



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

INTRODUCCION A LA REGRESION DINAMICA

T E S I S
QUE PARA OBTENER EL TITULO DE
ACTUARIO
PRESENTA:
LETICIA HERNANDEZ BELMONT



DIRECTOR: M. EN P. MARIA DEL PILAR ALONSO REYES



BACUIS 2001 SECC

200976



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



ESTADOS UNIDOS MEXICANOS

MAT. MARGARITA ELVIRA CHÁVEZ CANO
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

Introducción a la Regresión Dinámica

realizado por Leticia Hernández Belmont

con número de cuenta 9135278-3 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario M. en A.P. María del Pilar Alonso Reyes

Propietario M. en C. José Antonio Flores Díaz

Propietario Act. Jaime Vázquez Alamilla

Suplente Act. Yazmín Iliana Barcenás Orozco

Suplente Act. María Guadalupe Tzintzún Cervantes

Consejo Departamental de MATEMÁTICAS

M. en C. JOSÉ ANTONIO FLORES DÍAZ

Agradecimientos

A Dios:

Por darme la oportunidad de sentirme viva cada día, con salud, alegría, acompañada de una hermosa familia, fieles amigos y gratos compañeros.

Pero sobre todo, quiero darle las gracias por estar siempre conmigo en mi corazón y en mi pensamiento

A mis padres:

Hortensia Belmont y Manuel Hernández

Por darme todo su amor, su total apoyo y comprensión Gracias por transmitirme su experiencia a través de sabios consejos, actitudes ejemplares y grandes acciones. Los amo no sólo por todo lo que me han dado, que es mucho, sino por su gran calidad humana.

Esto es por ustedes y para ustedes.

A mi hermana:

Laura Hernández Belmont

Porque juntas hemos compartido la felicidad que la vida nos ha otorgado, enfrentado todo tipo de situaciones, aprendido de nuestras experiencias, por tu valiosa amistad y tu contagiosa alegría.

A mi abuela:

Adelina Cozaya

Por ser el pilar de la familia, brindarnos su casa y estar siempre dispuesta a ayudarnos, a escucharnos y a darnos un buen consejo.

A mis tíos y primos:

Por ofrecer a mi familia todo el apoyo que les ha sido posible en cualquier tipo de circunstancia, por estar con nosotros en las buenas y en las malas, por ser ejemplo de sencillez, honradez y sinceridad.

A mis amigos y compañeros de la escuela y del trabajo:

Por su amistad, por los momentos agradables que hemos pasado, por lo aprendido a su lado, gracias, ya que con todo eso me han ayudado a crecer como persona y como profesionalista.

A mi profesora Pilar Alonso:

Por su gran colaboración en la realización de esta tesis al dedicar su tiempo y comprensión. Agradezco sus consejos, correcciones y el compartir conmigo sus conocimientos.

A todos mis profesores:

Por todas las contribuciones que han hecho a mi vida profesional. Por dedicar tiempo a su carrera y ser portadores de los conocimientos adquiridos.

A la Universidad Nacional Autónoma de México.

Por la gran oportunidad que me dio de estudiar una carrera en sus bellas instalaciones, con profesores excepcionales, destacados y aplicados en sus actividades. Espero poner tu nombre muy en alto.

Leticia Hernández Belmont

ÍNDICE

Introducción	I
Capítulo I. Series de Tiempo	
1. Introducción	1
2. Series temporales vistas como procesos estocásticos	3
2.1 Proceso estocástico	3
2.2 Procesos estacionarios	4
2.3 Proceso de ruido blanco	10
2.4 Procesos integrados	11
2.5 Operador de rezago y de diferencia	21
2.6 Descomposición de Wold	23
3. Ecuaciones en diferencia	26
3.1 Ecuaciones en diferencia de orden uno	27
3.2 Ecuaciones en diferencia de orden dos	27
3.3 Ecuaciones en diferencia de orden k	32
4. Modelos estacionarios	36
4.1 Modelos autorregresivos (AR)	36
4.1.1. Función de autocorrelación parcial ($FACP$)	47
4.2 Procesos de media móvil (MA)	55
4.3 Procesos $ARMA$	64
5. Modelos no estacionarios	73
5.1 Modelos $ARIMA$	73
5.2 Transformaciones estabilizadoras de varianza	75
5.3 Procesos $ARIMA$ estacionales	77
5.4 Modelo multiplicativo estacional	82

Capítulo II. Regresión Dinámica

1. Introducción	83
2. Formulación de modelos de regresión dinámica	84
2.1 Hipótesis básicas	85
2.2 Conceptos fundamentales	86
2.3 Modelización de la función de transferencia	90
2.4 Función de respuesta a impulsos	93
2.5 Modelización del Proceso de Inercia	96
2.6 Cointegración	97
3. Construcción de un modelo de regresión dinámica	99
4. Identificación del modelo	101
4.1 Identificación del modelo	104
4.2 Identificación de un modelo de regresión dinámica	102
4.3 Estimador más usual de la función de correlación cruzada	110
4.4 Identificación de la función de transferencia	113
4.5 Identificación del proceso de inercia	115
5. Estimación	
5.1 Procedimiento básicos para la estimación	116
5.2 Estimación lineal	125
5.3 Estimación de máxima verosimilitud exacta	129
5.4 Obtención de valores iniciales	132
5.5 Estimación de modelos de regresión dinámicos	136
6. Verificación	
6.1 Sobre los errores	141
6.2 Sobre los parámetros	149
7. Predicción	
7.1 Predicción a partir de un modelo $ARMA(p, q)$	154
7.2 Pronóstico con modelos $ARIMA$	159
7.3 Pronóstico utilizando modelos de regresión dinámica	161
7.4 Intervalos de confianza	172
7.5 Adaptación de pronósticos	174

Capítulo III. Aplicación de modelos de Regresión Dinámica

1. Introducción	177
2. Modelo univariante para la serie inflación en México	178
2.1 Identificación del modelo	178
2.2 Estimación y verificación del modelo univariante para la serie inflación	191
2.3 Uso del modelo inflación	202
3. Modelos de regresión dinámica	204
3.1 Análisis previo de la serie tasas de interés real	204
3.2 Preblanqueado de la serie dependiente tasas de interés real	206
3.3 Estimación del modelo y verificación de los supuestos	213
3.4 Pronóstico	220
Conclusiones	225
Bibliografía	229

Introducción

Una de las técnicas estadísticas de mayor uso para describir, en algún sentido, el comportamiento de una variable, es sin duda el análisis de regresión. Los modelos desarrollados mediante esta técnica consideran que la variable de interés, denominada variable dependiente o respuesta, puede ser aproximada a partir de un conjunto de variables, conocidas como independientes o explicativas, que afecten el comportamiento de la misma. El caso extremo es que exista una *relación funcional* entre las variables, es decir que a partir del conocimiento de las variables independientes se pueda conocer exactamente el comportamiento de la variable respuesta, esta relación puede expresarse como:

$$Y_i = f(X_{i_1}, X_{i_2}, \dots, X_{i_k})$$

en donde Y_i representa a la variable respuesta y X_{i_j} , $j = 1, 2, \dots, k$ a las variables independientes. Sin embargo, no todos los factores o causas que afectan el comportamiento de la variable dependiente son observables o conocidas para el investigador, por lo que más que una relación funcional entre las variables se tiene una *relación estadística* o *estocástica*, en donde el conocimiento de una variable independiente sólo aproxima el comportamiento de la variable respuesta, quedando:

$$Y_t = f(X_{t_1}, X_{t_2}, \dots, X_{t_k}) + N_t$$

el primer término de la derecha de ésta igualdad representa a todas las variables independientes incluidas en el modelo, mientras que el segundo engloba a todas aquellas variables que por diversas circunstancias han sido excluidas[®].

En los cursos básicos de estadística generalmente se estudia el análisis de regresión clásico, el cual se caracteriza por considerar sólo situaciones estáticas en el sentido de que tanto las variables independientes como la dependiente están observadas en el mismo instante de tiempo dando como resultado que un cambio en alguna de las primeras, X_i , $i = 1, 2, \dots, K$, se refleje de manera simultánea en la segunda, Y . No obstante, estos modelos se ven limitados cuando alguna o todas las variables independientes presentan un orden natural con respecto al tiempo, ya que su comportamiento puede afectar a la variable respuesta no sólo en el instante mismo sino de alguna de las siguientes tres maneras:

- a) después de haber transcurrido un cierto periodo de tiempo,
- b) por varios periodos de tiempo, aunque el impacto se deje notar simultáneamente,
- c) o ambas cosas a la vez.

Estas situaciones se conocen como *dinámicas*, y para ellas se han venido desarrollando otro tipo de modelos, denominados *dinámicos*, los cuales tratan de incluir todos los datos históricos observados que afectan significativamente a la variable respuesta, aprovechando así toda la información posible que ayude a cumplir con los objetivos para los cuales se desarrolla un modelo, entre los cuales se encuentran: descripción de la variable respuesta, estimación de los parámetros, predicción y control.

Un ejemplo simple de un modelo dinámico es el siguiente:

$$Y_t = v X_{t-4} + a_t$$

en este caso el efecto de un cambio en X sólo se refleja en Y después de haber transcurrido cuatro periodos de tiempo, y el término N_t se reduce a un término de perturbación aleatoria, utilizado generalmente en modelos de regresión clásicos.

El ejemplo anterior no capta de ninguna manera la esencia de las situaciones dinámicas con las que se puede uno enfrentar, ya que la función que relaciona a las variables explicativas con la respuesta puede llegar a ser muy complicada o incluso desconocida, además de que el término a_t puede tener una estructura más complicada que la utilizada en modelos clásicos. Una manera un poco más general de ejemplificar un modelo dinámico es:

$$Y_t = v_0 X_{t-b} + v_1 X_{t-b-1} + v_2 X_{t-b-2} + \dots + a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \dots$$

en donde se tiene una única serie explicativa $\{X_t\}$ cuyo comportamiento afecta a la serie respuesta $\{Y_t\}$ mediante una función lineal y después de transcurridos b periodos de tiempo, prolongándose indefinidamente a partir de ese momento. En cuanto a los términos de perturbación, estos se encuentran autocorrelacionados, afectando a la serie respuesta también en forma lineal.

Como puede observarse el tipo de relaciones entre variables que abarcan estos modelos es mucho mayor que los modelos de regresión clásicos y es por lo cual representan una técnica de gran ayuda en el análisis de datos.

El objetivo del presente trabajo es despertar el interés del lector en la aplicación de los modelos de regresión dinámica a través de una introducción en donde se explica, de la

** El análisis de regresión está interesado en relaciones únicamente estadísticas o estocásticas por lo cual, a partir de este momento, son a las que se harán referencia*

manera más sencilla posible. los conceptos y procedimientos básicos para el desarrollo de un modelo de este tipo.

Es importante que el lector ya haya tenido contacto con algunos conceptos básicos de estadística descriptiva e inferencial, así como con la construcción de modelos de regresión clásica, esto último porque como posteriormente podrá darse cuenta una parte importante de lo estudiado en regresión clásica es aplicable —o puede extenderse con pequeños retoques— al caso dinámico. Además que no en todos los casos un modelo dinámico es la mejor opción, a pesar de que las variables sean dinámicas, ya que en algunas situaciones las ventajas que ofrece este modelo sobre un modelo de regresión clásica son mínimas. Por lo que resulta importante que el lector conozca ambos métodos y pueda decidir con facilidad el que mejor le convenga.

Para la construcción de un modelo es importante establecer y seguir una estrategia, la cual consiste básicamente de las siguientes cuatro etapas sugeridas por Box y Jenkins²:

1. **Identificación del modelo** (se busca el mejor modelo; el cual no sólo debe de ser adecuado para representar el comportamiento de los datos observados, sino que su elección debe ser sugerida por los mismos).
2. **Estimación de los parámetros implícitos en el modelo** (para lo cual se debe usar la técnica más eficiente que se conozca).
3. **Verificación de los supuestos** (para que los resultados puedan considerarse válidos)
4. **Uso del modelo** (para los fines que fue construido).

Para la etapa de estimación de los parámetros se han venido desarrollando varias técnicas sin embargo, todas ellas conllevan al mismo objetivo: simplificar la estructura del modelo para que el número de parámetros a estimar sea mínimo. La técnica que aquí se analizará esta basada en los modelos que se han venido desarrollando para el análisis de series de

² Introducidas por Box, G.E.P. y Jenkins, G. en su libro clásico *Time Series Analysis. Forecasting and Control* el cual apareció por primera vez en 1970.

tiempo con el enfoque sugerido por Box y Jenkins; la cual consiste en asumir que la trayectoria que sigue la variable independiente, X_t , puede ser modelada por una función, denominada *función de transferencia*, la cual tiene mucha similitud con un modelo *ARMA* para una serie de tiempo univariante, mientras que el término de perturbancia, N_t , puede ser representado por un modelo *ARIMA*, también modelo desarrollado para el análisis de series temporales.

Debido a que no todos los lectores están familiarizados con el análisis de series de tiempo y ya que sus conceptos se usarán ampliamente para el desarrollo de los modelos dinámicos se ha dedicado el primer capítulo a explicarlos brevemente; posteriormente en el segundo capítulo se explicará el desarrollo de la metodología a seguir para el análisis de regresión dinámica; y finalmente, en el capítulo tres se ilustran todos los conceptos con un ejemplo práctico.

Capítulo 1

Series de Tiempo

1. Introducción

Una *serie temporal o cronológica* es una secuencia de observaciones ordenada en el tiempo (en algunos casos el orden puede ser respecto a otra dimensión). Ejemplos de *series temporales* se encuentran en casi todas las áreas. En la vida diaria, se puede estar interesado en la variación de precios de los productos que se consumen, en las tasas mensuales de rendimiento que ofrecen los bancos o en los índices diarios de contaminación. En economía, ejemplos de series temporales son: las tasas de inflación, el Índice de Precios al

Consumidor, el Producto Nacional Bruto y otras
demás características económicas que se analizan. En demografía, se da seguimiento a la población, entre ellos el número de nacimientos y muertes. En medicina, un electrocardiograma puede ser considerado una serie de tiempo. Las utilidades anuales, los gastos en el mercado son variables de interés. Muchos fenómenos que se repiten a través del tiempo son de interés. Algunos ejemplos, de la gran variedad de series de tiempo que existen (fig. 1.1).

Kodak *ds*
digital science™

DEMO

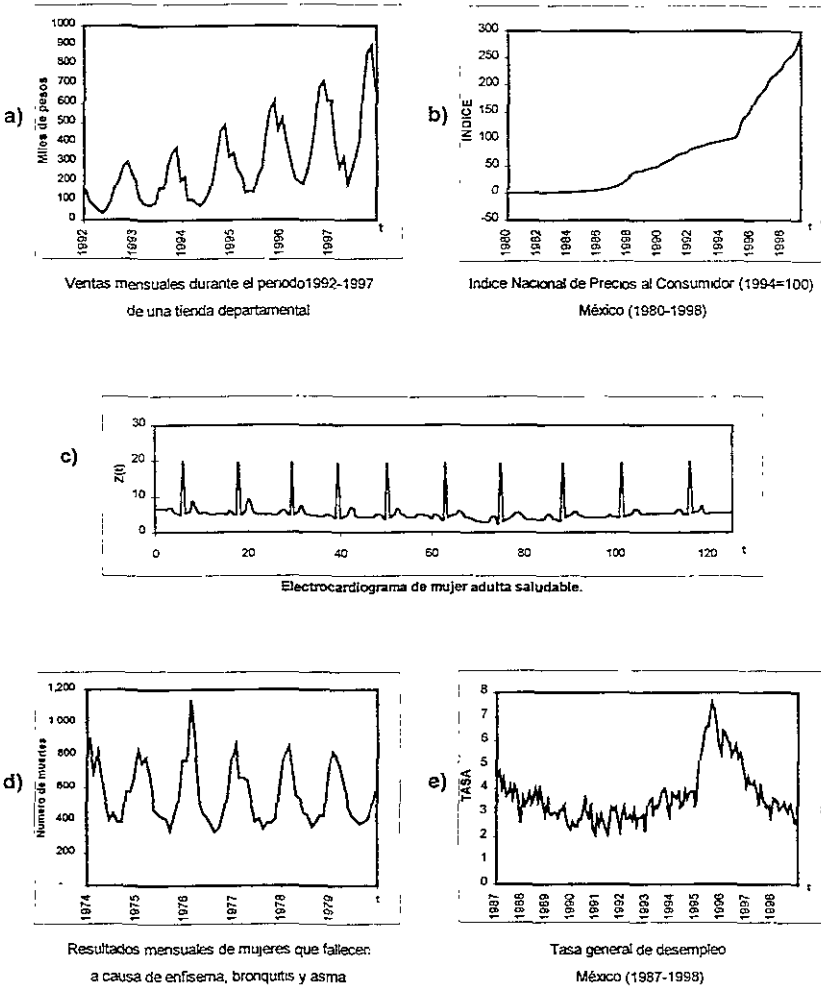


Figura 1.1

Hay varios objetivos para realizar un estudio de series temporales, el principal puede ser la *comprensión* del mecanismo que generó la serie, o la simple *descripción* de sus características sobresalientes a través de gráficas y medidas descriptivas, o quizá, la *predicción* de valores futuros que ayudarán en la planeación y toma de decisiones

(suponiendo que las condiciones futuras serán análogas a las pasadas), o tal vez, lo que se desea es *inferir* sobre el mecanismo probabilístico que generó la serie.

