.43242
Constant(3)	0.013732	0.0036179	3.7956	0.0001473
AR{1}(1,1)	-0.61416	0.086192	-7.1255	1.0373e-12
AR{1}(2,1)	0.020022	0.035281	0.56748	0.57039
AR{1}(3,1)	-0.02934	0.020376	-1.4399	0.14989
AR{1}(1,2)	-0.031053	0.26319	-0.11799	0.90608
AR{1}(2,2)	0.025708	0.10773	0.23862	0.8114
AR{1}(3,2)	0.23597	0.062219	3.7926	0.00014909
AR{1}(1,3)	-1.4408	0.43109	-3.3422	0.0008312
AR{1}(2,3)	0.67415	0.17646	3.8204	0.00013322
AR{1}(3,3)	0.016638	0.10191	0.16326	0.87031
AR{2}(1,1)	-0.27151	0.089829	-3.0225	0.0025069
AR{2}(2,1)	0.12873	0.03677	3.501	0.00046356
AR{2}(3,1)	0.095684	0.021236	4.5058	6.6128e-06
AR{2}(1,2)	0.63235	0.2748	2.3011	0.021384
AR{2}(2,2)	-0.058705	0.11248	-0.5219	0.60174
AR{2}(3,2)	0.00075403	0.064963	0.011607	0.99074
AR{2}(1,3)	2.0562	0.40705	5.0515	4.3847e-07
AR{2}(2,3)	0.024355	0.16662	0.14617	0.88379
AR{2}(3,3)	0.19916	0.096228	2.0697	0.038481

Figura 5.5: Estimación usando el modelo VAR(2).

VAR(1), pues los valores resultantes de los criterios de selección de orden son menores en el modelo VAR(2), y además, el modelo VAR(2) sí satisface la normalidad en los residuos que el otro modelo analizado no.

Por lo anterior se eligió trabajar con el modelo VAR(2).

Para los pronósticos se omitieron los últimos cuatro trimestres con el propósito de comparar los datos reales con los datos pronosticados. Además, se pronosticaron cuatro trimestres futuros más, de los cuales no se conocía alguna información. A continuación se muestran en el cuadro 5.1, los errores relativos, expresados en millones de pesos, para las tres variables; en el Cuadro 5.2 se presentan los cuatro pronósticos futuros y finalmente las gráficas que comparan los datos reales con los datos pronosticados. Cabe mencionar que en los resultados presentados, las transformaciones que se hicieron a los datos para conseguir la estacionariedad requerida, ya han sido invertidas.

Se puede decir que, en general, lo errores son pequeños. Note además, que los menores errores relativos corresponden al PIB de AS, seguido de los errores relativos del PIB de AT y finalmente PIB de AP, con los mayores

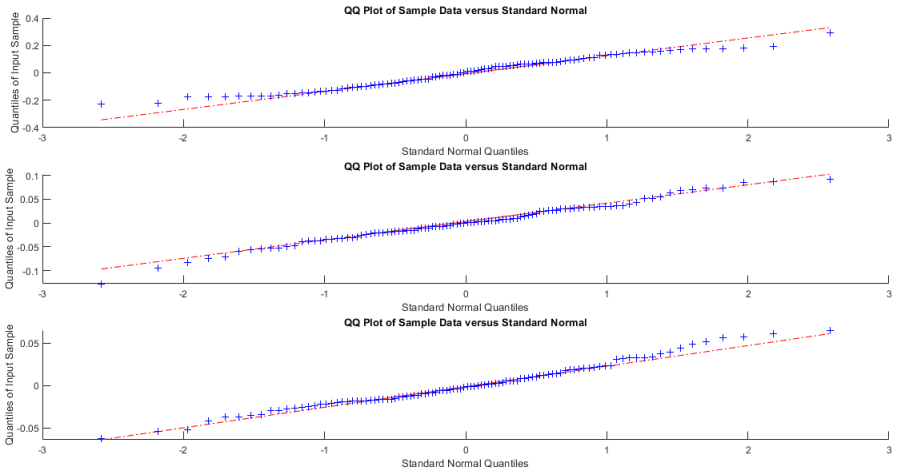

Figura 5.6: Q-Q plot de los residuos en el modelo VAR(2).

Trimestre	PIB AP	PIB AS	PIB AT
T1 19	0.42693	0.04401	0.09467
T2 19	0.14156	0.06749	0.0927
T3 19	0.74057	0.08366	0.12346
T4 19	0.2944	0.09073	0.15381

Cuadro 5.1: Errores Relativos.

errores relativos.

A partir de los resultados obtenidos, se puede concluir que el PIB de las actividades secundarias y terciarias crece para los cuatro trimestres de 2020; mientras que para los cuatro trimestres del mismo año, el PIB de las actividades primarias muestra crecimiento y decrecimiento en distintos puntos. Lo cual de podría indicar noticias positivas para la economía.

Podemos observar en las gráficas de *pronóstico contra valores reales* que los pronósticos más acertados son los del PIB de las actividades secundarias.

Se puede decir que trabajar con modelos VAR es relativamente fácil cuando las variables y el orden del modelo son pequeños, como en este caso.

Además, a partir de la tabla de errores relativos podemos decir que el pronóstico es bueno pues los valores de los errores son pequeños. Lo que nos

Trimestre	PIB AP	PIB AS	PIB AT
T1 20	1252000	8492000	17297000
T2 20	1113000	8711000	17742000
T3 20	1320000	8937000	18203000
T4 20	1175000	9178000	18694000

Cuadro 5.2: Pronósticos para los cuatro trimestres del 2020.

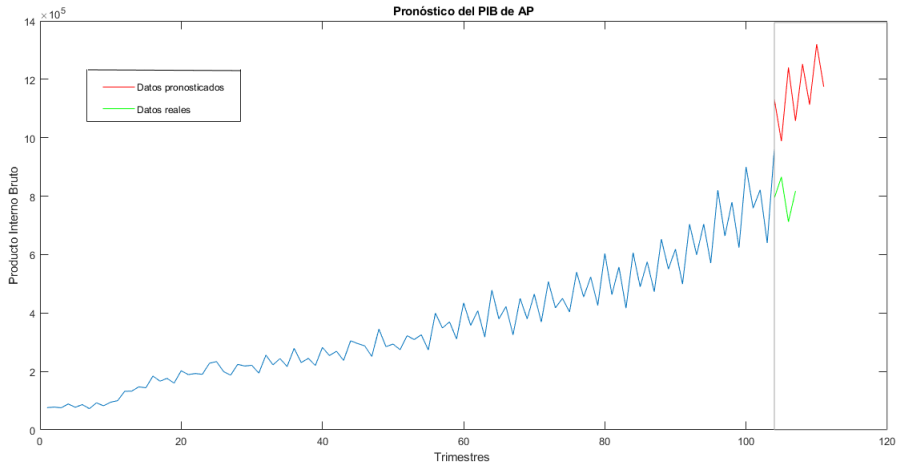

Figura 5.7: Pronóstico contra datos reales del PIB de Actividades Primarias.

permite decir que el modelo VAR es adecuado, al menos para este caso.

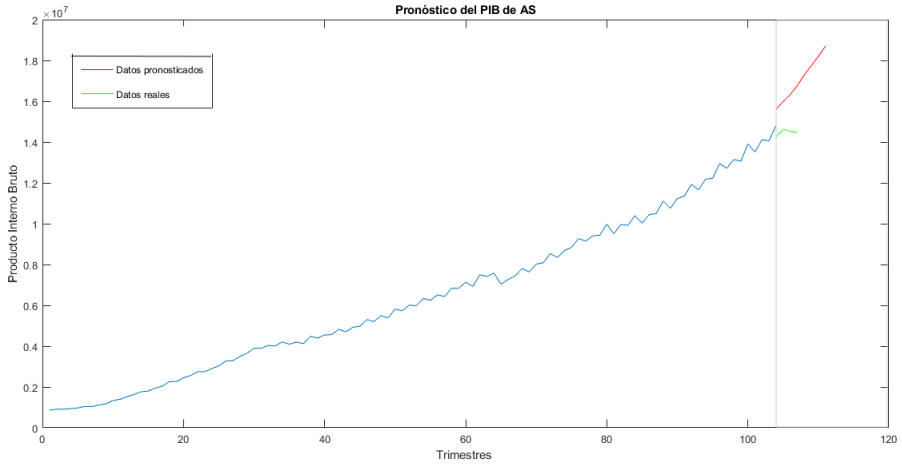

Figura 5.8: Pronóstico contra datos reales del PIB de Actividades Secundarias.

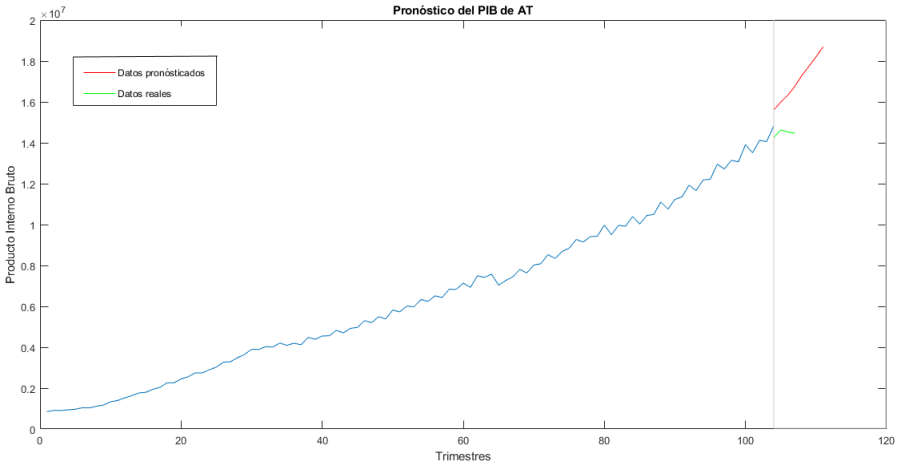

Figura 5.9: Pronóstico contra datos reales de PIB de Actividades Terciarias.

Resumen y conclusiones

En este trabajo se aborda un panorama general de la teoría de modelos autorregresivos vectoriales y se consideró además una aplicación relacionada con el PIB en México.

Primero se presentaron los conceptos y características que tiene un proceso de este tipo; cálculo de autocovarianzas, autocorrelaciones, estabilidad y estacionariedad. Depués, se abordaron métodos de estimación y sus propiedades. En seguida, se trataron métodos para pronosticar y sus cualidades. Además, se abordaron criterios para la selección de orden, y pruebas de normalidad y correlación de los residuos.

Finalmente, se explica el proceso de ajuste de un modelo AR(p) multidimensional, usando el método de Box-Jenkins, que involucra datos de actividades primarias, secundarias y terciarias, en México.

Con esto, se puede concluir que trabajar con un modelo autorregresivo vectorial es relativamente fácil cuando el orden del modelo es pequeño. Además, se puede notar que en el caso específico que se trabaja aquí, los pronósticos no son tan buenos simultáneamente, tienen un orden jerárquico respecto a la bondad. Es posible que éste sea el precio que se tenga que pagar por trabajar con múltiples series.

A partir de lo anterior, se espera que más adelante se pueda analizar e investigar más profundamente acerca de la relación entre la cantidad de series, el orden del proceso y la bondad de los pronósticos.

Apéndice A

Vectores y matrices

A.1. Determinantes

En las siguientes reglas, $A = (a_{ij})$ y $B = (b_{ij})$ son matrices $(m \times M)$ y c es un escalar.

(1) $det(I_m) = 1$.

(2) Si A es una matriz diagonal, $det(A) = a_{11} \cdot a_{22} \cdots a_{mm}$.

(3) Si A es una matriz diagonal superior, $det(A) = a_{11} \cdot a_{22} \cdots a_{mm}$.

(4) Si A tiene una fila o columna de ceros, $det(A) = 0$.

(5) Si A tiene dos filas o columnas iguales, entonces $det(A) = 0$.

(6) $det(cA) = c^m \;\; det(A)$.

(7) $det(AB) = det(A) \;\; det(B)$.

(8) Si C es una matriz $(m \times n)$, entonces $det(I_m + CC') = det(I_n + CC')$.

A.2. Forma Canónica de Jordan

Sea A una matriz de dimkensión $(m \times m)$ con eigenvalores $\lambda_1, \ldots \lambda_n$. Entonces existe una matriz no singular P tal que

$$P^{-1}AP = \begin{bmatrix} \Lambda_1 & & 0 \\ & \ddots & \\ 0 & & \Lambda_n \end{bmatrix} =: \Lambda$$

o $A = P\Lambda P^{-1}$, donde

$$\begin{bmatrix} \lambda_1 & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \ldots & \lambda_i \end{bmatrix}.$$

Esta descomposición es la forma canónica de Jordan. Como los eigenvalores de A pueden ser números complejos, Λ y P pueden ser matrices complejas. Si las raíces múltiples del polinomio característico existen, pueden aparecer más de una vez en la lista $\lambda_1, \ldots \lambda_n$.