La importancia del análisis de series de tiempo radica en el hecho de que utiliza como única información, para la construcción del modelo, la propia historia de la serie; la cual, puede ser *continua* o *discreta* dependiendo si las observaciones están hechas continuamente en el tiempo (*series de tiempo continuas*) o en intervalos específicos de tiempo (*series de tiempo discretas*). Además, que estos modelos reconocen el orden en el cual se hicieron las observaciones, ya que generalmente las observaciones sucesivas no son independientes.

2. Series temporales vistas como procesos estocásticos

2.1 Proceso estocástico

A partir de este momento se considerará sólo series de tiempo discretas con la característica adicional de que las observaciones están tomadas en instantes de tiempo igualmente espaciados (segundos, minutos, horas, días,..., etc.). Además, se supondrá que en cada instante de tiempo t hay una variable Z_t que toma al azar un sólo valor de entre todos los resultados posibles del fenómeno (estos valores, son los que constituyen la serie de tiempo).

A esta colección de variables aleatorias $\{Z_t\}$ ordenadas en el tiempo se le conoce como *proceso estocástico* y la serie observada se considera una *realización del proceso*.

Por ejemplo, en la *figura 1.2a*) se observan las tasas de inflación reales para el período (1950-1995). No obstante, estas tasas pudieron haber tenido otro comportamiento distinto como se muestra en la *figura 1.2b*) en donde se graficaron tres posibles realizaciones distintas del proceso además de la real.

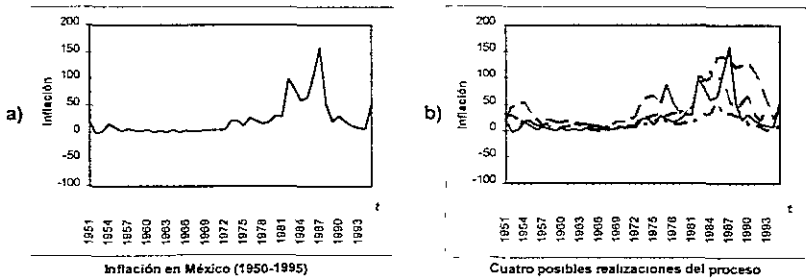


Figura 1.2

Para un proceso $\{Z_t\}$ dado; su función media está definida por

$$\mu_t = E(Z_t),$$

su función varianza

$$\sigma_t^2 = Var(Z_t) = E[(Z_t - \mu_t)^2],$$

y su autocovarianza^o entre los instantes de tiempo t_1 y t_2

$$\gamma(t_1, t_2) = Cov(Z_{t_1}, Z_{t_2}) = E[(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})].$$

2.2 Procesos estacionarios

Una manera de describir a un proceso estocástico es especificar a la función de densidad conjunta de todas las variables aleatorias que lo integran, pero en la práctica esto es muy complicado. Si tan sólo se tuvieran varias realizaciones del proceso se podría inferir sobre

^o Se llama autocovarianza a la covarianza entre dos observaciones de la misma serie.

las variables aleatorias pero cuando se dice, que la tasa de inflación al final de 1994 es una variable aleatoria, se tendrá una única observación sobre esta variable y no hay manera alguna de obtener otra. Esto, aunado a la dependencia que hay entre las variables obligan a definir algunas propiedades, como la *estacionariedad*, para simplificar el modelo.

Un proceso estocástico (serie de tiempo) se dice que es *estrictamente estacionario* si la distribución conjunta de cualquier conjunto de n variables correspondientes a los instantes de tiempo t_1, t_2, \dots, t_n es la misma que la de cualquier otro conjunto de n variables correspondientes a los instantes $t_{1+k}, t_{2+k}, \dots, t_{n+k}$.

Ya que n puede tomar cualquier valor, cuando $n=1$, la distribución de Z_t no depende de t , y no sólo la media y la varianza son constantes sino todos los momentos de orden superior.

Cuando $n=2$, se tiene una distribución conjunta de Z_{t_1} y Z_{t_2} igual a la de Z_{t_1+k} y Z_{t_2+k} con función de autocovarianza

$$\gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k)$$

al escribir $t_1 + k = t$ y $t_2 = t$,

$$\gamma_{-k} = \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k$$

la función de autocovarianzas sólo depende de $(t_2 - t_1) = k$, que se conoce como *rezago*.

En la práctica los fenómenos que se ajustan a este tipo de modelo son escasos por lo cual, es necesario definir la estacionariedad de la serie de una manera menos estricta.

Se dice que una serie de tiempo es *débilmente estacionaria* si:

- i) $E(Z_t) = \mu$
- ii) $Var(Z_t) = \sigma^2$
- iii) $\gamma_{-k} = \gamma(t, t-k) = \gamma(t, t+k) = \gamma_k$

donde, μ y σ^2 son constantes y la función autocovarianza sólo depende del rezago.

La serie queda definida en términos de su primer y segundo momentos y no se harán suposiciones para los momentos de orden superior. Si además se supone que $Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}$ tienen una distribución conjunta normal multivariada, la estacionariedad débil coincide con la estricta, debido a que la normal multivariada está completamente caracterizada por los momentos primero y segundo. Por lo tanto bastará conocer la media μ y la función de autocovarianza γ_k para describir completamente a la serie estacionaria.

No obstante, para evitar la influencia de las unidades de medida, es preferible trabajar con las autocorrelaciones de la serie, definidas a través de

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\sigma_t^2}$$

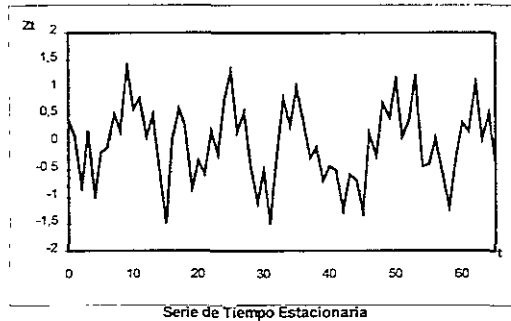


Figura 1.3

La varianza de Z_t en términos del rezago es igual a γ_0 , al igual que la varianza de Z_{t+k} .
Es así como, el coeficiente de correlación en el retardo k es

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

el cual, para valores diferentes de k , genera la función de autocorrelación (FAC).

Esta función, (FAC), tiene las siguientes propiedades

i) $\rho_0 = 1$

Prueba:

De la definición de coeficiente de correlación se sigue:

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1.$$

ii) $|\rho_k| \leq 1$

Prueba:

Para probar que el coeficiente de correlación en valor absoluto, entre las variables Z_t y Z_{t+k} , es siempre menor o igual a uno se hará uso de la varianza de la suma de estas dos variables estandarizadas y de su diferencia. La estandarización de variables consiste en restarle a cada una de ellas su media y dividir las entre su desviación estándar, debido a que estas dos variables forman parte de un mismo proceso estacionario ambas tendrán la misma media y la misma varianza. De esta manera la varianza de la que se partirá para la demostración es la siguiente:

$$\text{Var}\left(\frac{Z_t - \mu}{\sqrt{\gamma_0}} + \frac{Z_{t+k} - \mu}{\sqrt{\gamma_0}}\right) \quad (2.2.1)$$

nótese que se está utilizando la notación de rezago, por lo cual en lugar de utilizar σ^2 se utiliza γ_0 .

Desarrollando (2.2.1) de acuerdo a la definición de la varianza se tiene

$$\begin{aligned} \text{Var}\left(\frac{Z_t - \mu}{\sqrt{\gamma_0}} + \frac{Z_{t+k} - \mu}{\sqrt{\gamma_0}}\right) &= \\ &= \frac{1}{\gamma_0} [\text{Var}(Z_t - \mu) + \text{Var}(Z_{t+k} - \mu) + 2\text{Cov}(Z_t - \mu, Z_{t+k} - \mu)] \end{aligned} \quad (2.2.2)$$

utilizando las propiedades de la función varianza se tiene que $\text{Var}(Z_t - \mu) = \text{Var}(Z_t) + \text{Var}(\mu) = \text{Var}(Z_t)$, de igual manera que $\text{Var}(Z_{t+k} - \mu) = \text{Var}(Z_{t+k})$.

Ahora desarróllese la covarianza ente $Z_t - \mu$ y $Z_{t+k} - \mu$:

$$\text{Cov}[Z_t - \mu, Z_{t+k} - \mu] = E[(Z_t - \mu)(Z_{t+k} - \mu)] = \text{Cov}(Z_t, Z_{t+k}) \quad (2.2.3)$$

Es así como (2.2.2) se reduce a:

$$\begin{aligned} &\frac{1}{\gamma_0} [\text{Var}(Z_t) + \text{Var}(Z_{t+k}) + 2\text{Cov}(Z_t, Z_{t+k})] = \\ &= \frac{1}{\gamma_0} [\gamma_0 + \gamma_0 + 2\gamma_k] = 1 + 1 + 2\rho_k = 2(1 + \rho_k) \end{aligned}$$

Debido a que la varianza es una función no negativa, entonces.

$$2(1 + \rho_k) \geq 0$$

de donde se sigue que

$$\rho_k \geq -1$$

Considerando ahora, a la varianza entre la diferencia de las variables estandarizadas se obtendrá la otra parte de la desigualdad, es decir $\rho_k \leq 1$,

$$\begin{aligned} \text{Var}\left(\frac{Z_t - \mu}{\sqrt{\gamma_0}} - \frac{Z_{t+k} - \mu}{\sqrt{\gamma_0}}\right) &= \\ &= \frac{1}{\gamma_0} [\text{Var}(Z_t - \mu) + \text{Var}(Z_{t+k} - \mu) - 2\text{Cov}(Z_t - \mu, Z_{t+k} - \mu)] = \\ &= \frac{1}{\gamma_0} [\text{Var}(Z_t) + \text{Var}(Z_{t+k}) - 2\text{Cov}(Z_t, Z_{t+k})] = \\ &= \frac{1}{\gamma_0} [\gamma_0 + \gamma_0 - 2\gamma_k] = 1 + 1 - 2\rho_k = 2(1 - \rho_k) \end{aligned}$$

Recordando de nuevo que la varianza nunca es negativa, se tiene

$$\rho_k \leq 1.$$

- iii) $\rho_k = \rho_{-k}$ para todo k , es decir, la función autocorrelación es una función par y simétrica sobre $k = 0$.

Prueba.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_{-k}}{\gamma_0} = \rho_{-k}$$

Un diagrama de ρ_k contra k se llama *correlograma* y es muy usual graficarlo sólo para valores de k no negativos. En la figura 1.4 se presenta un correlograma en donde la dependencia entre observaciones tiende a cero al aumentar el retardo.

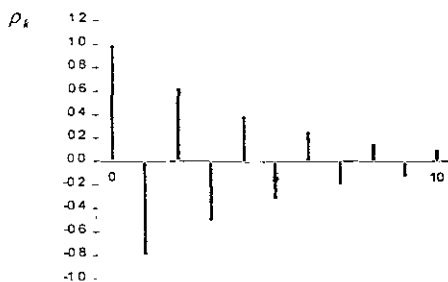


Figura 1.4

2.3 Proceso de Ruido Blanco

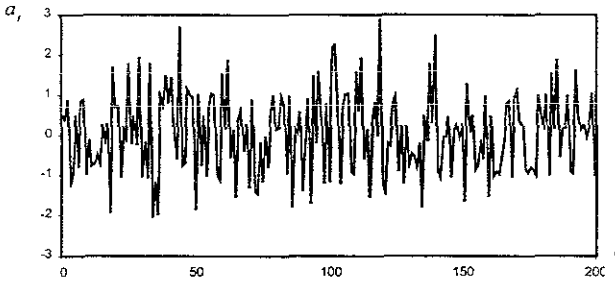
Se llama proceso de ruido blanco a la secuencia de variables aleatorias, usualmente denotada por a_t , con media constante μ_a , casi siempre igual a cero; varianza constante σ_a^2 ; y función de autocovarianza $\gamma_k = 0$ para todo $k \neq 0$. Este proceso, por definición estacionario, tiene función de autocovarianza

$$\gamma_k = \begin{cases} 0 & \text{si } k \neq 0 \\ \sigma_a^2 & \text{si } k = 0 \end{cases}$$

y función de autocorrelación

$$\rho_k = \begin{cases} 0 & \text{si } k \neq 0 \\ 1 & \text{si } k = 0 \end{cases}$$

Los procesos de este tipo son útiles en muchos casos, particularmente cuando se utilizan como componentes de procesos más complicados. En la *figura 1.5* se muestra una realización de un proceso de ruido blanco en donde las observaciones están normalmente distribuidas.



Una realización de un proceso de Ruido Blanco

Figura 1.5

2.4 Procesos integrados

El análisis de series temporales no puede estar limitado a series de tiempo estacionarias, por el hecho de que gran parte de las series que se encuentran en la práctica son no-estacionarias. Sin embargo, existen formas para convertir en estacionarias las series que no lo son. En muchos casos la sola diferenciación del proceso no-estacionario $\{Z_t\}$, da como resultado un nuevo proceso $\{W_t\}$ que si cumple con las características de estacionariedad, es decir

$$W_t = Z_t - Z_{t-1}$$

si $\{W_t\}$ todavía no es estacionario se repite la operación una vez más, obteniendo un nuevo proceso $\{Y_t\}$, donde

$$Y_t = W_t - W_{t-1}$$

de esta manera, se puede seguir diferenciando hasta encontrar un proceso estacionario, que se denominará *proceso integrado de orden h* por ser el resultado de h diferenciaciones.

Para expresar este tipo de relaciones se utiliza el operador ∇ , el cual está definido como:

$$\nabla Z_t = Z_t - Z_{t-1}$$

y por lo tanto se tiene que W_t puede escribirse como $W_t = \nabla Z_t$ y Y_t como $Y_t = \nabla \nabla Z_t = \nabla^2 Z_t$.

La aplicación de este tipo de procesos es de gran utilidad cuando existe alguna tendencia polinomial que describe a la serie. Considérese como ejemplo al siguiente proceso Z_t , que sigue una tendencia polinomial de orden uno

$$Z_t = a + bt + u_t$$

más un proceso estacionario $\{u_t\}$ (generalmente este proceso es de ruido blanco). Al aplicar el operador diferencia una vez se tiene

$$\nabla Z_t = Z_t - Z_{t-1} = a + bt + u_t - [a + b(t-1) + u_{t-1}] = b + u_t - u_{t-1} = b + \nabla u_t$$

para demostrar que éste nuevo proceso es estacionario obsérvese que ∇Z_t está compuesto por una constante b más la diferencia de dos procesos estacionarios, por lo tanto, bastará demostrar que la diferencia de procesos estacionarios es estacionaria.

Sean u_t y u_{t-1} dos procesos estacionarios, con media cero y varianza σ^2 , y considérese a la diferencia $\nabla u_t = u_t - u_{t-1}$.

Al aplicar la función esperanza y la función varianza, se tiene que

$$E(\nabla u_t) = E(u_t - u_{t-1}) = E(u_t) - E(u_{t-1}) = 0$$

$$Var(\nabla u_t) = Var(u_t - u_{t-1}) = Var(u_t) + Var(u_{t-1}) - 2Cov(u_t, u_{t-1}) = 2\sigma^2 - 2\gamma_1 = 2(\sigma^2 - \gamma_1)$$

la esperanza siempre será cero y la varianza constante. Además que la autocovarianza entre ∇u_t y ∇u_{t+k} sólo dependerá del retardo

$$\begin{aligned} Cov(\nabla u_t, \nabla u_{t+k}) &= E[(u_t - u_{t-1})(u_{t+k} - u_{t+k-1})] = E[u_t u_{t+k} - u_t u_{t+k-1} - u_{t-1} u_{t+k} + u_{t-1} u_{t+k-1}] = \\ &= E(u_t u_{t+k}) - E(u_t u_{t+k-1}) - E(u_{t-1} u_{t+k}) + E(u_{t-1} u_{t+k-1}) = 2\gamma_k - \gamma_{k+1} - \gamma_{k-1} \end{aligned}$$

Con esto queda demostrado que la tendencia polinomial de orden uno puede ser eliminada mediante la aplicación del operador ∇ . A este tipo de no-estacionariedad se le denomina no-estacionariedad homogénea.