La forma canónica de Jordan tiene aplicaciones importantes. Por ejemplo, implica que
$A^j = (P\Lambda P{-}1)^j = P\Lambda^j P^{-1}$
y se puede mostrar que

$$\Lambda_i^j = \begin{bmatrix} {\lambda^j}_i & \binom{j}{1}{\lambda^{j-1}}_i & \ldots & \binom{j}{r_i-1}{\lambda^{j-r_i+1}}_i \\ 0 & {\lambda^j}_i & \ldots & \binom{j}{r_i-1}{\lambda^{j-r_i+1}}_i \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & {\lambda^j}_i \end{bmatrix}$$

donde

$$\binom{p}{q} = \frac{q!}{(p-q)!q!}$$

denota un coeficiente binomial. Tenemos las siguientes reglas.
Reglas: Supongase que A es una matriz real $(m \times m)$ con eigenvalores $\lambda_1, \ldots, \lambda_n$ tal que todos tienen módulos menores que uno, esto es $|\lambda_i| < 1$ para $i = 1, \ldots, n$. Además, sea Λ y P como arriba.

(1) $A^j = P\Lambda^j P^{-1} \longrightarrow 0$.

(2) $\sum_{j=0}^{\infty} A^j = (I_m - A)^{-1}$ existe.

(3) La sucesión A^j, $j = 0, 1, 2, \ldots$, es absolutamente sumable, esto es $\sum_{j=0}^{\infty} |\alpha_{kl,j}|$ es finita para toda $k, l = 1, \ldots, m$, donde $\alpha_{kl,j}$ es un elemento de A^j.

A.3. Valores y vectores propios

Los eigenvalores de una matriz cuadrada $(m \times m)$ son las raices del polinomio en λ dado por $det(A - \lambda I_m)$ o $det(\lambda I_m) - A$. El determinante es frecuentemente llamado el determinante característico y el polinomio es llamado el polinomio característico de A. Un número λ_i es un eigenvalor de A, si las columnas de $(A - \lambda_i I_m)$ son linealmente independientes. Consecuentemente, existe un vector $(m \times 1)$ $v_i \neq 0$ tal que

$$(A - \lambda_i I_m)v_i = o \quad o \quad Av_i = \lambda_i v_i \tag{A.3.1}$$

Un vector con esta propiedad es un eigenvector o vector característico de A asociado con el eigenvalor λ_i. Por supuesto, cualquier múltiplo escalar de v_i distinto de cero también es un eigenvector de A asociado a λ_i.

Reglas:

(1) Si A es simétrica, entonces todos los eigenvalores son números reales.

(2) Los eigenvalores de una matriz diagonal son sus elementos diagonales.

(3) Los eigenvalores de una matriz triángular son sus elementos diagonales.

(4) Una matriz $(m \times m)$ tiene a lo más m eigenvalores.

(5) Sean $\lambda_1, \ldots, \lambda_m$ los eigenvalores de la matriz $(m \times m)$ A, entonces $|A| = \lambda_1 \cdots \lambda_m$, esto es, el determinante es el producto de los eigenvalores.

(6) Sean λ_i y λ_i eigenvalores ditintos de A asociados con los vectores propios v_i y v_j respectivamente. Entonces v_i y v_j son linealmente independientes.

(7) Todos los eigenvalores de la matriz A $(m \times m)$ tiene módulos menores que 1 si y sólo si $det(I_m - A_z) \neq 0$ para $|z| \leq 1$, esto es, el polinomio $det(I_m - A_z) \neq 0$ tiene raíces no complejas en y sobre el círculo unitario.

A.4. El producto de Kronecker

Sea $A = (a_{ij})$ y $B = b_{ij}$ matrices $(m \times n)$ y $(p \times q)$, respectivamnete. La matriz $(mp \times nq)$

$$A \otimes B := \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix} \tag{A.4.1}$$

es el producto de Kronecker o producto directo de A y B.
Reglas: En las siguientes reglas, se asumen dimensiones adecuadas.

(1) $A \otimes B \neq B \otimes A$ en general.

(2) $(A \otimes B)' = A' \otimes B'$.

(3) $A \otimes (B + C) = A \otimes B + A \otimes C$.

(4) $(A \otimes B)(C \otimes D) = AC \otimes BD$.

(5) Si A y B son invertibles, entonces $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

(6) Si A y B son matrices cuadradas con eigenvalores λ_A, λ_B, respectivamente, y los eigenvectores asociados v_A, v_B respectivamente, entonces $\lambda_A \lambda_B$ es un eigenvalor de $A \otimes B$ con eigenvector $v_A \otimes v_B$.

(7) Si A y B son matrices cuadradas $(m \times m)$ y $(n \times n)$, respectivamente, entonces $|A \otimes B| = |A|^n |B|^m$.

(8) Si A y B son matrices cuadradas,
$tr(A \otimes B) = tra(A)tra(B)$.

(9) $(A \otimes B)^+ = A^+ \otimes B^+$.

A.5. Los operadores vec y vech, y las matrices relacionadas

Sea $A = (a_1, ..., a_n)$ una matriz $(m \times n)$ con columnas a_1 $(m \times 1)$, el operador *vec* transforma a A en un vector $(mn \times 1)$ al apilar las columnas, esto es,

$$vec(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

Reglas asociadas al operador vec: Sean A, B, C matrices con dimensiones apropiadas

(1) $vec(A + B) = vec(A) + vec(B)$.

(2) $vec(ABC) = (C' \otimes A)vec(B)$.

(3) $vec(AB) = (I \otimes A)vec(B) = (B' \otimes I)vec(A)$.

(4) $vec(ABC) = (I \otimes AB)vec(C) = (C'B' \otimes I)vec(A)$.

(5) $vec(B')'vec(A) = tr(BA) = tr(AB) = vec(A')'vec(B)$.

(6)

$$\begin{aligned} tr(ABC) &= vec(A')'(C' \otimes I)vec(B) \\ &= vec(A')'(I \otimes B)vec(C) \\ &= vec(B')'(A' \otimes I)vec(C) \\ &= vec(B')'(I \otimes C)vec(A) \\ &= vec(C')'(B' \otimes I)vec(A) \\ &= vec(C')'(I \otimes A)vec(B). \end{aligned}$$

El operador *vech* está relacionado con el operador *vec*. Solo se apilan los elementos en y debajo de la diagonal principal de una matriz cuadrada.

$$vech \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \\ \alpha_{22} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix} \tag{A.5.1}$$

En general, si A es una matriz de dimensión $(m \times M)$, $vech(A)$ es un vector de dimensión $m(m+1)/2$.