En general un proceso que sigue una tendencia polinomial de orden h

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_h t^h + u_t = \sum_{i=0}^h \beta_i t^i + u_t$$

donde $\{u_t\}$ es un proceso estacionario, será un proceso integrado de orden h .

Para demostrar lo anterior serán necesarios los siguientes resultados:

Teorema 1 (teorema de Pascal). Sean n, m números enteros y no negativos; entonces:

$$\binom{n}{m} = \binom{n-1}{m} + \binom{n-1}{m-1}$$

en donde $\binom{n}{m} = \frac{n!}{(n-m)!m!}$.

Teorema 2 (teorema del binomio). Sean a y b números reales y n entero, $n \geq 0$; entonces:

$$(a-b)^n = \sum_{i=0}^n \binom{n}{i} (-1)^i a^{n-i} b^i$$

Primero, se demostrará que por cada vez que se aplica el operador ∇ , se disminuye un grado la tendencia polinomial del proceso Z_t . Para ello se aplica una vez el operador ∇ al proceso:

$$\begin{aligned} \nabla Z_t &= \sum_{i=0}^h \beta_i t^i + u_t - \sum_{i=0}^h \beta_i (t-1)^i - u_{t-1} \\ &= \sum_{i=0}^h \beta_i [t^i - (t-1)^i] + \nabla u_t = \sum_{i=0}^h \beta_i \left[t^i - \sum_{j=0}^i \binom{i}{j} (-1)^j t^{i-j} \right] + \nabla u_t = \\ &= \sum_{i=0}^h \beta_i \left[t^i - t^i - \sum_{j=1}^i \binom{i}{j} (-1)^j t^{i-j} \right] + \nabla u_t = \sum_{i=0}^h \beta_i \left[- \sum_{j=1}^i \binom{i}{j} (-1)^j t^{i-j} \right] + \nabla u_t \end{aligned}$$

utilizando la definición de $\binom{i}{j}$, se puede observar la siguiente propiedad:

$$\binom{i}{j} = \frac{i!}{(i-j)!j!} = \frac{i \cdot (i-1)!}{j \cdot (i-j)!(j-1)!} = \frac{i}{j} \binom{i-1}{j-1}$$

de esta manera,

$$\sum_{i=0}^h \beta_i \left[-\sum_{j=1}^i \binom{i}{j} (-1)^j t^{i-j} \right] + \nabla u_t = \sum_{i=0}^h \beta_i \left[-\sum_{j=1}^i \frac{i}{j} \binom{i-1}{j-1} (-1)^j t^{i-j} \right] + \nabla u_t$$

Cuando $i=0$, el término se anula, por lo tanto la expresión anterior se reduce a:

$$\begin{aligned} &= \sum_{i=1}^h \beta_i \left[(-1) \sum_{j=1}^i \frac{i}{j} \binom{i-1}{j-1} (-1)^j t^{i-j} \right] + \nabla u_t = \sum_{i=0}^{h-1} \beta_{i+1} \left[\sum_{j=0}^{i-1} \frac{i+1}{j+1} \binom{i}{j} (-1)^j t^{i-j} \right] + \nabla u_t = \\ &= \sum_{i=0}^{h-1} \beta_{i+1} \left[\sum_{j=0}^{i-1} \binom{i+1}{j+1} (-1)^j t^{i-j} \right] + \nabla u_t = \sum_{i=0}^{h-1} \sum_{j=0}^{i-1} \binom{i+1}{j+1} \beta_{i+1} (-1)^j t^{i-j} + \nabla u_t = \end{aligned}$$

a partir de aquí se puede deducir un proceso con tendencia polinomial de un orden menor, de la siguiente manera:

$$\sum_{i=0}^{h-1} \sum_{j=0}^{i-1} \binom{i+1}{j+1} \beta_{i+1} (-1)^j t^{i-j} + \nabla u_t = \sum_{k=0}^{h-1} \beta_k^{(1)} t^k + \nabla u_t \quad (2.4.1)$$

en donde $\beta_k^{(1)}$ representan a los coeficientes del nuevo polinomio. El superíndice (1) denota que son el resultado de la primera diferenciación y se obtienen de igualar potencias en (2.4.1); por ejemplo, el coeficiente, correspondiente al término t^0 , se obtiene cuando $i=j$ para $i=0,1,\dots,h-1$, es decir,

$$\beta_0^{(1)} = \beta_1 - \beta_2 + \beta_3 - \dots + (-1)^{h-1} \beta_h$$

el coeficiente correspondiente a t^i , se obtendrá de igual manera, sólo que ahora la condición será: $i = j + 1$ con $i = 1, 2, \dots, h - 1$

$$\beta_1^{(1)} = 2\beta_2 - 3\beta_3 + 4\beta_4 - \dots + (-1)^h h\beta_h$$

y así sucesivamente hasta llegar al coeficiente correspondiente a t^{h-1} obtenido cuando $i = j + h - 1$ con $i = h - 1$

$$\beta_{h-1}^{(1)} = h\beta_h$$

Si se aplica de nuevo el operador diferencia, ∇ , se tiene

$$\begin{aligned} \nabla^2 Z_t &= \sum_{k=0}^{h-1} \beta_k^{(1)} t^k + \nabla u_t - \sum_{k=0}^{h-1} \beta_k^{(1)} (t-1)^k - \nabla u_{t-1} = \sum_{k=0}^{h-1} \beta_k^{(1)} [t^k - (t-1)^k] + \nabla u_t - \nabla u_{t-1} = \\ &= \sum_{k=0}^{h-1} \beta_k^{(1)} \left[t^k - \sum_{j=0}^k \binom{k}{j} (-1)^j t^{k-j} \right] + \nabla^2 u_t = \sum_{k=0}^{h-1} \beta_k^{(1)} \left[t^k - t^k - \sum_{j=1}^k \binom{k}{j} (-1)^j t^{k-j} \right] + \nabla^2 u_t = \\ &= \sum_{k=0}^{h-1} \beta_k^{(1)} \sum_{j=1}^k \binom{k}{j} (-1)^j t^{k-j} + \nabla^2 u_t = \sum_{k=0}^{h-1} \beta_k^{(1)} \sum_{j=1}^k \frac{k}{j} \binom{k-1}{j-1} (-1)^j t^{k-j} + \nabla^2 u_t = \\ &= \sum_{k=1}^{h-1} \sum_{j=1}^k \frac{k}{j} \binom{k-1}{j-1} (-1)^j t^{k-j} = \sum_{k=1}^{h-1} \sum_{j=1}^k \binom{k}{j} \beta_k^{(1)} (-1)^j t^{k-j} + \nabla^2 u_t = \\ &= \sum_{k=0}^{h-2} \sum_{j=0}^{k-1} \binom{k+1}{j+1} \beta_k^{(1)} (-1)^j t^{k-j} + \nabla^2 u_t \end{aligned}$$

al igualar las potencias con el siguiente polinomio de orden $h-2$ se obtienen los coeficientes $\beta_l^{(2)}$ para $l = 0, 1, \dots, h - 2$

$$\sum_{t=0}^{h-2} \beta_t^{(2)} t^t + \nabla^2 u_t$$

con:

$$\beta_0^{(2)} = \beta_1^{(1)} - \beta_{21}^{(1)} + \beta_3^{(1)} - \dots + (-1)^{h-2} \beta_{h-1}^{(1)}$$

$$\beta_1^{(2)} = 2\beta_{21}^{(1)} - 3\beta_{31}^{(1)} + 4\beta_4^{(1)} - \dots + (-1)^{h-1} (h-1) \beta_{h-1}^{(1)}$$

⋮

$$\beta_{h-2}^{(2)} = (h-1) \beta_{h-1}^{(1)}$$

supóngase que al aplicar el operador diferencia $h-1$ veces, se obtiene:

$$\nabla^{h-1} Z_t = \sum_{s=0}^1 \beta_s^{(h-1)} t^s + \nabla^{h-1} u_t$$

como puede observarse este último proceso todavía tiene una tendencia polinomial de orden uno, por lo cual será necesario una diferenciación más

$$\begin{aligned} \nabla \nabla^{h-1} Z_t &= \nabla^h Z_t = \sum_{s=0}^1 \beta_s^{(h-1)} t^s + \nabla^{h-1} u_t - \sum_{s=0}^1 \beta_s^{(h-1)} (t-1)^s + \nabla^{h-1} u_{t-1} = \\ &= \sum_{s=0}^1 \beta_s^{(h-1)} [t^s - (t-1)^s] + \nabla^{h-1} (u_t - u_{t-1}) = \beta_1^{(h-1)} [t - (t-1)] + \nabla^{h-1} \nabla u_t = \\ &= \beta_1^{(h-1)} + \nabla^h u_t = \beta_0^{(h)} + \nabla^h u_t \end{aligned}$$

con:

$$\begin{aligned}
 \beta_0^{(h)} &= \beta_1^{(h-1)} \\
 &= 2\beta_2^{(h-2)} \\
 &= 2 \cdot 3 \cdot \beta_3^{(h-3)} \\
 &\quad \vdots \\
 &= 2 \cdot 3 \cdot \dots \cdot (h-1)\beta_{h-1}^{(1)} \\
 &= 2 \cdot 3 \cdot \dots \cdot (h-1)h\beta_h = h!\beta_h
 \end{aligned}$$

Es así como:

$$\nabla^h Z_t = (h)!d + \sum_{j=0}^h (-1)^j u_{t-j} = (h)!d + \nabla^h u_t$$

La tendencia polinomial de orden h ha sido eliminada mediante la aplicación sucesiva de h veces el operador ∇ . Ahora se debe demostrar que efectivamente $\nabla^h Z_t$ es un proceso estacionario, es decir que $\nabla^h u_t$ cumple con las características de estacionariedad ya que básicamente $\{\nabla^h Z_t\}$ está integrado por la diferencia de procesos estacionarios $\{\nabla^h u_t\}$.

Se comprueba primero, por inducción que el proceso integrado de orden h , $\nabla^h Z_t$, puede ser factorizado como:

$$\nabla^h Z_t = \sum_{i=0}^h \binom{h}{i} (-1)^i Z_{t-i} \tag{2.4.2}$$

Para ello aplíquese una vez el operador diferencia al proceso Z_t , obteniéndose:

$$\nabla Z_t = Z_t - Z_{t-1} \tag{2.4.3}$$

el cual, puede ser reescrito como:

$$\nabla Z_t = \sum_{i=0}^1 \binom{1}{i} (-1)^i Z_{t-i}$$

Para ver la factorización de manera más clara, aplíquese a (2.4.3) una vez más el operador ∇ ,

$$\nabla \nabla Z_t = \nabla^2 Z_t = \nabla(Z_t - Z_{t-1}) = Z_t - Z_{t-1} - (Z_{t-1} - Z_{t-2}) = Z_t - 2Z_{t-1} + Z_{t-2}$$

que corresponde a:

$$\nabla^2 Z_t = \sum_{i=0}^2 \binom{2}{i} (-1)^i Z_{t-i}$$

supóngase que la factorización es válida para $h-1$, es decir

$$\nabla^{h-1} Z_t = \sum_{i=0}^{h-1} \binom{h-1}{i} (-1)^i Z_{t-i} \quad (2.4.4)$$

Ahora, para demostrar que (2.4.2) es válido para h aplíquese ∇ a (2.4.4)

$$\begin{aligned} \nabla \nabla^{h-1} Z_t &= \nabla^h Z_t = \nabla \left(\sum_{i=0}^{h-1} \binom{h-1}{i} (-1)^i Z_{t-i} \right) = \sum_{i=0}^{h-1} \binom{h-1}{i} (-1)^i Z_{t-i} - \sum_{i=0}^{h-1} \binom{h-1}{i} (-1)^i Z_{t-i-1} = \\ &= \sum_{i=0}^{h-1} \binom{h-1}{i} (-1)^i Z_{t-i} + \sum_{i=1}^h \binom{h-1}{i-1} (-1)^i Z_{t-i} = \sum_{i=1}^{h-1} \left[\binom{h-1}{i} + \binom{h-1}{i-1} \right] (-1)^i Z_{t-i} + Z_t + (-1)^h Z_{t-h} \end{aligned}$$

utilizando el teorema de Pascal y ya que $\binom{h}{0} = 1$ y $\binom{h}{h} = 1$ la expresión anterior se reduce

a:

$$\sum_{i=0}^h \binom{h}{i} (-1)^i Z_{t-i},$$

por lo tanto (2.4.2) es válida para cualquier entero no negativo.

Entonces, al diferenciar h veces el proceso estacionario u_t , se tiene

$$\nabla^h u_t = \sum_{i=0}^h \binom{h}{i} (-1)^i u_{t-i},$$

aplicar ahora el operador esperanza, para demostrar que la media es constante

$$E(\nabla^h u_t) = \sum_{i=0}^h \binom{h}{i} (-1)^i E(u_{t-i}) = \sum_{i=0}^h \binom{h}{i} (-1)^i \mu = \mu \sum_{i=0}^h \binom{h}{i} (-1)^i = \mu(1-1)^h = 0$$

de igual manera calcúlese la autocovarianza del proceso

$$\begin{aligned} \text{Cov}(\nabla^h u_t, \nabla^h u_{t+k}) &= E[\nabla^h u_t \nabla^h u_{t+k}] = E\left\{ \left[\sum_{i=0}^h \binom{h}{i} (-1)^i u_{t-i} \right] \left[\sum_{j=0}^h \binom{h}{j} (-1)^j u_{t+k-j} \right] \right\} = \\ &= \sum_{i=0}^h \sum_{j=0}^h \binom{h}{i} \binom{h}{j} (-1)^{i+j} E(u_{t-i} u_{t+k-j}) = \sum_{i=0}^h \sum_{j=0}^h \binom{h}{i} \binom{h}{j} (-1)^{i+j} \gamma_{i-j+k} \end{aligned}$$

la cual sólo depende del rezago, en particular cuando $k=0$ se obtiene la

$$\text{Var}(\nabla^h u_t) = \sum_{i=0}^h \sum_{j=0}^h \binom{h}{i} \binom{h}{j} (-1)^{i+j} \gamma_{i-j}, \text{ que es constante, ya que no depende ni de } t, \text{ ni de } k.$$

Con esto queda demostrado, que la h -ésima diferencia de un proceso que sigue una tendencia polinomial de orden h es estacionaria, además que la estacionariedad de un proceso se conserva a pesar del uso del operador ∇ .

Puede seguir aplicándose el operador diferencia ∇ a un proceso que ya es estacionario, y este seguirá siéndolo, sin embargo, el hecho de sobrediferenciar ocasiona la pérdida de observaciones innecesariamente, ya que cada vez que se aplica el operador ∇ se pierde una observación.

2.5 Operador de rezago y de diferencia.

El *operador diferencia* ∇ , introducido ya en la sección anterior, sirve para denotar el grado de diferenciación o integración al que fue sometido el proceso original para que cumpliera con las características de estacionariedad. Este operador relaciona al proceso original Z_t , con el obtenido después de diferenciarlo k veces de la siguiente manera:

$$Z_t = \nabla^k Z_t$$

Otro operador, denominado *operador de rezago*, y denotado como B (del ingles Backward) está definido mediante la relación

$$BZ_t = Z_{t-1}$$

por la aplicación sucesiva de este operador

$$BZ_t = Z_{t-1}$$

$$B^2 Z_t = B(Z_{t-1}) = Z_{t-2}$$

$$\vdots$$

$$B^k Z_t = B^{k-1}(Z_{t-1}) = B^{k-2}(Z_{t-2}) = \dots = Z_{t-k}$$

se obtiene la variable retardada k periodos. La necesidad de utilizar éste operador se notará ampliamente en las secciones subsecuentes.

Ambos operadores, de rezago y diferencia, están relacionados entre ellos, recuérdese que un proceso integrado de orden k , puede ser factorizado como sigue (*ver sección 2.4*):

$$\nabla^k Z_t = \sum_{i=0}^k \binom{k}{i} (-1)^i Z_{t-i} \quad (2.5.1)$$

al utilizar el operador de rezago en (2.5.1), se tiene:

$$\nabla^k Z_t = \sum_{i=0}^k \binom{k}{i} (-1)^i B^i Z_t \quad (2.5.2)$$

en donde, de acuerdo al *teorema del binomio* se puede observar la forma de un binomio elevado a la k -ésima potencia,

$$\nabla^k Z_t = (1 - B)^k Z_t \quad (2.5.3)$$

de esta manera se concluye, que el operador diferencia está relacionado con el operador de rezago, mediante:

$$\nabla = 1 - B$$

y que la aplicación de k veces el operador ∇ no es otra cosa que un binomio elevado a la k -ésima potencia,

$$\nabla^k = (1 - B)^k \quad (2.5.4)$$

2.6 Descomposición de Wold

Los modelos para series que se estudiarán aquí están basados en la idea de que una serie de tiempo estacionaria, cuyos valores sucesivos pueden ser altamente dependientes, puede considerarse generada a partir de variables aleatorias no correlacionadas a_t , mediante la siguiente relación

$$\tilde{Z}_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \quad (2.6.1)$$

en donde $\tilde{Z}_t = Z_t - \mu$.