A.5.1. Matrices de eliminación, duplicación y conmutación

Los operadores *vec* y *vech* están relacionados por la matriz de eliminación, $\boldsymbol{L}_m$, y la matriz de duplicación, $\boldsymbol{D}_m$. La primera es una matriz de dimensión $(\frac{1}{2}m(m+1) \times m^2)$ tal que, para una matriz cuadrada A de dimensión $(m \times m)$,

$$vech(A) = \boldsymbol{L}_m vec(A). \tag{A.5.2}$$

La matriz de duplicación $\boldsymbol{D}_m$ es de dimensión $(m^2 \times \frac{1}{2}m(m+1))$ y es definida de tal forma que, para cualquier matriz simétrica A de dimensión $(m \times m)$,

$$vec(A) = \boldsymbol{D}_m vech(A). \tag{A.5.3}$$

Como el rango de $\boldsymbol{D}_m$ es $m(m+1)/2$, la matriz $\boldsymbol{D}_m^{'}\boldsymbol{D}_m$ es invertible. Entonces, multiplicando (A.5.3) por $(\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m^{'}$, se tiene

$$(\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m^{'} vec(A) = vech(A). \tag{A.5.4}$$

Sin embargo, $(\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m^{'} \neq \boldsymbol{L}_m$ en general, pues (A.5.4) se cumple sólo para matrices simétricas, mientras que (A.5.2) se cumple para cualquier matriz cuadrada.

La matriz de conmutación, $\boldsymbol{K}_{mn}$, es otra matriz que ocasionalmente es útil para tratar con el operador *vec*. $\boldsymbol{K}_{mn}$ es una matriz de dimensión $(mn \times mn)$ definida de tal forma que para cualquier matriz A de dimensión $(m \times n)$,

$$vec(A') = \boldsymbol{K}_{mn} vec(A)$$

o, equivalentemente,

$$vec(A) = \boldsymbol{K}_{mn} vec(A').$$

Reglas:

(1) $\boldsymbol{L}_m \boldsymbol{D}_m = I_{m(m+1)/2}$.

(2) $\boldsymbol{K}_{mm} \boldsymbol{D}_m = \boldsymbol{D}_m$.

(3) $\boldsymbol{K}_{m1} = \boldsymbol{K}_{1m} = I_m$.

(4) $\boldsymbol{K}'_{mn} = \boldsymbol{K}_{mn}^{-1} = \boldsymbol{K}_{nm}$

(5) $tr\ \boldsymbol{K}_{mm} = m$.

(6) $det(\boldsymbol{K}_{mn}) = (-1)^{mn(m-1)(n-1)/4}$.

(7) $tr(\boldsymbol{D}_m^{'}\boldsymbol{D}_m) = m^2,\ tr(\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1} = m(m+3)/4$.

(8) $det(\boldsymbol{D}_m^{'}\boldsymbol{D}_m) = 2^{m(m-1)/2}$.

(9) $tr(\boldsymbol{D}_m\boldsymbol{D}_m^{'}) = m^2$.

(10) $\left|\boldsymbol{D}_m(A \otimes A)\boldsymbol{D}_m^{'}\right| = 2^{m(m-1)/2}\left|A\right|^{m+1}$, donde A es una matriz de dimensión $(m \times m)$.

(11) $(\boldsymbol{D}_m^{'}(A \otimes A)\boldsymbol{D}_m)^{-1} = (\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m^{'}(A^{-1} \otimes A^{-1})\boldsymbol{D}_m(\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1}$, si A es una matriz no singular de dimensión $(m \times m)$.

(12) $\boldsymbol{L}_m\boldsymbol{L}_m^{'} = I_{m(m+1)/2}$.

(13) $\boldsymbol{L}_m\boldsymbol{L}_m^{'}$ y $\boldsymbol{L}_m\boldsymbol{K}_{mm}\boldsymbol{L}_m^{'}$ idempotente.

Sean A y B matrices triangulares inferiores. Se cumplen las siguientes reglas:

(14) $\boldsymbol{L}_m(A \otimes B)\boldsymbol{L}_m^{'}$ es una matriz triangular inferior.

(15) $\boldsymbol{L}_m^{'}\boldsymbol{L}_m(A' \otimes B)\boldsymbol{L}_m^{'} = (A' \otimes B)\boldsymbol{L}_m^{'}$.

(16) $\left[\boldsymbol{L}_m(A' \otimes B)\boldsymbol{L}_m^{'}\right]^s = \boldsymbol{L}_m((A')^s \otimes B^s)\boldsymbol{L}_m^{'}$ para $s = 0, 1, \ldots$ y para $s = \ldots, -2, -1$, si A^{-1} y B^{-1} existen.

Sean G, F y $\boldsymbol{b}$ de dimensión $(m \times n), (p \times q)$ y $(p \times q)$. Entonces los siguientes resultados se cumplen:

(17) $\boldsymbol{K}_{pm}(G \otimes F) = (F \otimes G)\boldsymbol{K}_{qn}$.

(18) $\boldsymbol{K}_{pm}(G \otimes F)\boldsymbol{K}_{nq} = F \otimes G$.

(19) $\boldsymbol{K}_{pm}(G \otimes \boldsymbol{b}) = \boldsymbol{b} \otimes G$.

(20) $\boldsymbol{K}_{pm}(\boldsymbol{b} \otimes G) = G \otimes \boldsymbol{b}$.

(21) $vec(G \otimes F) = (I_n \otimes \boldsymbol{K}_{qm} \otimes I_p)(vec(G) \otimes vec(F))$.

(22) $(\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m^{'}\boldsymbol{K}_{mm} = (\boldsymbol{D}_m^{'}\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m^{'}$.