Esta idea queda fundamentada en la *representación general de procesos estacionarios* obtenida por Wold en 1938. Este autor demostró que todo proceso débilmente estacionario Z_t que no contenga componentes deterministas puede expresarse en función de a_t mediante (2.6.1) donde $\{a_t\}$ es un proceso de ruido blanco.

Los coeficientes ψ_i tienen que cumplir con ciertas condiciones para que el proceso sea estacionario. La esperanza de \tilde{Z}_t tiene que ser constante al igual que la varianza, y la función de autocovarianza solo debe depender del retardo.

Debido a que a_t es un proceso de ruido blanco, se puede verificar de inmediato que la esperanza de la *representación lineal general*, es constante:

$$E(\tilde{Z}_t) = E(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots) = E(a_t) + \psi_1 E(a_{t-1}) + \psi_2 E(a_{t-2}) + \dots = 0$$

La varianza de \tilde{Z}_t será:

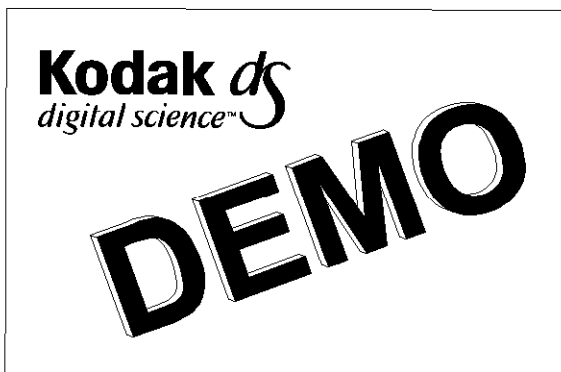
$$Var(\tilde{Z}_t) = Var(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots) = Var \left[\sum_{i=0}^{\infty} \psi_i a_{t-i} \right]$$

con $\psi_0 = 1$, y por lo tanto

$$\text{Var}(Z_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2$$

entonces, para que el proceso Z_t sea estacionario, la serie $\{\psi_i^2\}$ deberá ser convergente.

En el caso de la autocovarianza del proceso Z_t , ésta puede calcularse multiplicando de ambos lados de (2.6.1) por Z_{t-k} y aplicando el operador esperanza, de la siguiente manera:



$$E \left[\sum_{i=0}^{\infty} \psi_i a_{t-i} Z_{t-k} \right]$$

en t , se obtiene:

$$\left[\sum_{i=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t-k-j} \right]$$

Los términos diferentes de cero, serán aquellos en donde

$i = j + k$. De esta manera, la función de autocovarianzas estará dada por:

$$\gamma_k = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

la cual sólo depende del rezago.

Por lo tanto, un proceso de la forma (2.6.1) se dirá que es estacionario sólo si $\sum_{i=0}^{\infty} \psi_i^2 < \infty$.

Dentro de los casos particulares de la representación general o descomposición de Wold se encuentran los modelos de media móvil, los cuales se representan mediante la expresión

$$Z_t = \theta_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.6.2)$$

en donde θ_0 denota a la media del proceso Z_t , a_t un proceso de ruido blanco, y $\theta_1, \theta_2, \dots, \theta_q$ son constantes que sirven para relacionar a las sucesiones $\{Z_t\}$ y $\{a_t\}$ ⁹.

Asimismo, los modelos autorregresivos, que se definen como

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (2.6.3)$$

en donde θ_0 y $\phi_i, i = 1, \dots, p$ son constantes a determinar y a_t un proceso de ruido blanco.

Las características más importantes de los procesos de media móvil y autorregresivos se presentaran posteriormente, por el momento sólo se mencionará que al utilizar el operador de retardo B , un modelo autorregresivo de orden p , (2.6.3), puede representarse como:

$$Z_t = \theta_0 + \phi_1 B Z_t + \phi_2 B^2 Z_t + \dots + \phi_p B^p Z_t + a_t$$

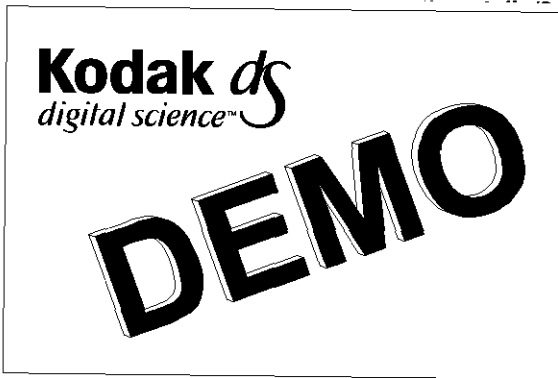
de donde se obtiene:

$$Z_t - \phi_1 B Z_t - \phi_2 B^2 Z_t - \dots - \phi_p B^p Z_t = \theta_0 + a_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = \theta_0 + a_t$$

$$\phi_p(B) Z_t = \theta_0 + a_t$$

⁹Se denotan ahora con θ para indicar que son un número finito de coeficientes de rezago



2) se le puede representar de manera más
 B

$$\theta_2 B^2 a_t - \dots - \theta_q B^q a_t$$

$$(\theta_2 B^2 - \dots - \theta_q B^q) a_t$$

$$= \theta_q(B) a_t$$

donde a $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ y $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ se les conoce como polinomios de rezago de orden p y q respectivamente.

3. Ecuaciones en diferencia

Las ecuaciones en diferencia son los análogos discretos de las ecuaciones diferenciales, juegan un papel importante en el estudio de las series temporales ya que algunas de sus propiedades dependen de las características de las raíces de una ecuación de este tipo. Es por eso que se hará un paréntesis para dar una breve introducción enfocada básicamente a los métodos para obtener sus soluciones.

Una ecuación lineal en diferencia de orden k esta dada por la ecuación

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_k x_{t-k} = f_t \tag{3.1.0}$$

donde $\phi_i, i = 1, \dots, k$, son constantes a determinar. Si la función f_t es idénticamente cero se dice que la ecuación es homogénea y no-homogénea si $f_t \neq 0$. A menos que se indique otra cosa, las ecuaciones que aquí se estudiarán son las ecuaciones lineales homogéneas en diferencia con coeficientes constantes.

3.1 Ecuaciones en diferencia de orden uno

El caso más simple y por ello el más fácil de resolver es la ecuación en diferencia de orden uno

$$x_t - \phi x_{t-1} = 0 \quad (3.1.1)$$

la cual puede ser reescrita como

$$x_t = \phi x_{t-1} \quad (3.1.2)$$

e iterando sucesivamente se obtiene su solución general

$$x_t = \phi x_{t-1} = \phi (\phi x_{t-2}) = \dots = \phi^t x_0. \quad (3.1.3)$$

Se denomina a (3.1.3) *solución general* de la ecuación lineal homogénea (3.1.1) ya que a partir de ella se tiene un número infinito de soluciones para cada valor de x_0 . Dada una condición inicial x_0 la solución se vuelve única.

3.2 Ecuaciones en diferencia de orden dos

CASO 1

Considérese la ecuación lineal homogénea en diferencia de orden dos:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = 0 \quad (3.2.1)$$

Ahora, se demuestra que su solución general es

$$x_t = A_1 G_1^t + A_2 G_2^t \quad (3.2.2)$$

en donde A_1 y A_2 son constantes que dependen de las condiciones iniciales y G_i , $i = 1, 2$ son las raíces de una ecuación del siguiente estilo:

$$y^2 - \phi_1 y - \phi_2 = 0 \quad (3.2.3)$$

Para demostrarlo sustitúyase (3.2.2) en (3.2.1), de la siguiente manera:

$$\begin{aligned} & A_1 G_1' + A_2 G_2' - \phi_1 (A_1 G_1^{t-1} + A_2 G_2^{t-1}) - \phi_2 (A_1 G_1^{t-2} + A_2 G_2^{t-2}) = \\ & = A_1 G_1^{t-2} (G_1^2 - \phi_1 G_1 - \phi_2) + A_2 G_2^{t-2} (G_2^2 - \phi_1 G_2 - \phi_2) = 0 \end{aligned}$$

La última igualdad se debe a que si se observa, G_i es una de las raíces de la ecuación (3.2.3), entonces se cumple que:

$$G_1^2 - \phi_1 G_1 - \phi_2 = 0$$

de igual manera, cuando $y = G_2$

$$G_2^2 - \phi_1 G_2 - \phi_2 = 0$$

Con esto queda demostrado que si G_i , $i = 1, 2$ satisface (3.2.3) entonces (3.2.2) satisface (3.2.1).

En términos del operador de retardo B , (3.2.1) se expresa como

$$(1 - \phi_1 B - \phi_2 B^2)x_t = 0 \quad (3.2.4)$$

en donde puede observarse que la ecuación en diferencia de orden dos puede expresarse como el producto de $1 - \phi_1 B - \phi_2 B^2$ por x_t . Además, si se factoriza B^2 en el primer factor del producto de (3.2.4) obsérvese que éste tiene la forma de (3.2.3):

$$B^2(B^{-2} - \phi_1 B^{-1} - \phi_2)x_t = 0$$

De esta manera, se dice que (3.2.2) es la solución general de la ecuación en diferencia $x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = 0$, en donde G_i^{-1} $i=1,2$, son las raíces de la ecuación $1 - \phi_1 B - \phi_2 B^2 = 0$; ya que si G_i^{-1} satisface $B^{-2} - \phi_1 B^{-1} - \phi_2 = 0$ entonces G_i satisface $B^2 - \phi_1 B - \phi_2 = 0$.

A $1 - \phi_1 B - \phi_2 B^2 = 0$ se le conoce como *ecuación auxiliar* asociada a la ecuación en diferencia y sus raíces están dadas por

$$G_1^{-1} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}, \quad G_2^{-1} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

Hasta ahora se ha supuesto que G_1^{-1} y G_2^{-1} son raíces reales y diferentes, es decir que $\phi_1^2 + 4\phi_2 > 0$, hay que analizar ahora los dos casos restantes en donde $\phi_1^2 + 4\phi_2 = 0$ y $\phi_1^2 + 4\phi_2 < 0$.

CASO 2

Cuando $\phi_1^2 + 4\phi_2 = 0$ la ecuación auxiliar sólo tiene una raíz, $G^{-1} = \frac{\phi_1}{2}$. Por lo que hay que demostrar que la solución general de (3.2.2) puede escribirse como $(A_1 + tA_2)G'$. Es decir que si G' es solución de (3.2.1) entonces tG' también lo es.

Para comprobarlo sustitúyase tG' en (3.2.1)

$$tG' - \phi_1(t-1)G'^{-1} - \phi_2(t-2)G'^{-2} = tG' - \phi_1 B t G' - \phi_2 B^2 t G' = tG'(1 - \phi_1 B - \phi_2 B^2)$$

ahora escríbase a $1 - \phi_1 B - \phi_2 B^2$ en términos de su raíz doble G , es decir como $(1 - GB)^2$, entonces

$$\begin{aligned} tG'(1 - \phi_1 B - \phi_2 B^2) &= tG'(1 - GB)^2 = tG'(1 - GB)(1 - GB) = \\ &= [tG' - tG'^{-1}B](1 - GB) = [tG' - (t-1)G'](1 - GB) = G'(1 - GB) = G' - G' = 0 \end{aligned}$$

se tiene dos soluciones G' y tG' independientes, ya que una no es múltiplo constante de la otra, y se puede expresar la solución general de (3.2.1) como $(A_1 + tA_2)G'$, en donde A_1 y A_2 quedarán determinadas por las condiciones iniciales.

CASO 3

Si $\phi_1^2 + 4\phi_2 < 0$, las raíces de $1 - \phi_1 B - \phi_2 B^2 = 0$ son complejas y G_1 y G_2 son de la forma

$$G_1 = \alpha + i\beta \quad \text{y} \quad G_2 = \alpha - i\beta$$

donde $\alpha = -\frac{\phi_1}{2\phi_2}$ y $\beta = \frac{\sqrt{-\phi_1^2 - 4\phi_2}}{2\phi_2}$, ya que el inverso multiplicativo de un número

complejo $\alpha + i\beta$ está dado por $\frac{\alpha}{\alpha^2 + \beta^2} + i \frac{-\beta}{\alpha^2 + \beta^2}$.

Como las soluciones de la ecuación en diferencia son G_1^t y G_2^t se debe explicar como calcular potencias de números complejos. Esto es más fácil cuando los números están escritos en forma polar.

$$\text{Sea} \quad r = \sqrt{\alpha^2 + \beta^2} \quad \theta = \tan^{-1} \frac{\beta}{\alpha}$$

$$\alpha = r \cos \theta \quad \beta = r \operatorname{sen} \theta$$

usando las dos últimas relaciones, se tiene

$$G_1 = r(\cos \theta + i \operatorname{sen} \theta), \quad G_2 = r(\cos \theta - i \operatorname{sen} \theta)$$

La fórmula de Euler dice que $e^{i\theta} = \cos \theta + i \operatorname{sen} \theta$ con lo cual se puede expresar a G_1 y G_2 como

$$G_1 = r e^{i\theta} \quad \text{y} \quad G_2 = r e^{-i\theta}$$

Entonces

$$G_1' = r' e^{i\theta} = r' (\cos t\theta + i \operatorname{sen} t\theta)$$

y

$$G_2' = r' e^{-i\theta} = r' (\cos t\theta - i \operatorname{sen} t\theta)$$

son las dos soluciones independientes de (3.2.1).

Si G_1' y G_2' son soluciones entonces cualquier combinación lineal entre ellas también lo es, por lo tanto $\frac{1}{2}(G_1' + G_2')$ es solución, lo mismo que $\frac{1}{2i}(G_1' - G_2')$ las cuales son independientes y pueden ser escritas como $G_{1\prime} = r' \cos t\theta$ y $G_{2\prime} = r' \operatorname{sen} t\theta$, respectivamente. Quedando la solución general como

$$x_t = A_1 G_1^t + A_2 G_2^t$$

3.3 Ecuaciones en diferencia de orden k

En esta sección se generalizará para ecuaciones en diferencia de orden $k > 2$ los resultados que se obtuvieron en la sección anterior. Sea

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_k x_{t-k} = 0$$

cuya ecuación auxiliar asociada es:

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_k B^k = 0, \tag{3.3.1}$$

si, las raíces de (3.3.1) son de multiplicidad uno, la solución general es

$$A_1 G_1^t + \dots + A_k G_k^t = \sum_{j=1}^k A_j G_j^t \tag{3.3.2}$$

para demostrarlo se sustituye (3.3.2) en la ecuación en diferencia de orden $k - \text{ésimo}$

$$\sum A_i G_i^t - \phi_1 \sum A_i G_i^{t-1} - \dots - \phi_k \sum A_i G_i^{t-k} = \sum_{i=1}^k A_i G_i^{t-k} (G_i^k - \phi_1 G_i^{k-1} - \dots - \phi_k) = 0$$

Pero cuando la raíz G_i^{-1} , $i = 1, \dots, k$ es de multiplicidad m_i , la solución esta dada por

$$x_t = \sum_{i=1}^N G_i^t \sum_{j=0}^{m_i-1} t^j A_{ij} \tag{3.3.3}$$

en donde m_i indica la multiplicidad de la raíz G_i^{-1} y N el número de raíces diferentes de la ecuación auxiliar.

Es así como:

$$\sum_{i=1}^N m_i = k \quad \text{y} \quad \prod_{i=1}^N (1 - G_i B)^{m_i} = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_k B^k .$$

Para demostrarlo serán de utilidad los siguientes dos lemas:

Lema 3.3.1

Sea $Z_t = At^j$, m un entero no negativo, A una constante y j un entero no negativo menor que m . Entonces $(1 - B)^m Z_t = 0$.

Prueba:

Para $k = 1$, se demuestra que $(1 - B)^k Z_t = 0$.

Ya que, j es un entero no negativo menor que k , en este caso j sólo puede ser cero, por lo tanto $Z_t = At^0 = A$. Sustituyendo $Z_t = A$ en $(1 - B)^m Z_t$, se tiene $(1 - B)A = A - A = 0$.