A.6. Diferenciación de matrices y vectores

En lo siguiente, se asumirá que todas las derivadas existen y son continuas. Sea $f(\beta)$ una función escalar que depende del vector $\beta = (\beta_1, \ldots, \beta_n)$. $(n \times 1)$

$$\frac{\partial f}{\partial \beta} := \begin{bmatrix} \frac{\partial f}{\partial \beta_1} \\ \vdots \\ \frac{\partial f}{\partial \beta_n} \end{bmatrix}, \frac{\partial f}{\partial \beta'} := \begin{bmatrix} \frac{\partial f}{\partial \beta_1}, \cdots, \frac{\partial f}{\partial \beta_n} \end{bmatrix}$$

son vectores de derivadas parciales de primer orden, de dimensión $(n \times 1)$ y $(1 \times n)$, respectivamente,

$$\frac{\partial^2 f}{\partial \beta \partial \beta'} := \begin{bmatrix} \frac{\partial^2 f}{\partial \beta_i \beta_j} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \beta_n \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_n \partial \beta_n} \end{bmatrix}$$

es la *matriz Hessiana* $(n \times n)$ de derivadas parciales de segundo orden. Si $f(A)$ es una función escalar de una matriz $A = (a_{ij})$, entonces

$$\frac{\partial f}{\partial A} := \begin{bmatrix} \frac{\partial f}{\partial a_{ij}} \end{bmatrix}$$

es una matriz de derivadas parciales, de dimensión $(m \times n)$. Si la matriz A depende del escalar β, entonces

$$\frac{\partial A}{\partial \beta} := \begin{bmatrix} \frac{\partial a_{ij}}{\partial \beta} \end{bmatrix}$$

es una matriz de dimensión $(m \times n)$. Si $y(\beta) = (y_1(\beta), \ldots, y_m(\beta))'$ es un vector de dimensión $(m \times 1)$ que depende del vector β de dimensión $(n \times 1)$, entonces

$$\frac{\partial^2 y}{\partial \beta'} := \begin{bmatrix} \frac{\partial y_1}{\partial \beta_1} & \cdots & \frac{\partial y_1}{\beta_n} \\ \vdots & & \vdots \\ \frac{\partial y_m}{\beta_1} & \cdots & \frac{\partial y_m}{\beta_n} \end{bmatrix}$$

es una matriz de dimensión $(m \times n)$ y

$$\frac{\partial y'}{\partial \beta} := \left(\frac{\partial y}{\partial \beta'} \right)'.$$

Las siguientes dos proposiciones son necesarias para derivar reglas de diferenciación de matrices y vectores.

Proposición A.1 (Regla de la cadena para la diferenciación vectorial)**.** *Sean α y β vectores de dimensión $(m \times 1)$ y $(n \times 1)$, respectivamente, y suponga que $h(\alpha)$ y $g(\beta)$ son de dimensión $(p \times 1)$ y $(m \times 1)$, correspondientemente. Entonces, con $\alpha = g(\beta)$,*

$$\frac{\partial (g(\beta))}{\partial \beta'} = \frac{\partial (h)}{\partial \alpha'} \frac{\partial g(\beta)}{\partial \beta'} \quad (p \times n).$$

Proposición A.2 (Reglas del producto para la diferenciación vectorial)**.** *(a) Si β $(m \times 1)$, $a(\beta) = (a_1(\beta), \ldots, a_n(\beta))'$ y $c(\beta) = (c_1(\beta), \ldots, c_p(\beta))'$ son de dimensión $(m \times 1)$, $(n \times 1)$ y $(p \times 1)$, respectivamente, y $A = (a_{ij})$ de dimensión $(n \times p)$ no depende de β. Entonces*

$$\frac{\partial \left[a(\beta)' A c(\beta)\right]}{\partial \beta'} = c(\beta)' A' \frac{\partial a(\beta)}{\partial \beta'} + a(\beta)' A \frac{c(\beta)}{\partial \beta'}.$$

(b) Si β es un escalar, $A(\beta)$ es de dimensión $(m \times n)$ y $B(\beta)$ es de dimensión $(n \times q)$, entonces

$$\frac{\partial AB}{\partial \alpha \beta} = \frac{\partial A}{\partial \beta} B + A \frac{\partial B}{\partial \beta}.$$

(c) Si β, $A(\beta)$ y $B(\beta)$ son de dimensión $(m \times 1)$, $(m \times n)$ y $(n \times q)$, respectivamente, entonces

$$\frac{\partial vec(AB)}{\partial \beta'} = (I_q \otimes A) \frac{\partial vec(B)}{\partial \beta'} + (B' \otimes I_n) \frac{\partial vec(A)}{\partial \beta'}.$$

Ahora las siguientes reglas se pueden verificar:

(1) Sea A una matriz y β un vector de dimensión $(m \times n)$ y $(n \times 1)$, respectivamente. Entonces

$$\frac{\partial A\beta}{\partial \beta'} = A \qquad y \qquad \frac{\partial \beta' A'}{\partial \beta} = A'.$$

(2) Sean A y β de dimensión $(m \times m)$ y $(m \times 1)$, respectivamente. Entonces

$$\frac{\partial \beta' A \beta}{\partial \beta} = (A + A')\beta \qquad \frac{\partial \beta' A \beta}{\partial \beta'} = \beta'(A + A').$$

(3) Si A y β son de dimensión $(m \times m)$ y $(m \times 1)$, respectivamente, entonces

$$\frac{\partial^2 \beta' A \beta}{\partial \beta \partial \beta'} = A + A'.$$

(4) Si A es una matriz simétrica de dimensión $(m \times m)$ y β es un vector de dimensión $(m \times 1)$, entonces

$$\frac{\partial^2 \beta' A \beta}{\alpha \beta \alpha \beta'} = 2A.$$

(5) Sea Ω una matriz simétrica $(n \times n)$ y $c(\beta)$ un vector de dimensión $(n \times 1)$ que depende del vector β de dimensión $n \times 1$. Entonces

$$\frac{\partial c(\beta)' \Omega c(\beta)}{\partial \beta'} = 2c(\beta)' \Omega \frac{c(\beta)}{\partial \beta'}$$

y

$$\frac{\partial c(\beta)' \Omega c(\beta)}{\partial \beta'} = 2 \left[\frac{\partial c(\beta)'}{\partial \beta} \Omega \frac{\partial c(\beta)'}{\partial \beta'} + [c(\beta)' \Omega \otimes I_m] \frac{\partial vec(\partial c(\beta)' / \alpha \beta}{\partial \beta'} \right].$$

En particular, si y es un vector y X una matriz de dimensión $(n \times 1)$ y $(n \times m)$,

$$\frac{\partial (y - X\beta)' \Omega (y - X\beta)}{\partial \beta'} = -2(y - X\beta)' \Omega X$$

y

$$\frac{\partial^2 (y - X\beta)' \Omega (y - X\beta)}{\partial \beta \partial \beta'} = 2X' \Omega X.$$

(6) Suponga que β, $B(\beta)$ son de dimensión $(m \times 1)$, $(n \times p)$, respectivamente, y A y C son matrices de dimensión $(k \times n)$ y $(p \times q)$, que no dependen de β. Entonces

$$\frac{\partial vec(ABD)}{\partial \beta'} = (C' \otimes A) \frac{\partial vec(B)}{\partial \beta'}.$$