Asúmase que para $k = m - 1$ la igualdad $(1 - B)^{m-1} Z_t = (1 - B)^{m-1} At^j = 0$, $j < m - 1$ es válida. Ahora, se demuestra que se cumple para $k = m$, con $Z_t = At^j$ y $j < m$

$$\begin{aligned} (1 - B)^m Z_t &= (1 - B)^{m-1} (1 - B) At^j = (1 - B)^{m-1} [At^j - BA t^j] = \\ &= (1 - B)^{m-1} [At^j - A(t-1)^j] = (1 - B)^{m-1} (-A) [(t-1)^j - t^j] \end{aligned}$$

desarrollando el binomio $(t - 1)^j$, el término anterior puede expresarse como

$$\begin{aligned} (1 - B)^{m-1} (-A) \left\{ \sum_{i=0}^j \binom{j}{i} (-1)^{j-i} t^i - t^j \right\} &= \\ = (1 - B)^{m-1} \left\{ -A \sum_{i=0}^{j-1} \binom{j}{i} (-1)^{j-i} t^i \right\} &= \end{aligned}$$

$$= -\sum_{i=0}^{j-1} (1-B)^{m-1} A \binom{j}{i} (-1)^{j-i} t^i$$

En cada uno de los sumandos se pueden utilizar la hipótesis de inducción, ya que i no puede ser mayor a m , ($i = 0, 1, \dots, j-1$), y para cada i fija se tiene que $A \binom{j}{i} (-1)^{j-i}$ es constante. Es así como, $(1-B)^{m-1} A \binom{j}{i} (-1)^{j-i} t^i = 0$ para cada $i = 0, 1, \dots, j-1$ y por lo tanto se concluye que $(1-B)^m Z_t = 0$.

Lema 3.3.2

Sea $Z_t = G' t^j$, m un entero no negativo, $G \neq 0$ y j es un entero no negativo menor que m . Entonces $(1-GB)^m Z_t = 0$.

Prueba:

Notar que, si $m = 1$

$$\begin{aligned} (1-GB)Z_t &= (1-GB)G' t^j = G' t^j - BG'^{j+1} t^j \\ G' t^j - G' (t-1)^j &= G' [t^j - (t-1)^j] = G' (t^j - Bt^j) = G' (1-B)t^j \end{aligned}$$

si, a los resultados anteriores se le multiplica nuevamente por $(1-GB)$, se obtiene:

$$\begin{aligned} (1-GB)(1-GB)Z_t &= (1-GB)(1-GB)G' t^j = (1-GB)G' (1-B)t^j = \\ &= (1-B)[(1-GB)G' t^j] = (1-B)G' (1-B)t^j = G' (1-B)^2 t^j \end{aligned}$$

y así sucesivamente hasta obtener la expresión general para m

$$(1-GB)^m Z_t = G' (1-B)^m t^j$$

por el Lema 3.3.1, $(1-B)^m t' = 0$ con $A=1$, por lo tanto

$$(1-GB)^m Z_t = G'(1-B)^m t' = 0.$$

Regresando a la solución de la ecuación en diferencia de orden k

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_k x_{t-k} = 0 \quad (3.3.4)$$

sustituir (3.3.3) en (3.3.4) para comprobar que es solución

$$\sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} t'^j A_y + \phi_1 \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} (t-1)^j A_y + \phi_2 \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} (t-2)^j A_y + \dots + \phi_k \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} (t-k)^j A_y$$

utilizando el operador de retardo B , se tiene

$$\begin{aligned} & \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} t'^j A_y + \phi_1 B \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} t'^j A_y + \phi_2 B^2 \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} t'^j A_y + \dots + \phi_k B^k \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} t'^j A_y \\ &= \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} t'^j A_y (1 + \phi_1 B + \phi_2 B^2 + \dots + \phi_k B^k) \\ &= \sum_{i=1}^N G'_i \sum_{j=0}^{m_i-1} t'^j A_y (1 - G_1 B)^{m_1} (1 - G_2 B)^{m_2} \dots (1 - G_N B)^{m_k} = 0 \end{aligned}$$

ya que para cada i y j fijas $A_y t'^j G'_i (1 - G_i B)^{m_i} = 0$ por el Lema 3.3.2.

Se nota que una raíz compleja de la ecuación auxiliar aparece siempre en pares, es decir, que si $(\alpha + i\beta)$ es solución, también lo es su conjugado $(\alpha - i\beta)$. Entonces para cada par de raíces complejas de multiplicidad m , la solución de la ecuación homogénea en diferencia

puede contener las secuencias $r^t \cos \theta t$, $r^t \sin \theta t$, $tr^t \cos \theta t$, $tr^t \sin \theta t$, ..., $t^{m-1}r^t \cos \theta t$, $t^{m-1}r^t \sin \theta t$.

4. Modelos estacionarios

En esta sección se analizará más a fondo las características de los modelos autorregresivos y de media móvil, así como la relación dual que hay entre ellos. Posteriormente se indicará como la combinación lineal de estas dos clases de modelos dan origen a modelos más generales: los modelos *ARMA*.

4.1 Modelos autorregresivos (AR)

Como ya se había mencionado un proceso autorregresivo de orden p , denotado por $AR(p)$, está dado por la ecuación:

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (4.1.1)$$

en donde θ_0 , $\phi_i, i = 1, \dots, p$ son constantes a determinar y a_t un proceso de ruido blanco independiente de Z_{t-k} para todo k positivo.

A primera vista puede parecer un modelo de regresión múltiple, sin embargo Z_t no está explicando en términos de variables independientes sino en términos de sus propios valores rezagados.

Si se supone que Z_t es un proceso estacionario, al aplicar la función de medias a (4.1.1), se obtiene el valor de θ_0

$$\mu = \theta_0 + \phi_1 \mu + \phi_2 \mu + \dots + \phi_p \mu$$

$$\theta_0 = (1 - \phi_1 - \phi_2 - \dots - \phi_p) \mu \quad (4.1.2)$$

Sustituyendo (4.1.2) en (4.1.1) y llamando $\tilde{Z}_t = Z_t - \mu$, se tiene el proceso expresado en desviaciones a su media:

$$\begin{aligned} Z_t - \mu &= \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + a_t \\ \tilde{Z}_t &= \phi_1\tilde{Z}_{t-1} + \phi_2\tilde{Z}_{t-2} + \dots + \phi_p\tilde{Z}_{t-p} + a_t \end{aligned} \quad (4.1.3)$$

la cual es la expresión más usada de un $AR(p)$.

La función de autocovarianzas entre Z_{t-k} y Z_t está dada por:

$$E[(Z_{t-k} - \mu)(Z_t - \mu)] = E(\tilde{Z}_{t-k}\tilde{Z}_t)$$

utilizando la expresión (4.1.3)

$$\begin{aligned} E(\tilde{Z}_{t-k}\tilde{Z}_t) &= E[\tilde{Z}_{t-k}(\phi_1\tilde{Z}_{t-1} + \phi_2\tilde{Z}_{t-2} + \dots + \phi_p\tilde{Z}_{t-p} + a_t)] \\ &= \phi_1 E(\tilde{Z}_{t-k}\tilde{Z}_{t-1}) + \phi_2 E(\tilde{Z}_{t-k}\tilde{Z}_{t-2}) + \dots + \phi_p E(\tilde{Z}_{t-k}\tilde{Z}_{t-p}) + E(\tilde{Z}_{t-k}a_t) \end{aligned}$$

por ser a_t independiente de Z_{t-k} para $k > 0$ y γ_k una función par, se tiene

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \dots + \phi_p\gamma_{k-p} \quad k > 0 \quad (4.1.4)$$

Dividiendo (4.1.4) entre γ_0 , para obtener el coeficiente de autocorrelación de acuerdo a su definición, se obtiene

$$\frac{\gamma_k}{\gamma_0} = \phi_1 \frac{\gamma_{k-1}}{\gamma_0} + \phi_2 \frac{\gamma_{k-2}}{\gamma_0} + \dots + \phi_p \frac{\gamma_{k-p}}{\gamma_0} \quad k > 0$$

de donde se deduce la siguiente ecuación en diferencia de orden p

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad k > 0 \quad (4.1.5)$$

cuya solución de acuerdo a la sección (3) es:

$$\rho_k = \sum_{i=1}^N G_i^k \sum_{j=1}^{m_i-1} k^j A_{ij} \quad k > 0 \quad (4.1.6)$$

en donde A_{ij} son constantes a determinar a partir de las condiciones iniciales y G_i^{-1} son las raíces de multiplicidad m_i de la ecuación característica $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$ con $\sum_{i=1}^N m_i = p$. Si $m_i = 1$ para toda i , las raíces G_i^{-1} son distintas y (4.1.6) se reduce a

$$\rho_k = \sum_{i=1}^p A_i G_i^k \quad k > 0 \quad (4.1.7)$$

Tanto en (4.1.6) como en (4.1.7) se observa que G_i , $i = 1, \dots, N$ tiene que ser menor que uno en módulo para todo i , de lo contrario ρ_k crecería sin límite conforme k aumenta y no se cumpliría la condición $\rho_k < 1$, si $G_i = 1$ para alguna i el componente G_i^k no decrece y ρ_k no tiende a cero para ningún retardo, este caso es especial y se analizará por separado en secciones precedentes.

Esta última condición, G_i menor que uno en módulo para toda i , como se verá posteriormente es una condición de estacionariedad ya que como se puede observar no todos los procesos autorregresivos tienen por qué ser estacionarios.

En el caso de que existan G_i^{-1} raíces complejas la estructura de ρ_i se tornará más complicada debido a que contendrá términos senoidales.

Para poder determinar las constantes A_y de (4.1.6) ó A_1, \dots, A_p de (4.1.7) se requerirá de p condiciones iniciales que se pueden obtener mediante el conocimiento de ϕ_i , $i = 1, \dots, p$, y resolviendo el sistema de ecuaciones:

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_2 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p\end{aligned}$$

el cual surge de particularizar la ecuación (4.1.5) para $k = 1, \dots, p$. De igual manera, si de antemano se conoce los coeficientes de autocorrelación ρ_k para $k = 1, \dots, p$, utilizando el mismo sistema quedarán determinados los coeficientes ϕ_i , $i = 1, \dots, p$. A este sistema de ecuaciones se le conoce como ecuaciones de Yule-Walker.

Para ilustrar estos procedimientos, considérese un proceso $AR(2)$

$$\mathcal{Z}_t = \phi_1 \mathcal{Z}_{t-1} + \phi_2 \mathcal{Z}_{t-2} + a_t \quad k > 0 \quad (4.1.8)$$

con función de autocorrelación

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad (4.1.9)$$

Si G_1^{-1} y G_2^{-1} son las raíces reales de la ecuación característica entonces G_1 y G_2 deben de ser menores que uno, en módulo, para que el proceso sea estacionario, es decir:

$$\frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2} < 1 \quad (4.1.10)$$

lo que da la región de estacionariedad

$$\phi_1 + \phi_2 < 1$$

$$\phi_1 - \phi_2 > -1$$

$$\phi_2 < 1$$

para el proceso $AR(2)$.

La solución general de (4.1.9) está dada por

$$\rho_k = A_1 G_1^k + A_2 G_2^k \quad (4.1.11)$$

en donde A_1 y A_2 serán determinadas como sigue:

particularizando (4.1.11) para $k=0$ y $k=1$, se tiene

$$\rho_0 = A_1 + A_2 = 1 \quad (4.1.12)$$

$$\rho_1 = A_1 G_1 + A_2 G_2 \quad (4.1.13)$$

y de la primera ecuación de Yule-Walker

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad (4.1.14)$$

con lo cual sustituyendo (4.1.14) en (4.1.13)

$$\frac{\phi_1}{1-\phi_2} = A_1 G_1 + (1-A_1) G_2$$

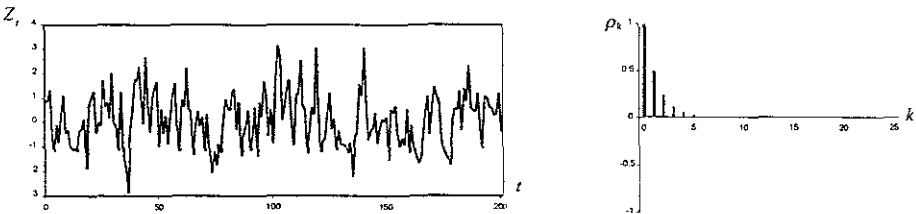
despejando A_1 de la ecuación anterior y A_2 de (4.1.12) se obtiene

$$A_1 = \frac{\left[\frac{\phi_1}{1-\phi_2} - G_2 \right]}{(G_1 - G_2)}$$

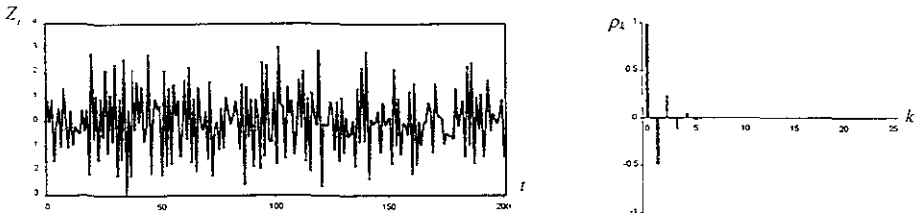
$$A_2 = 1 - A_1$$

Tomando en cuenta todo lo anterior, se concluye que; ρ_k decrecerá conforme aumente el retardo k , en forma exponencial si G_1 y G_2 son reales, y en forma senoidal si G_1 y G_2 son complejas.

A continuación (Figura 4.1) se muestran dos realizaciones de series con su correspondiente correlograma, generadas por $Z_t = 0.5Z_{t-1} + a_t$ y $Z_t = -0.5Z_{t-1} + a_t$, respectivamente.



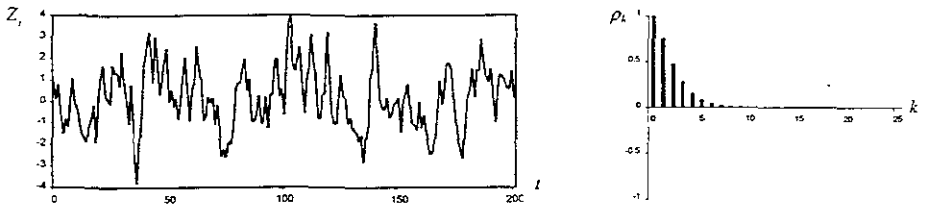
$$AR(1), \phi = 0.5$$



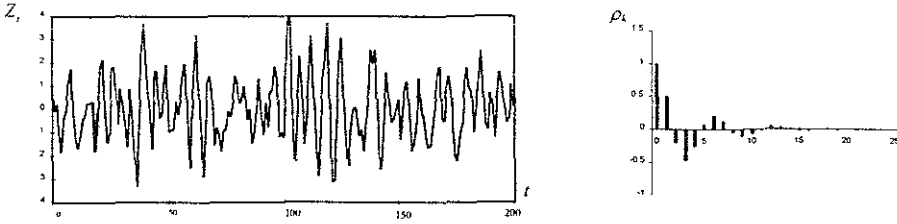
$$AR(1), \phi = -0.5$$

Figura 4.1

De las siguientes figuras, la primera corresponde a la serie generada por $Z_t = 0.9Z_{t-1} - 0.2Z_{t-2} + a_t$, cuya función de autocorrelación muestra un decaimiento exponencial, debido a que las raíces de su ecuación característica son reales; la segunda corresponde a una serie generada $Z_t = 0.8Z_{t-1} - 0.6Z_{t-2} + a_t$, en donde se observa que las raíces de su ecuación característica son complejas debido al comportamiento oscilatorio que muestra su función de autocorrelación.



$$AR(2), \phi_1 = 0.9 \text{ y } \phi_2 = -0.2$$



$$AR(2), \phi_1 = 0.8 \text{ y } \phi_2 = -0.6$$

Figura 4.2

Representación general

De acuerdo a la representación general obtenida por Wold, un proceso autorregresivo será estacionario sólo si puede ser escrito como:

$$\tilde{Z}_t = \psi(B)a_t$$

en donde, a_t es un proceso de ruido blanco y la serie $\psi_1^2, \psi_2^2, \psi_3^2 \dots$ es convergente.

A continuación se demostrará que la condición de que las raíces de la ecuación característica $\phi(B)=0$, para el modelo $\phi_p(B)\tilde{Z}_t = a_t$, estén fuera del círculo unitario es necesaria y suficiente para que éste pueda ser representado de manera general como un

modelo estacionario $\tilde{Z}_t = \psi(B)a_t$, con $\psi(B) = \frac{1}{\phi(B)}$.

Para ello, escríbase al modelo autorregresivo, como:

$$\tilde{Z}_t = \frac{1}{\phi_p(B)} a_t = \psi(B)a_t \quad (4.1.15)$$

con $\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \dots)$ tal que $\phi(B)\psi(B) = 1$.

En términos de las raíces del polinomio $\phi_p(B)$, (4.1.15) se escribe como sigue:

$$Z_t = \frac{1}{\phi(B)} a_t = \left[\prod_{i=1}^p \frac{1}{(1 - G_i B)} \right] a_t$$

Desarrollése el cociente $\prod_{i=1}^p \frac{1}{(1 - G_i B)}$ en fracciones parciales, de la siguiente manera:

$$Z_t = \left[\sum_{i=1}^p \frac{k_i}{(1 - G_i B)} \right] a_t$$

entonces la condición de estacionariedad $\{\psi_i^2$ convergente es equivalente ahora a $G_i < 1$ para toda i .

El valor inverso de $(1 - G_i B)$ con $G_i < 1$, de acuerdo a la expresión para la suma infinita de una progresión geométrica, se define como:

$$(1 - G_i B)^{-1} = \sum_{j=0}^{\infty} G_i^j B^j$$

entonces:

$$Z_t = \sum_{i=1}^p \left[k_i \sum_{j=0}^{\infty} G_i^j B^j \right] a_t = \sum_{j=0}^{\infty} \left[\sum_{i=1}^p k_i G_i^j \right] a_{t-j}$$

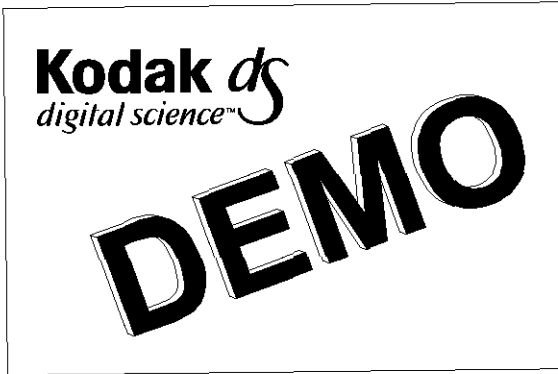
al aplicar la varianza, se tiene:

$$\text{Var}(\tilde{Z}_t) = \sum_{j=0}^{\infty} \left[\sum_{i=1}^p k_i G_i^j \right]^2 \sigma_\varepsilon^2$$

La cual es finita si $|G_i| < 1$.

Los coeficientes ψ_j de la representación general se calculan de manera que

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{p-1} B^{p-1} - \phi_p B^p) (1 + \psi_1 B + \psi_2 B^2 + \dots) = 1$$



$$\dots - \phi_{p-1} (B^{p-1} + \psi_1 B^p + \psi_2 B^{p+1} + \dots) -$$

$$^2 + \dots) = 1$$

se obtiene

$$1 = 1 + (\psi_1 - \phi_1)B + (\psi_2 - \phi_1\psi_1 - \phi_2)B^2 + \dots + (\psi_{p-1} - \phi_1\psi_{p-2} - \phi_2\psi_{p-3} - \dots - \phi_{p-1})B^{p-1} +$$

$$+ (\psi_p - \phi_1\psi_{p-1} - \phi_2\psi_{p-2} - \dots - \phi_{p-1}\psi_1 - \phi_p)B^p + \dots$$

así que, para que se cumpla la igualdad

$$\psi_1 = \phi_1$$

$$\psi_2 = \phi_1\psi_1 + \phi_2$$

⋮

$$\psi_{p-1} = \phi_1\psi_{p-2} + \phi_2\psi_{p-3} + \dots + \phi_{p-1}$$

$$\psi_p = \phi_1\psi_{p-1} + \phi_2\psi_{p-2} + \dots + \phi_{p-1}\psi_1 + \phi_p$$

y, en general

$$\psi_i = \phi_1 \psi_{i-1} + \phi_2 \psi_{i-2} + \dots + \phi_{i-1} \psi_1 + \phi_i \quad \text{Para } i \geq 3$$

con $\phi_i = 0$ para $i > p$.

Tómese como ejemplo de nuevo al proceso $AR(2)$

$$Z_t = \frac{1}{(1 - \phi_1 B - \phi_2 B^2)} \alpha_t = (1 + \psi_1 B + \psi_2 B^2 + \dots) \alpha_t,$$

lo cual implica que

$$(1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \dots) = 1$$

es decir que los coeficientes deben cumplir con las siguientes condiciones,

$$\begin{aligned} \psi_1 &= \phi_1 \\ \psi_2 &= \psi_1 \phi_1 + \phi_2 \\ \psi_3 &= \psi_2 \phi_1 + \psi_1 \phi_2 \\ &\vdots \end{aligned}$$

en general para $j \geq 2$, se tiene

$$\psi_j = \psi_{j-1} \phi_1 + \psi_{j-2} \phi_2 \quad (\psi_0 = 1)$$

En el caso de tener un proceso $AR(1)$, los coeficientes ψ_i serán igual a ϕ^i , $i = 1, 2, \dots$, y se pueda escribir al proceso como:

$$Z_t = (1 + \psi_1 B + \psi_2 B^2 + \dots) a_t$$

Con lo cual se concluye que un proceso *AR* de orden uno será el equivalente de un *MA* de orden infinito.

4.1.1 Función de autocorrelación parcial (*FACP*)

La dependencia entre las observaciones se da de diferentes maneras, por ejemplo en un modelo *AR*(1) el efecto de Z_{t-1} en Z_t es directo y resulta irrelevante considerar a Z_{t-2} , ya que la dependencia entre Z_t y Z_{t-2} (eliminando de ambos los efectos de Z_{t-1}) es nula. En cambio en un *AR*(2) los efectos de Z_{t-2} en Z_t se transmiten, una parte por Z_{t-1} y otra directamente de Z_{t-2} . En este último proceso no se puede ignorar a Z_{t-2} ya que en un futuro puede traer consecuencias el hecho de desperdiciar una parte de información.

Para determinar el orden de un proceso autorregresivo, no serán suficientes las características y propiedades que se analizaron en la sección anterior, se requerirá de otro instrumento que permita la identificación de una manera más clara y simple. A este instrumento se le conoce como *coeficiente de autocorrelación parcial*. Este coeficiente, permite la elección adecuada del orden de un proceso autorregresivo evitando así, la pérdida de información ocasionada por considerar un proceso autorregresivo de orden menor ó la sobreconsideración de variables cuyo efecto directo es nulo.

La labor de la función de autocorrelación parcial consiste en medir la relación lineal que hay entre Z_t y Z_{t-k} , $k > 0$, eliminando de antemano los efectos de las variables intermedias $Z_{t-1}, \dots, Z_{t-k+1}$. Cuando la función es diferente de cero indica que hay un efecto directo entre Z_t y Z_{t-k} que no se transmite a través de las variables $Z_{t-1}, \dots, Z_{t-k+1}$, y lo más adecuado será considerar a esta variable en el modelo.

De esta manera un proceso $AR(p)$ tendrá los primeros p coeficientes de autocorrelación parciales distintos de cero, e igual a cero los coeficientes asociados a $p+1, p+2, \dots$

Para calcular el coeficiente de correlación parcial entre \tilde{Z}_{t-1} y \tilde{Z}_{t-k} se procederá de la siguiente manera:

1) Eliminar de \tilde{Z}_t el efecto de $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots, \tilde{Z}_{t-k+1}$ mediante la regresión

$$\tilde{Z}_t = \alpha_1 \tilde{Z}_{t-1} + \alpha_2 \tilde{Z}_{t-2} + \dots + \alpha_{k-1} \tilde{Z}_{t-k+1} + u_t \quad (4.1.1.1)$$

en donde u_t recogerá la parte de \tilde{Z}_t no común con $\tilde{Z}_{t-1}, \dots, \tilde{Z}_{t-k+1}$. Procediendo de igual manera que en la sección anterior, multiplicando de ambos lados de (4.1.1.1) por \tilde{Z}_{t-i} y aplicando el operador esperanza, se obtiene la ecuación:

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + \dots + \alpha_{k-1} \gamma_{i-k+1} \quad i > 0 \quad (4.1.1.2)$$

y las ecuaciones de Yule-Walker, estarán dadas por:

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-2} + \dots + \alpha_{k-1} \rho_{i-k+1} \quad i > 0$$

obtenidas de dividir (4.1.1.2) sobre γ_0 . Particularizando para $i=1, 2, \dots, k-1$ se obtiene el siguiente sistema de ecuaciones:

$$\begin{aligned} \rho_1 &= \alpha_1 \rho_0 + \alpha_2 \rho_1 + \dots + \alpha_{k-1} \rho_{k-2} \\ \rho_2 &= \alpha_1 \rho_1 + \alpha_2 \rho_0 + \dots + \alpha_{k-1} \rho_{k-3} \\ &\vdots \\ \rho_{k-1} &= \alpha_1 \rho_{k-2} + \alpha_2 \rho_{k-3} + \dots + \alpha_{k-1} \rho_0 \end{aligned}$$

que en términos matriciales se puede escribir de la siguiente manera:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \dots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix} \quad (4.1.1.3)$$

A través de este sistema de ecuaciones y conociendo los valores de ρ_i , se pueden obtener los parámetros α_i , $i = 1, \dots, k-1$.

2) De igual manera, mediante la regresión

$$\tilde{Z}_{t-k} = \beta_1 \tilde{Z}_{t-k+1} + \beta_2 \tilde{Z}_{t-k+2} + \dots + \beta_{k-1} \tilde{Z}_{t-1} + v_t$$

se eliminará de \tilde{Z}_{t-k} el efecto de $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots, \tilde{Z}_{t-k+1}$.

De nuevo v_t contendrá la parte de \tilde{Z}_{t-k} no común con $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots, \tilde{Z}_{t-k+1}$. Y resolviendo el sistema:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{bmatrix} \quad (4.1.1.4)$$

se conocerán los parámetros β_i , $i = 1, \dots, k$. Obsérvese que si α_i satisface (4.1.1.3), también lo hace con (4.1.1.4) con lo cual $\alpha_i = \beta_i$, ($1 \leq i \leq k-1$).

3) El coeficiente de correlación simple entre u_t y v_t será el coeficiente de autocorrelación parcial de orden k que se denotará por P_k .

$$P_k = \frac{Cov(u_i, v_i)}{\sqrt{Var(u_i)Var(v_i)}}$$

de (4.1.1.1) se obtendrá la $Var(u_i)$

$$\begin{aligned} Var(u_i) &= Var(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1}) = \\ &= E[(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1})(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1})] = \\ &= E[Z_i(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1})] - \alpha_1 E[Z_{i-1}(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1})] \\ &\quad - \dots - \alpha_{k-1} E[Z_{i-k+1}(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1})] \\ &= \gamma_0 - \alpha_1 \gamma_1 - \alpha_2 \gamma_2 - \dots - \alpha_{k-1} \gamma_{k-1} - \alpha_1 (\gamma_1 - \alpha_1 \gamma_0 - \alpha_2 \gamma_1 - \dots - \alpha_{k-1} \gamma_{k-2}) - \\ &\quad - \dots - \alpha_{k-1} (\gamma_{k-1} - \alpha_1 \gamma_k - \dots - \alpha_{k-1} \gamma_0) \end{aligned}$$

el término $\alpha_1 (\gamma_1 - \alpha_1 \gamma_0 - \dots - \alpha_{k-1} \gamma_{k-2})$ de ésta última expresión será igual a cero, ya que para $i=1$ se obtiene de (4.1.1.2) la igualdad $\gamma_1 = \alpha_1 \gamma_0 + \alpha_2 \gamma_1 + \dots + \alpha_{k-1} \gamma_{k-2}$. Y así sucesivamente si se particulariza para $i=2, 3, \dots, k-1$ la ecuación (4.1.1.2), se llega finalmente a la siguiente expresión para la $Var(u_i)$

$$Var(u_i) = \gamma_0 - \alpha_1 \gamma_1 - \alpha_2 \gamma_2 - \dots - \alpha_{k-1} \gamma_{k-1} \quad (4.1.1.5)$$

Además $Var(u_i) = Var(v_i)$ debido a que $\alpha_i = \beta_i$, para $i=1, 2, \dots, k-1$.

Desarrólese ahora la $Cov(u_i, v_i)$:

$$Cov(u_i, v_i) = E[(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1})(Z_{i+k} - \beta_1 Z_{i-k+1} - \beta_2 Z_{i-k+2} - \dots - \beta_{k-1} Z_{i+k-1})]$$

haciendo uso, una vez más, de la igualdad $(1 \leq i \leq k-1)$, se tiene

$$\begin{aligned} \text{Cov}(u_i, v_i) &= E[(Z_i - \alpha_1 Z_{i-1} - \alpha_2 Z_{i-2} - \dots - \alpha_{k-1} Z_{i-k+1})(Z_{i-k} - \alpha_1 Z_{i-k+1} - \alpha_2 Z_{i-k+2} - \dots - \alpha_{k-1} Z_{i-1})] \\ &= E[Z_{i-k}(Z_i - \alpha_1 Z_{i-1} - \dots - \alpha_{k-1} Z_{i-k+1}) - \alpha_1 Z_{i-k+1}(Z_i - \alpha_1 Z_{i-1} - \dots - \alpha_{k-1} Z_{i-k+1}) - \dots \\ &\quad - \alpha_{k-1} Z_{i-1}(Z_i - \alpha_1 Z_{i-1} - \dots - \alpha_{k-1} Z_{i-k+1})] \\ &= \gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1 - \alpha_1 (\gamma_{k-1} - \alpha_1 \gamma_{k-2} - \dots - \alpha_{k-1} \gamma_0) - \dots - \alpha_{k-1} (\gamma_1 - \alpha_1 \gamma_0 - \dots - \alpha_{k-1} \gamma_{k-2}) \end{aligned}$$

utilizando de nuevo la ecuación (4.1.1.2), válida para $i > 0$ se concluye:

$$\text{Cov}(u_i, v_i) = \gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1$$

Por lo tanto

$$\begin{aligned} P_k &= \frac{\text{Cov}(u_i, v_i)}{\sqrt{\text{Var}(u_i)} \sqrt{\text{Var}(v_i)}} = \frac{\text{Cov}(u_i, v_i)}{\text{Var}(u_i)} = \frac{\gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1}{\gamma_0 - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_{k-1}} = \\ &= \frac{\rho_k - \alpha_1 \rho_{k-1} - \dots - \alpha_{k-1} \rho_1}{1 - \alpha_1 \rho_1 - \dots - \alpha_{k-1} \rho_{k-1}} \end{aligned} \quad (4.1.1.6)$$

Por otro lado, utilizando la regla de Cramer para resolver el sistema de ecuaciones (4.1.1.3), se pueden conocer los valores de α_i , en términos de los coeficientes de autocorrelación

ρ_i :

$$\alpha_1 = \frac{\begin{matrix} \rho_1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-3} & \dots & 1 \end{matrix}}{\begin{matrix} 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{matrix}}, \quad \alpha_2 = \frac{\begin{matrix} 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_1 & \rho_2 & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-1} & \dots & 1 \end{matrix}}{\begin{matrix} 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{matrix}}$$

en general:

$$\alpha_i = \frac{\begin{matrix} 1 & \rho_1 & \dots & \rho_{i-2} & \rho_1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \dots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \dots & \rho_{k-i} & \rho_{k-1} & \rho_{k-i-2} & \dots & 1 \end{matrix}}{\begin{matrix} 1 & \rho_1 & \dots & \rho_{i-2} & \rho_{i-1} & \rho_1 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \dots & \rho_{i-3} & \rho_{i-2} & \rho_{i-1} & \dots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \dots & \rho_{k-i} & \rho_{k-i-1} & \rho_{k-i-2} & \dots & 1 \end{matrix}}$$

sustituyendo α_i , $i=1, \dots, k-1$ en (4.1.1.6), se puede ver a P_k como la expansión de la siguiente expresión en términos de la última columna.

$$P_k = \frac{\begin{matrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{matrix}}{\begin{matrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{matrix}} \quad (4.1.1.7)$$

En seguida se demostrará que P_k es igual a ϕ_{kk} , en donde ϕ_{kk} es el k -ésimo coeficiente de regresión del siguiente modelo

$$\tilde{Z}_{t+k} = \phi_{k1}\tilde{Z}_{t+k-1} + \phi_{k2}\tilde{Z}_{t+k-2} + \dots + \phi_{kk}\tilde{Z}_t + e_{t+k}$$

en donde e_{t+k} es un término de error no correlacionado con \tilde{Z}_{t+k-j} con $j > 0$.