(7) Suponga que β, $A(\beta)$ y $D(\beta)$ son de dimensión $(m \times 1)$, $(n \times p)$ y $(q \times r)$, respectivamente, y C de dimensión $(p \times q)$ no depende de β. Entonces

$$\frac{\partial vec(ACD)}{\partial \beta'} = (I_r \otimes AC) \frac{\partial vec(D)}{\partial \beta'} + (D'C' \otimes I_n) \frac{\partial vec(A)}{\partial \beta'}.$$

(8) Si β y $A(\beta)$ de dimensión $(m \times 1)$ y $(n \times n)$, entonces para cualquier entero positivo h,

$$\frac{\partial vec(A^h)}{\partial \beta'} = \left[\sum_{i=0}^{h-1} (A')^{h-1-i} \otimes A^i\right] \frac{vec(A)}{\partial \beta'}$$

(9) Si A es una matriz no singular de dimensión $(m \times m)$, entonces

$$\frac{\partial vec(A^{-1})}{\partial vec(A)'} = -(A^{-1})' \otimes A^{-1}.$$

(10) Si $A = (a_{ij})$ es una matriz de dimensión $(m \times m)$, entonces

$$\frac{\partial tr(A)}{\partial A} = I_m.$$

(11) Si $A = (a_{ij})$ y $B = (b_{ij})$ son de dimensión $(m \times n)$ y $(n \times m)$, entonces

$$\frac{\partial tr(AB)}{\partial A} = B'.$$

(12) Suponga que A es una matriz de dimensión $(m \times n)$ y, B y C son de dimensión $(m \times m)$ y $(n \times m)$, respectivamente. Entonces

$$\frac{\partial tr(ABC)}{\partial A} = B'C'.$$

(13) Sean A, B, C, D de dimensión $(m \times n)$, $(n \times n)$, $(m \times n)$ y $(n \times m)$, respectivamente. Entonces

$$\frac{\partial tr(DABA'C)}{\partial A} = CDAB + D'C'AB'.$$

(14) Sean A, B y C matrices de dimensión $(m \times m)$ y suponga que A es una matriz no singular. Entonces

$$\frac{\partial tr(BA^{-1}C)}{\partial A} = -(A^{-1}CBA^{-1})'.$$

(15) Sea $A = (a_{ij})$ una matriz de dimensión $(m \times m)$. Entonces

$$\frac{\partial |A|}{\partial A} = \left(A^{adj}\right)',$$

donde A^{adj} es la matriz adjunta de A.

(16) Si A es una matriz no singular de dimensión $(m \times m)$ con $|A| > 0$, entonces

$$\frac{\partial ln\,|A|}{\partial A} = (A')^{-1}.$$

Apéndice B

Convergencia estocástica y distribuciones asintóticas

B.1. Conceptos de convergencia estocástica

Sea x_T, $T = 1, 2, \cdots$, una sucesión de variables aleatorias escalares definidas en el espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$. La sucesión x_T converge en probabilidad a la variable aleatoria x, definida en el mismo espacio de probabilidad, si para todo $\epsilon > 0$.

$$\lim_{x \to \infty} \mathcal{P}(|x_T - x| > \epsilon|) = 0 \tag{B.1.1}$$

o, equivalentemente,

$$\lim_{T \to \infty} \mathcal{P}(|x_T - x| < \epsilon|) = 1. \tag{B.1.2}$$

Este tipo de convergencia estocástica es abreviada como

$$plim\ x_T = x \quad o \quad x_T \xrightarrow{p} y_t. \tag{B.1.3}$$

La sucesión x_T converge con probabilidad uno a la variable aleatoria x, si para todo $\epsilon > 0$

$$\mathcal{P}\left(\lim_{T \to \infty} |x_T - x| < \epsilon\right) = 1. \tag{B.1.4}$$

Este tipo de convergencia estocástica puede escribirse como $x_T \xrightarrow{a.s} y_t$.

La sucesión x_T converge en media cuadrática a x, $x_T \xrightarrow{c.m} y_t$, si

$$\lim_{T \to \infty} E(x_T - x)^2 = 0. \tag{B.1.5}$$

Este tipo de convergencia requiere que la media y la varianza de las variables aleatorias exista.

Proposición B.1 (Propiedades de convergencia en probabilidad y en Distribución)**.** *Suponga que x_T y y_T son sucesiones de vectores aleatorios $(K \times 1)$, A_T es una sucesión de matrices aleatorias $(K \times K)$, x es un vector aleatorio $(K \times 1)$, c es un vector fijo $(K \times 1)$ y A es una matriz fija $K \times K$*

(1) Si $plim x_T$, $plim y_T$, y $plim A_T$ existe, entonces

- *(a) $plim(x_T \pm y_T) = plim x_T \pm plim y_T$;*
- *(b) $plim(c' x_T) = c'(plim x_T)$;*
- *(c) $plim x'_T y_T = (plim x_T)'(plim y_T)$;*
- *(d) $A_T x_T = plim(A_T) plim(x_T)$.*

(2) Si $x_T \xrightarrow{d} x$ y $plim(x_T - y_T) = 0$, entonces $y_T \to^d x$.

(3) si $x_T \xrightarrow{d} x$ y $plim\ y_T = c$, entonces

- *(a) $x_T \pm y_T \xrightarrow{d} x \pm c$;*
- *(b) $y'_T x_T \xrightarrow{d} c' x$.*

(4) Si $x_T \xrightarrow{d} x$ y $plim A_T = A$, entonces $A_T x_T \xrightarrow{d} Ax$.

(5) Si $x_T \xrightarrow{d} x$ y $plim A_T = 0$, entonces $plim A_T x_T = 0$.

B.2. Sumas infinitas de variables aleatorias

La representación MA de un proceso VAR es frecuentemente una suma infinita de vectores aleatorios. Como en un estudio de sumas infinitas de número reales, debemos especificar qué queremos decir con una suma infinita. El concepto de convergencia absoluta es básico en lo siguiente. Una sucesión doblemente infinita de números reales $\{a_i\}$, $i = 0, \pm 1, \pm 2, \ldots$, es absolutamente sumable si

$$\lim_{x \to \infty} \sum_{i=-n}^{n} |a_i| \tag{B.2.1}$$

existe y es finito. El límite es usualmente denotado por $\sum_{i=-\infty}^{\infty} |a_i|$.

Proposición B.2. *Supongamos que A_i es una sucesión de matrices reales $(k \times k)$ absolutamente sumable y z_t es una sucesión de variables aleatorias K-dimensionales que satisfacen*
$E(z_t^{'} z_t) \leq c$, $t = 0 \pm 1, \pm 2, \ldots,$
para alguna constante finita c. Entonces existe una sucesión de variables aleatorias k-dimensionales y_t tal que

$$\sum_{i=-n}^{n} nA_i z_{t-1} \xrightarrow[n\to\infty]{c.m} y_t. \tag{B.2.2}$$

La sucesión se determina de manera única, excepto en un conjunto de probabilidad cero.

Proposición B.3. *Sean z_t una sucesión de variables aleatorias K-dimensionales que satisfacen*

$$E(z_t^{'} z_t) \leq c, \quad t = 0 \pm 1, \pm 2, \ldots, \tag{B.2.3}$$

A_i, B_i sucesiones de matrices $(K \times K)$ absolutamente sumables y

$$y_t = \sum_{i=-\infty}^{\infty} A_i z_{t-i} \quad y \quad x_t = \sum_{i=-\infty}^{\infty} B_i z_{t-i}, \tag{B.2.4}$$

entonces

$$E(y_t) = \lim_{n\to\infty} \sum_{i=-n}^{n} A_i E(z_{t-i}) \tag{B.2.5}$$

y

$$E(y_t x_t^{'}) = \lim_{n\to\infty} \sum_{i=-n}^{n} A_i z_{t-1} \sum_{j=-n}^{n} A_i E(z_{t-1} z_{t-1}') B'_j, \tag{B.2.6}$$

donde el límite de la sucesión de matrices es la matriz de los límites de las sucesiones de los elementos individuales.

Proposición B.4 (Leyes débiles de los grandes números)**.** *(1) (Teorema de Khinchine)*
Sea $\{x_t\}$ una sucesión de variables aleatorias i.i.d con $E(x_t) = \mu < \infty$.
Entonces
$\overline{x_T} := \frac{1}{T} \sum_{t=1}^{T} x_t \xrightarrow{p} \mu$.

(2) Sea $\{x_t\}$ una sucesión de variables aleatorias con $E(x_t) = \mu < \infty$ y $E|x_T|^{1+\epsilon} \leqslant c < \infty$(t=1,2,...) para algún $\epsilon > 0$ y una constante finita c. Entonces $\overline{x_T} \xrightarrow{d} \mu$.

(3) (Teorema de Chebyshev) Sea $\{x_t\}$ *una sucesión de variables aleatorias no correlacionadas con* $E(x_t) = \mu < \infty$ *y* $lim_{T\to\infty}E(\overline{x}_T - \mu)^2 = 0$ *. Entonces* $\overline{x}_T \xrightarrow{d} \mu$.

(4) (Corolario del Teorema de Chebyshev) Sea $\{x_t\}$ *una sucesión de variables aleatorias independientes con* $E(x_t) = \mu < \infty$ *y* $Var(x_t) \leq c < \infty$ $(t = 1, 2, ...)$ *para alguna constante finita c.*
Entonces $\overline{x}_T \xrightarrow{d} \mu$.

(5) (LLN para diferencias de Martingala) Sea $\{x_t\}$ *una sucesión de diferencia de Martingala, estrictamente estacionaria, con* $E|x_t| < \infty$ *(t=1,2,...). Entonces* $\overline{x}_T \xrightarrow{d} 0$.

(6) (Proceso estacionario) Sea $\{x_t\}$ *un proceso etocástico con* $E(x_t) = \mu < \infty$ *y* $E\left[(x_t - \mu)(x_{t-j} - \mu)\right] = \gamma_j (t = 1, 2, ...)$ *tal que* $\sum_{j=0}^{\infty} |\gamma_j| < \infty$. *Entonces* $\overline{x}_T \xrightarrow{c.m} \mu$ *y, por lo tanto* $\overline{x}_T \xrightarrow{p} \mu$, *y* $lim_{T\to\infty}TE(\overline{x}_T - \mu)^2 = \sum_{j=-\infty}^{\infty} \gamma_j$.

B.3. Leyes de grandes números y teoremas del límite central

Proposición B.5 (Teoremas del límite central)**.** *(1) (Lindeberg-LevyCLT) Sea* x_t *una sucesión de vectores aleatorios* $K-$*dimensionales i.i.d con media* μ *y matriz de covarianza* Σ_x

$$\sqrt{T}(\overline{x}_T - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma_x).$$

(2) (CLT para matrices de diferencia de Martingala)
Sea $\{x_{T,t} = (x_{1T,t}, \cdots, x_{KT,t})'\}$ *una matriz de diferencias de martigala con matriz de covarianza* $E(x_{T,t}x'_{T,t}) = \Sigma_{Tt}$ *tal que* $T^{-1}\sum_{t=1}^{T}\Sigma_{Tt} \to \Sigma$, *donde* Σ *es definida positiva. Además, supongamos que* $T^{-1}\sum_{t=1}^{T} x_{T,t}x'_{T,t} \xrightarrow{p} \Sigma$ *y* $E(x_{iT,t}x_{jT,t}x_{kT,t}x_{lT,t}) < \infty$ *para todo t y T y todo* $1 \leq i, j, k, l \leq K$. *Entonces*

$$\sqrt{T}(\overline{x}_T - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sum_{j=-\infty}^{\infty} \Gamma_x(j)\right),$$

donde $\Gamma_x(j) := E\left[(x_t - \mu)(x_{t-j} - \mu)'\right]$.

B.4. Propiedades asintóticas estándar de estimadores y estadísticas de prueba

Proposición B.6 (Propiedades Asintóticas de los estimadores). *Sea $\widehat{\beta}$ un estimador del vector β con $\sqrt{T}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}(0,\Sigma)$. Entonces se cumplen las siguientes reglas:*

(1) Si $plim\widehat{A} = A$ entonces $\sqrt{T}\widehat{A}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}(0, A\Sigma A')$.

(2) Si $R \neq 0$ es una matriz $(M \times K)$, entonces β con $\sqrt{T}\left(R\widehat{\beta}-R\beta\right) \xrightarrow{d} \mathcal{N}(0, R\Sigma R')$.

(3) Si $g(\beta) = (g_1(\beta), \cdots, g_m(\beta))'$ es una función vectorial continuamente diferenciable con $\partial g/\partial\beta' \neq 0$, entonces

$$\sqrt{T}\left[g(\widehat{\beta}-g(\beta))\right] \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial g(\beta)}{\beta'}\Sigma\frac{\partial g(\beta)'}{\beta}\right).$$

Si $\frac{\partial g(\beta)}{\beta'} = 0$, $\sqrt{T}\left[g(\widehat{\beta}-g(\beta))\right] \xrightarrow{d} 0$.

(4) Si Σ es no singular, $T(\widehat{\beta}-\beta)'\Sigma^{-1}(\widehat{\beta}-\beta) \xrightarrow{d} \chi^2(K)$.