Utilizando de nuevo las ecuaciones de Yule-Walker y aplicando la regla de Cramer sucesivamente se obtienen los valores de ϕ_{kk}

$$\phi_{11} = \rho_1$$

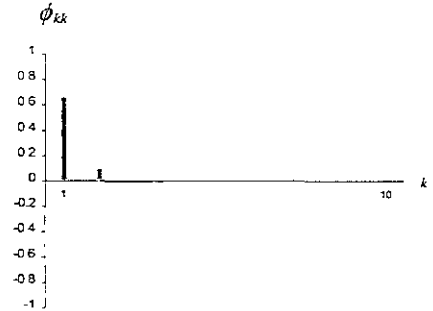
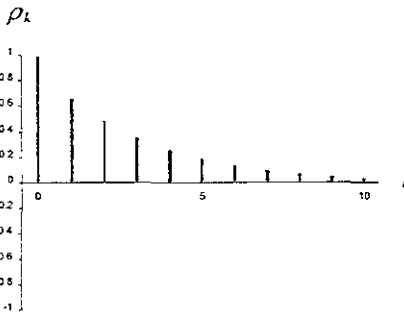
$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

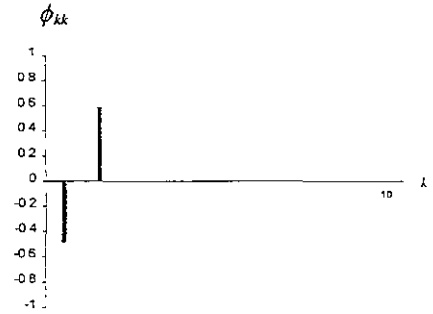
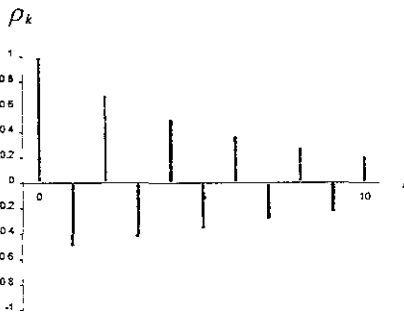
$$\vdots$$

$$\begin{array}{cccccc}
 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\
 \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\
 \vdots & \vdots & \vdots & & \vdots & \vdots \\
 \phi_{kk} = \frac{\rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k}{1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1}} \\
 \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\
 \vdots & \vdots & \vdots & & \vdots & \vdots \\
 \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1
 \end{array} \tag{4.1.1.8}$$

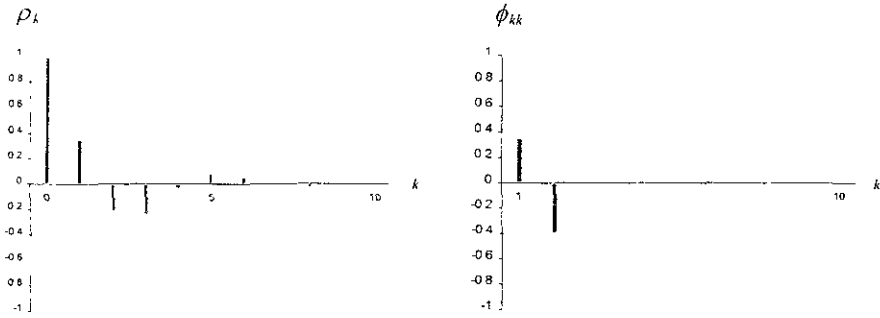
comparando (4.1.1.8) con (4.1.1.7) se concluye que P_k es igual a ϕ_{kk} . A esta sucesión se le denomina *función de autocorrelación parcial* (FACP). Ver fig. 4.1.1, en donde se muestra los diferentes comportamientos de la FAC y la FACP de un proceso autorregresivo de orden dos $(1 - \phi_1 B - \phi_2 B^2)Z_t = a_t$,



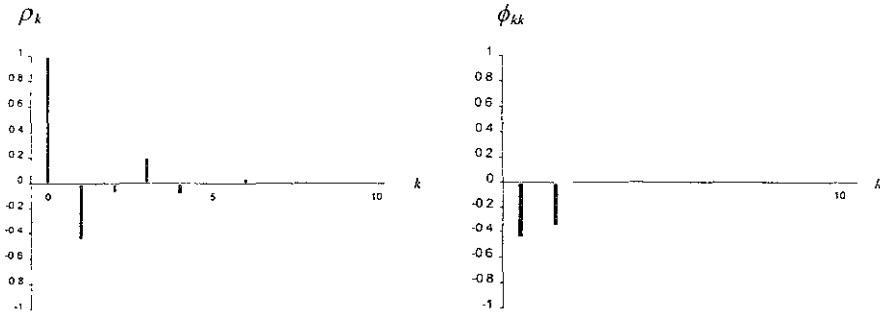
$\phi_1 = 0.6, \phi_2 = 0.1$



$$\phi_1 = -0.2, \phi_2 = 0.6$$



$$\phi_1 = 0.5, \phi_2 = -0.4$$



$$\phi_1 = -0.6, \phi_2 = -0.35$$

Figura 4.1.1

4.2 Procesos de media móvil (MA)

Un proceso de media móvil de orden q se define como:

$$Z_t = \theta_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

en donde a_t es un proceso de ruido blanco y $\theta, i=1, 2, \dots, q$ son constantes a determinar.

El valor de θ_0 se obtiene aplicando la función esperanza al proceso $MA(q)$

$$\mu = E(Z_t) = E(\theta_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}) = \theta_0$$

por lo tanto θ_0 representa a la media del proceso Z , y éste podrá ser representado en desviaciones a su media como sigue:

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \tag{4.2.1}$$

con $\tilde{Z}_t = Z_t - \mu$.

Obsérvese de inmediato que un proceso de media móvil será siempre estacionario, debido a que es un caso particular de la representación general o descomposición de Wold, en donde $\psi_0 = 1$, $\psi_1 = -\theta_1$, ..., $\psi_q = -\theta_q$, $\psi_{q+1} = 0$, $\psi_{q+2} = 0, \dots$, cumpliéndose así, la condición de estacionariedad $\sum_{i=0}^{\infty} \psi_i^2 < \infty$.

Utilizando el operador de retardo B en (4.2.1), se tiene

$$\tilde{Z}_t = (1 - B\theta_1 - B^2\theta_2 - \dots - B^q\theta_q) a_t \tag{4.2.2}$$

Si se aplica la función esperanza y la función varianza a (4.2.1) se obtiene

$$E(\tilde{Z}_t) = 0 \quad \text{y} \quad \gamma_0 = Var(\tilde{Z}_t) = \sigma_a^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$$

que no dependen del tiempo. Para calcular la función de autocovarianzas, multiplíquese de ambos lados de (4.2.1) por \tilde{Z}_{t-k}

$$\tilde{Z}_i \tilde{Z}_{i-k} = a_i \tilde{Z}_{i-k} - \theta_1 a_{i-1} \tilde{Z}_{i-k} - \dots - \theta_q a_{i-q} \tilde{Z}_{i-k}$$

sustituir a \tilde{Z}_{i-k} en términos de a_i

$$\begin{aligned} \tilde{Z}_i \tilde{Z}_{i-k} = & a_i (a_{i-k} - \theta_1 a_{i-k+1} - \dots - \theta_q a_{i-k+q}) - \theta_1 a_{i-1} (a_{i-k} - \theta_1 a_{i-k+1} - \dots - \theta_q a_{i-k+q}) - \dots \\ & - \theta_q a_{i-q} (a_{i-k} - \theta_1 a_{i-k+1} - \dots - \theta_q a_{i-k+q}) \end{aligned}$$

aplicar ahora el operador esperanza

$$\begin{aligned} E(\tilde{Z}_i \tilde{Z}_{i-k}) = & E[a_i (a_{i-k} - \theta_1 a_{i-k+1} - \dots - \theta_q a_{i-k+q})] - \theta_1 E[a_{i-1} (a_{i-k} - \theta_1 a_{i-k+1} - \dots - \theta_q a_{i-k+q})] - \dots \\ & - \theta_q E[a_{i-q} (a_{i-k} - \theta_1 a_{i-k+1} - \dots - \theta_q a_{i-k+q})] \end{aligned}$$

debido a que el proceso a_i es un proceso de ruido blanco, se obtiene:

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma_a^2, & \text{si } k=1, \dots, q \\ 0, & \text{si } k \geq q+1 \end{cases}$$

que dependen únicamente del retardo y de donde se sigue que la función de autocorrelación es:

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{si } k=1, \dots, q \\ 0 & \text{si } k \geq q+1 \end{cases}$$

Esta función jugará un papel importante en la determinación del orden del proceso MA , ya que será diferente de cero solo para $k \leq q$.

Se dirá que un proceso es *invertible* si la influencia del pasado decrece con el tiempo. Para aclarar esta propiedad tómesese como ejemplo a dos procesos de media móvil de orden uno

$$i) \quad \tilde{Z}_t = a_t - \theta a_{t-1}$$

$$ii) \quad \tilde{Z}_t = a_t - \frac{1}{\theta} a_{t-1}$$

es fácil demostrar que su función de autocorrelación es la misma para ambos procesos. Para ello multiplíquese de ambos lados del primer proceso por \tilde{Z}_{t-k} , de donde se obtiene

$$\tilde{Z}_t \tilde{Z}_{t-k} = a_t \tilde{Z}_{t-k} - \theta a_{t-1} \tilde{Z}_{t-k}$$

sustituyendo \tilde{Z}_{t-k} en función de a_t , sólo en el término de la derecha, y aplicando el operador esperanza:

$$E(\tilde{Z}_t \tilde{Z}_{t-k}) = E[a_t (a_{t-k} - \theta a_{t-k-1}) - \theta a_{t-1} (a_{t-k-1} - \theta a_{t-k-2})] =$$

$$= E(a_t a_{t-k}) - \theta E(a_t a_{t-k-1}) - \theta E(a_{t-1} a_{t-k-1}) - \theta^2 E(a_{t-1} a_{t-k-2})$$

de aquí que la función de autocovarianza

$$\gamma_1 = -\theta \sigma_a^2$$

$$\gamma_2 = 0$$

$$\vdots$$

dividiendo entre la varianza, $\gamma_0 = \sigma_a^2(1 + \theta^2)$, se obtiene la función de autocorrelación

$$\rho_k = \begin{cases} \frac{-\theta}{1+\theta^2} & k=1 \\ 0 & k>1 \end{cases}$$

siguiendo el mismo procedimiento se llega a que la función de autocorrelación del segundo proceso es:

$$\rho_k = \begin{cases} \frac{-1}{1+\frac{1}{\theta^2}} & k=1 \\ 0 & k>1 \end{cases}$$

que es igual a la del primer proceso. Sin embargo solo uno de ellos es invertible.

Sustituyendo sucesivamente a_{t-1}, a_{t-2}, \dots , en función de $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots$, se tiene que

$$i) \tilde{Z}_t = a_t - \theta \tilde{Z}_{t-1} + \theta^2 \tilde{Z}_{t-2} - \dots$$

$$ii) \tilde{Z}_t = a_t - \frac{1}{\theta} \tilde{Z}_{t-1} + \frac{1}{\theta^2} \tilde{Z}_{t-2} - \dots$$

si $\theta < 1$, la influencia de la variable retrasada sobre la variable actual será menor conforme el retardo aumente y se dirá que el proceso es invertible. No cumpliendo esta propiedad el proceso del inciso ii).

El considerar sólo procesos invertibles asegurará que a cada función de autocorrelación le corresponderá un único proceso de media móvil.

Así como un proceso MA es siempre estacionario, un proceso AR será siempre invertible ya que la influencia de la variable retrasada disminuye con el paso del tiempo. Por lo tanto, un proceso de media móvil será invertible si éste puede reescribirse como un proceso autorregresivo.

Sin embargo, para que un proceso MA sea invertible tendrá que cumplir con ciertas condiciones las cuales son las análogas a las condiciones de estacionariedad para un proceso autorregresivo. Es decir que las raíces de la ecuación característica $\theta(B)=0$ deberán estar fuera del círculo unitario.

Para demostrar que un modelo de media móvil será invertible sólo si las raíces de la ecuación característica están fuera del círculo unitario, considérese un proceso MA de orden q :

$$Z_t = \theta_q(B)\alpha_t \tag{4.2.3}$$

el cual, se puede escribir como:

$$\pi(B)Z_t = \frac{1}{\theta_q(B)}Z_t = \alpha_t \tag{4.2.4}$$

en donde $\pi(B)=1-\pi_1B-\pi_2B^2-\dots$ es un polinomio de rezago de orden infinito que satisface la ecuación $\theta_q(B)\pi(B)=1$.

A $\frac{1}{\theta_q(B)}$ se le puede escribir en términos de sus raíces

$$\frac{1}{\theta_q(B)} = \prod_{i=1}^q \frac{1}{(1-G_iB)}$$

y desarrollando en fracciones parciales

$$\theta^{-1}(B) = \sum_{i=1}^q \frac{\lambda_i}{(1 - G_i B)}$$

si $G_i < 1$ se define el valor inverso de $(1 - G_i B)$, como:

$$(1 - G_i B)^{-1} = \sum_{j=0}^{\infty} G_i^j B^j$$

y (4.2.4) puede ser expresado como

$$\sum_{k=0}^{\infty} \left(\sum_{i=1}^q \lambda_i G_i^k \right) Z_{t-k} = a_t$$

en donde se observa que $\sum_{i=1}^q \lambda_i G_i^k$ decrece conforme el retardo k aumenta.

Por simplicidad, se ha supuesto que las raíces de $\theta_q(B) = 0$ son todas distintas, pero estos resultados se pueden extender para el caso en el que existan raíces iguales.

Los coeficientes del polinomio $\pi(B)$ se obtienen de la relación

$$\pi(B)\theta(B) = 1$$

es decir, que los coeficientes del polinomio $\pi(B)$ deberán satisfacer las ecuaciones siguientes:

$$\pi_1 = -\theta_1$$

$$\pi_2 = \pi_1\theta_1 - \theta_2$$

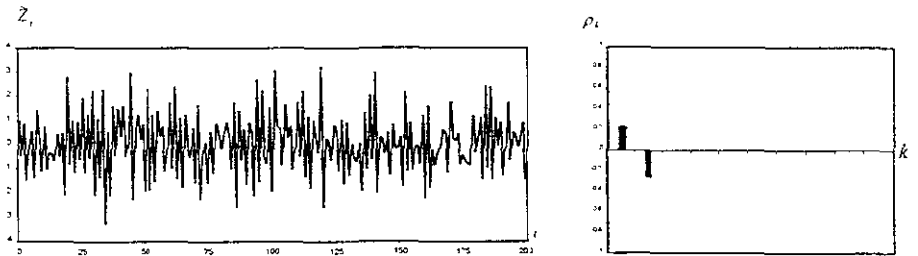
$$\pi_3 = \pi_2\theta_1 + \pi_1\theta_2 - \theta_3$$

⋮

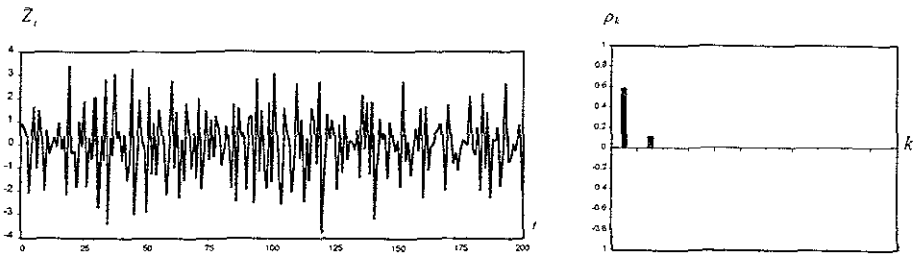
$$\pi_j = \pi_{j-1}\theta_1 + \pi_{j-2}\theta_2 + \dots + \pi_1\theta_{j-1} - \theta_j, \quad j > q$$

en donde $\theta_j = 0$ para $j > q$.

En el caso de tener un proceso *MA* de primer orden éste será el equivalente a un proceso *AR* de orden infinito. Este resultado es el dual del obtenido en la sección anterior. Esta dualidad también se da entre la *FAC* y la *FACP*; el proceso *AR*(*p*) tiene una *FAC* que va disminuyendo conforme aumenta el retardo y una *FACP* limitada a *p* retardos, mientras que el proceso *MA*(*q*) tiene una *FAC* limitada a *q* retardos y una *FACP* que disminuirá conforme aumenta el retardo. Ver fig. 4.2.1.