(5) Si Σ es no singular y $plim\widehat{\Sigma} = \Sigma$, entonces $T(\widehat{\beta}-\beta)'\widehat{\Sigma}^{-1}(\widehat{\beta}-\beta) \xrightarrow{d} \chi^2(K)$.

(6) Si $\Sigma = QA$, donde Q es simétrica, idempotente de rango n y A es definida positiva, entonces $T(\widehat{\beta}-\beta)'A^{-1}(\widehat{\beta}-\beta) \xrightarrow{d} \chi^2(n)$.

Apéndice C

Distribuciones normales multidimensionales

C.1. Normal Multivariada

Un vector K-dimensional de variables aleatorias continuas $y = (y_1, \ldots, y_k)'$ tiene una distribución normal multivariante con vector media $\mu = (\mu_1, \ldots, \mu_K)$ y matriz de covarianza Σ, esto es,
$u_t \sim N(\mu, \Sigma)$,
si su distribución tiene la función de densidad de probabilidad

$$f(y) = \frac{1}{(2\pi)^{K/2}} |\Sigma|^{-1/2} exp\left[-\frac{1}{2}(y-\mu)'\Sigma^{-1(y-\mu)} \right]. \tag{C.1.1}$$

Alternativamente, $u_t \sim N(\mu, \Sigma)$, para cualquier vector de dimensión K, c, tal que $c'y \sim N(c'\mu, c'\Sigma c)$.
Los siguientes resultados respecto a la normal multivariada y distribuciones relacionadas son muy útiles

Proposición C.1 (Distribución marginal de una normal multivariada)**.** *Sean y_1 y y_2 dos vectores aleatorios tales que* $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$, *entonces* $y_1 \sim N(\mu_1, \Sigma_{11})$.

Proposición C.2 (Transformación lineal de un vector aleatorio normal)**.** *Supongamos que $y \sim N(\mu, \Sigma)$ es de dimensión $(K \times 1)$, A es una matriz de dimensión $(M \times K)$ y c un vector de dimensión $(M \times 1)$. Entonces*
$x = Ay + c \sim N(A\mu + c, A\Sigma A')$.

Proposición C.3 (Transformación lineal de un vector aleatorio normal multivariante)**.** *Suponga que* $y \backsim \mathcal{N}(\mu, \Sigma)$ *es* $(K \times 1)$*,* A *es una matriz* $(M \times K)$ *y* c *un vector* $(M \times 1)$*. Entonces*
$x = Ay + c \sim \mathcal{N}(A\mu + c, A\Sigma A')$.

C.2. Distribuciones Relacionadas

Supongase que $y \sim N(0, I_K)$. La distribución de $z = y'y$ es una distribución chi-cuadrada con K grados de libertad,
$z \sim \chi^2(K)$.

Proposición C.4 (Distribuciones de Formas cuadráticas)**.** *(1) Suponga que* $y \sim N(0, I_K)$ *y* A *es una matriz de dimensión* $(K \times K)$ *simétrica e idempotente con* $rang\,(A) = n$*. Entonces* $y'Ay \sim \chi^2(n)$.

(2) Si $y \sim N(0, \Sigma)$*, donde* Σ *es una matriz de dimensión* $(K \times K)$ *definida positiva, entonces* $y'\Sigma^{-1}y \sim \chi^2(K)$.

(3) Sea $y \sim N(0, QA)$*, donde* Q *es una matriz idempotente de dimensión* $(K \times K)$ *con* $rang(Q) = n$ *y* A *es una matriz de dimensión* $(K \times K)$ *definida positiva. Entonces* $y'A^{-1}y \sim \chi^2(n)$.

(4) Sea $y \sim N(0, \Sigma)$*, donde* Σ *es la matriz de covarianza de dimensión* $(K \times K)$*, no singular. Además, sea* A *una matriz de dimensión* $(K \times K)$ *no singular con* $rang(A) = n$*. Entonces* $y'Ay \sim \chi^2(n) \Rightarrow A\Sigma A = A$
y $A\Sigma A = A \Rightarrow y'Ay \sim \chi^2(n)$.

Bibliografía

[1] Akaike, H., *Information theory and an extension of the maximum likelihood principle*, en B.N. Petrov y F. Csáki (eds), 2nd International Symposium on Information Theory, 1973.

[2] Aragón, M.G., *Análisis de la radiación solar en el municipio de Tlaxco-Tlaxcala usando la metodología de BOx-Jekins(Tesis de licenciatura en matemáticas*, Tesis de licenciatura, 2018.

[3] Hannan, E.J., *The statistical theory of linear systems*, en P.R Krishnaiah (ed.), Developments in Statistics, Academic Press, New York, 1979, pp. 83-121.

[4] López, Ana M. *El papel de la información económica como generador de conocimiento en el proceso de predicción: Comparaciones empíricas del crecimiento del PIB regional. Estudios de Economía Aplicada*, Fecha de Consulta: 23 de Enero de 2021. Disponible en: https://www.redalyc.org/articulo.oa?id=301/30147485004

[5] Lütkepohl, H., *New introduction to multiple time series*, Springer, 2005 .

[6] Pastor, Alfredo.*La ciencia Humilde. Economía para ciudadanos.* Barcelona: Editorial Noema, 2008.

[7] Quinn, B., *Order determination for a multivariate autoregression*, Journal of the Royal Statistical Society, 1980.

[8] Ruiz-Ramírez, Juan, Hernández-Rodríguez, Gabriela Eréndira y Díaz Córdoba, Miriam de los Ángeles. (2014). *Determinación del tamaño mínimo de la serie de tiempo en el pronóstico del PIB trimestral en México*. Observatorio de la Economía Latinoamericana, Nº. 178. [Fecha de consulta: 11 de enero de 2020] en: http://www.eumed.net/cursecon/ecolat/mx/2013/pib-trimestral-mexico.html.

[9] Sánchez, J., *Relación entre los modelos de series de tiempo univariadas con los modelos de series de tiempo multivariadas*, Tesis de Licenciatura, 2013.

[10] Schwarz, G., *Estimating the dimension of a model*, Annals of Statistics, 1978.

[11] Sitio de página web: https://www.inegi.org.mx/temas/pib/, fecha de consulta: 28 de noviembre de 2019.

[12] Sitio de página web: https://docplayer.es/44600228-Econometria-ii-grado-en, -finanzas-y-contabilidad.html, fecha de consulta: 23 de marzo de 2020.

[13] Sitio de página web: https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/, fecha de consulta: 26 de marzo de 2020.