Serie generada por el proceso $Z_t = a_t + 0.5a_{t-1} + 0.36a_{t-2}$

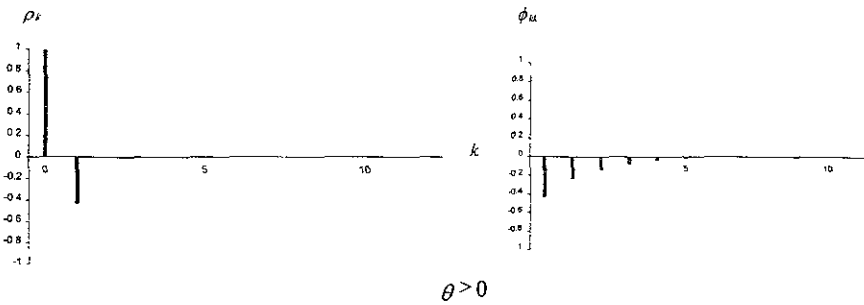


Serie generada por el proceso $Z_t = a_t + 0.9a_{t-1} + 0.22a_{t-2}$

Figura 4.2.1

A continuación, Fig. 4.2.2, se presenta la FAC^{\otimes} y la $FACP^{\otimes}$ de procesos $MA(1)$:

$$Z_t = (1 - \theta B)a_t,$$



⁰¹ Función de autocorrelación.

⁰² Función de autocorrelación parcial.

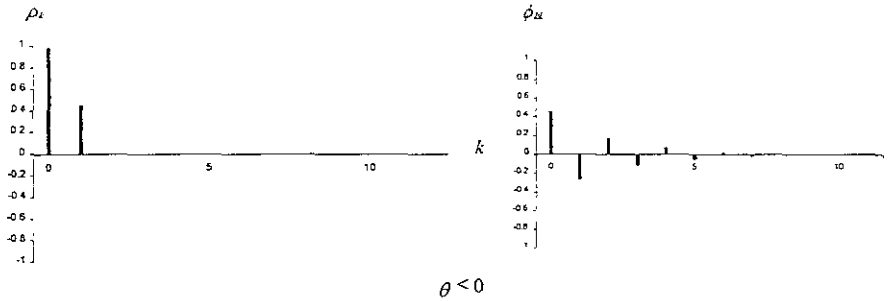


Figura 4.2.2

4.3 Procesos ARMA

De las secciones anteriores se concluye que un proceso estacionario $AR(p)$ puede representarse mediante un proceso $MA(\infty)$ y la explicación del fenómeno será la misma. Sin embargo, si se cuenta con un número finito de observaciones se preferirá trabajar con un modelo que involucre un número finito de parámetros que con un modelo de orden infinito, ya que entre más parámetros haya en el modelo la estimación de éstos será menos eficiente. De igual manera que se preferirá trabajar con un modelo invertible $MA(q)$ que con su representación $AR(\infty)$. A pesar de esto, el modelo puede contener aún un número bastante grande de parámetros, por lo cual una mezcla de componentes autorregresivos y de media móvil darán un nuevo modelo cuyo número de parámetros sea menor. Este modelo se denomina *autorregresivo de media móvil (ARMA)* y al ahorro de parámetros se le conoce como *principio de parsimonia*.

El caso general de un proceso $ARMA(p, q)$ (con p componentes autorregresivos y q de media móvil), se representa mediante:

$$\phi_p(B)Z_t = \theta_q(B)\alpha_t \tag{4.3.1}$$

donde $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ y $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$.

Para que el proceso sea estacionario e invertible se requiere que las raíces de $\phi_p(B)=0$ y de $\theta_q(B)=0$ estén fuera del círculo unitario respectivamente, además, de que se está suponiendo que $\phi_p(B)=0$ y $\theta_q(B)=0$ no tienen raíces iguales.

Si un proceso *ARMA* es estacionario e invertible puede representarse como un $AR(\infty)$ o como un $MA(\infty)$ debido a su estructura mixta,

$$\pi(B)Z_t = \frac{\phi_p(B)}{\theta_q(B)} Z_t = a_t \quad \text{ó} \quad \psi(B)a_t = \frac{\theta_q(B)}{\phi_p(B)} Z_t = Z_t$$

en donde los coeficientes de $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$; resultarán de

$$\theta_q(B)\pi(B) = \phi_p(B)$$

y los coeficientes de $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$; de

$$\phi_p(B)\psi(B) = \theta_q(B)$$

Reescribiendo la ecuación (4.3.1) como

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

si el proceso es estacionario $E(Z_t) = 0$; multiplicando por Z_{t-k}

$$Z_t Z_{t-k} = (\phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}) Z_{t-k}$$

$$Z_t Z_{t-k} = \phi_1 Z_{t-1} Z_{t-k} + \dots + \phi_p Z_{t-p} Z_{t-k} + a_t Z_{t-k} - \theta_1 a_{t-1} Z_{t-k} - \dots - \theta_q a_{t-q} Z_{t-k}$$

tomando esperanzas:

$$E(Z_t, Z_{t-k}) = E(\phi_1 Z_{t-1} Z_{t-k} + \dots + \phi_p Z_{t-p} Z_{t-k} + a_t Z_{t-k} - \theta_1 a_{t-1} Z_{t-k} - \dots - \theta_q a_{t-q} Z_{t-k})$$

por lo tanto,

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + E(Z_{t-k} a_t) - \theta_1 E(Z_{t-k} a_{t-1}) - \dots - \theta_q E(Z_{t-k} a_{t-q})$$

en donde como Z_{t-k} está afectada por a_{t-k} , a_{t-k-1} , ..., pero es independiente de a_{t-k+1} , a_{t-k+2} , se tiene que

$$E(Z_{t-k} a_{t-i}) = 0 \quad \text{para } k > i$$

y por lo tanto

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} \quad \text{para } k > q$$

de aquí, dividiendo entre γ_0 se obtiene

$$\frac{\gamma_k}{\gamma_0} = \phi_1 \frac{\gamma_{k-1}}{\gamma_0} + \dots + \phi_p \frac{\gamma_{k-p}}{\gamma_0} \quad \text{para } k > q$$

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} \quad \text{para } k > q \quad (4.3.2)$$

La solución de esta ecuación en diferencia, como se vio en la sección (3), está dada por:

$$\rho_k = \sum_{i=1}^p G_i^k \sum_{j=1}^{m_i-1} k^j A_{ij}$$

en donde A_j son constantes a determinar a partir de p condiciones iniciales y G_i^{-1} son las raíces de multiplicidad m_i de $\phi(B)=0$.

Por lo tanto, las primeras q autocorrelaciones tendrán cualquier estructura ya que involucran tanto a parámetros autorregresivos como de media móvil y a partir de $q+1$ la *FAC* se comportará de acuerdo a (4.3.2) y a las p condiciones iniciales; decreciendo de manera exponencial y/o senoidal.

La función de autocorrelación parcial se comporta de manera análoga a la *FAC*; por lo cual es difícil de determinar el orden de un proceso *ARMA* a partir de su función de autocorrelación simple o parcial.

Estos resultados se observan de manera fácil en un proceso *ARMA*(1,1) que es el más sencillo y se utiliza mucho en la práctica.

Este modelo está definido como:

$$(1 - \phi B)Z_t = (1 - \theta B)a_t \quad (4.3.3)$$

se asume $|\phi| < 1$ para que sea estacionario, $|\theta| < 1$ para que sea invertible y además $\phi = \theta$.

Cuando $\phi = 0$ el modelo (4.3.3) se reduce a un *MA*(1)

$$Z_t = a_t - \theta a_{t-1}$$

y a un *AR*(1) cuando $\theta = 0$.

$$Z_t = \phi Z_{t-1} + a_t$$

Para obtener la función de autocovarianzas del modelo $ARMA(1,1)$, se multiplica de ambos lados de (4.3.3) por \tilde{Z}_{t-k}

$$\tilde{Z}_t \tilde{Z}_{t-k} = \phi \tilde{Z}_{t-1} \tilde{Z}_{t-k} + a_t \tilde{Z}_{t-k} - \theta a_{t-1} \tilde{Z}_{t-k}$$

y tomando esperanzas, se obtiene

$$E(\tilde{Z}_t \tilde{Z}_{t-k}) = E(\phi \tilde{Z}_{t-1} \tilde{Z}_{t-k} + a_t \tilde{Z}_{t-k} - \theta a_{t-1} \tilde{Z}_{t-k})$$

$$\gamma_k = \phi \gamma_{k-1} + E(\tilde{Z}_{t-k} a_t) - \theta E(\tilde{Z}_{t-k} a_{t-1})$$

Más específicamente, para $k = 0$

$$\gamma_0 = \phi \gamma_1 + E(\tilde{Z}_t a_t) - \theta E(\tilde{Z}_t a_{t-1})$$

de donde se obtiene:

$$\begin{aligned} \gamma_0 &= \phi \gamma_1 + E[(\phi \tilde{Z}_{t-1} + a_t - \theta a_{t-1}) a_t] - \theta E[(\phi \tilde{Z}_{t-1} + a_t - \theta a_{t-1}) a_{t-1}] = \\ &= \phi \gamma_1 + \sigma_a^2 - \phi \theta E[(\phi \tilde{Z}_{t-2} + a_{t-1} - \theta a_{t-2}) a_{t-1}] + \theta \sigma_a^2 = \\ &= \phi \gamma_1 + \sigma_a^2 - \phi \theta \sigma_a^2 + \theta^2 \sigma_a^2 = \phi \gamma_1 + \sigma_a^2 - \theta (\phi + \theta) \sigma_a^2 \end{aligned}$$

para $k = 1$, la función de autocorrelación del proceso $ARMA(1,1)$ se calcula como:

$$\gamma_1 = \phi \gamma_0 + E(\tilde{Z}_{t-1} a_t) - \theta E(\tilde{Z}_{t-1} a_{t-1}) = \phi \gamma_0 - \theta \sigma_a^2$$

para dejar a la función de autocorrelación sólo en términos de ϕ y θ , sustitúyase el valor de γ_1 en γ_0 , es así como:

$$\gamma_0 = \frac{(1 + \theta^2 - 2\phi\theta)}{(1 - \phi^2)} \sigma_a^2$$

de igual manera, ahora sustitúyase el valor de γ_0 en γ_1 , para obtener:

$$\gamma_1 = \frac{(\phi - \theta)(1 - \phi\theta)}{(1 - \phi^2)} \sigma_a^2$$

es así, como la función de autocorrelación de un proceso $ARMA(1,1)$ se puede resumir de la siguiente manera:

$$\gamma_k = \begin{cases} \frac{(1 + \theta^2 - 2\phi\theta)}{(1 - \phi^2)} \sigma_a^2 & k = 0 \\ \frac{(\phi - \theta)(1 - \phi\theta)}{(1 - \phi^2)} \sigma_a^2 & k = 1 \\ \phi\gamma_{k-1} & k \geq 2 \end{cases}$$

De donde se sigue que la *FAC* para el modelo $ARMA(1,1)$ es:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{(\phi - \theta)(1 - \phi\theta)}{1 - 2\phi\theta + \theta^2} & k = 1 \\ \phi\rho_{k-1} & k > 1 \end{cases}$$

Esta función tendrá un valor inicial que depende de los parámetros autorregresivos y de media móvil y después un decrecimiento geométrico que depende sólo de ϕ .

Para calcular la *FACP* no hay que olvidar que el modelo $ARMA(1,1)$, $(1 - \phi B)\tilde{Z}_t = (1 - \theta B)\alpha_t$, puede escribirse como:

$$(1 - \theta B)^{-1}(1 - \phi B)\tilde{Z}_t = \alpha_t,$$

de donde se obtiene, dado que $\theta < 1$:

$$\sum_{i=0}^{\infty} \theta^i B^i (1 - \phi B)\tilde{Z}_t = \alpha_t,$$

y por lo tanto

$$\tilde{Z}_t = (\phi - \theta)\tilde{Z}_{t-1} + \theta(\phi - \theta)\tilde{Z}_{t-2} + \theta^2(\phi - \theta)\tilde{Z}_{t-3} + \dots + \alpha_t,$$

donde se observa que la *FACP*, análoga a la *FAC*, tendrá un valor inicial para $k=1$ que depende de los dos parámetros, θ y ϕ ; y a partir de $k > 1$ el efecto directo de \tilde{Z}_{t-k} sobre \tilde{Z}_t decrecerá geométricamente, con decrecimiento θ .

En la *figura 4.3.1* se muestran algunos patrones que siguen la *FAC* y la *FACP* para un modelo $ARMA(1,1)$.

5. Modelos no estacionarios

5.1 Modelos *ARIMA*

Un proceso integrado de orden d , como se vio en la sección 3.2, es aquel que al diferenciarlo d veces da como resultado un proceso estacionario, es decir, si \tilde{Z}_t es un proceso integrado de orden d entonces el proceso $\{w_t\}$ dado por

$$w_t = \nabla^d \tilde{Z}_t \quad (5.1.1)$$

será estacionario. Al proceso $\{w_t\}$ se le puede aplicar toda la teoría que hasta ahora se ha desarrollado y obtener un modelo *ARMA*(p, q)

$$\phi_p(B)w_t = \theta_q(B)a_t$$

si $\{w_t\}$ es reemplazado por $\nabla^d \tilde{Z}_t$, en la ecuación anterior, se tiene

$$\phi_p(B)\nabla^d \tilde{Z}_t = \theta_q(B)a_t$$

este modelo, será capaz de representar a cierta clase de procesos no-estacionarios y se denomina *autorregresivo e integrado de media móvil*, de forma abreviada modelo *ARIMA*(p, d, q) en donde el parámetro p se refiere al orden del polinomio autorregresivo, d al orden de integración del proceso \tilde{Z}_t y q al orden del polinomio de media móvil. Se denomina integrado debido a que al aplicar ∇^d al proceso \tilde{Z}_t da como resultado una suma infinita (ó integración) de términos w_t ; por ejemplo, supóngase el caso más sencillo

$$w_t = \nabla \tilde{Z}_t$$

el operador inverso de ∇ está definido en términos del operador B mediante la relación: $\nabla = 1 - B$ (ver sección 2.5), entonces

$$\nabla^{-1} = (1 - B)^{-1} = 1 + B + B^2 + \dots$$

de aquí que

$$\tilde{Z}_t = \nabla^{-1} w_t = w_t + w_{t-1} + w_{t-2} + \dots$$

Si se expresa al modelo *ARIMA* sólo en términos del operador de retardo B , se tiene

$$\phi_p(B)(1 - B)^d \tilde{Z}_t = \theta_q(B) \alpha_t$$

en donde a $\phi_p(B)(1 - B)^d$ se le puede considerar como un polinomio de retardo autorregresivo generalizado $\rho(B)$ de orden $p + d$, con d raíces sobre el círculo unitario y las p restantes fuera de él

$$\rho(B) \tilde{Z}_t = \theta_q(B) \alpha_t$$

A pesar de que el operador de retardo autorregresivo tiene varias raíces unitarias, el modelo seguirá siendo estacionario aunque la dependencia entre las variables disminuya muy lentamente con el paso del tiempo.

Si se recuerda, la función de autocorrelación simple de un modelo *ARMA*(p, q), está dada por

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad \text{para } k > q$$

cuya solución (si las raíces de $\phi(B)$ son todas distintas) es:

$$\rho_k = \sum_{i=1}^p A_i G_i^k$$

donde G_i^{-1} $i=1, \dots, p$ son las raíces de $\phi_p(B)=0$ y $G_i < 1$.

Cuando una de las raíces G_i es igual ó muy próxima a la unidad, dígase $G_i \cong 1 - \varepsilon$ con ε muy pequeño, entonces para k grande los términos $A_i G_i^k$ debidos a las otras raíces serán cero ya que G_i^k tiende a cero conforme k aumenta, mientras que el término $A_i G_i^k$ con $G_i = 1 - \varepsilon$ tenderá a A_i conforme ε tienda a cero. De esta manera la *FAC* para k grande seguirá una tendencia lineal en lugar de un rápido decrecimiento geométrico.

Estos resultados se pueden generalizar para el caso en que las raíces de $\phi(B)$ no son todas distintas y cuando más de una de ellas es muy cercana a la unidad.

5.2 Transformaciones estabilizadoras de varianza.

Los procesos integrados pueden resolver el problema de media no-estacionaria sin embargo, en algunos casos, no serán suficientes para resolver los de varianza no-estacionaria. Cuando la varianza no-constante se debe únicamente a que la media presenta una dependencia del tiempo; ambos problemas, media y varianza no-estacionarias, quedaran resueltos mediante un número adecuado de diferenciaciones del proceso. Pero cuando la no-estacionariedad de la varianza persiste a pesar de que el proceso sea estacionario en la media lo más adecuado es utilizar una transformación estabilizadora de la varianza.

Es común que la varianza cambie como cambia la media, presentando una relación del tipo:

$$\text{Var}(Z_t) = cf(\mu_t)$$

para alguna constante c positiva y una función f . Se quiere ahora, encontrar una función T que transforme a la serie Z_t , de tal manera que la serie $T(Z_t)$ tenga varianza constante. Entonces, si $T(Z_t)$ es una función cuya primera derivada existe, puede obtenerse la siguiente aproximación a partir del teorema de Taylor utilizando sólo el primer orden de aproximación alrededor del punto μ_t ,

$$T(Z_t) \cong T(\mu_t) + T'(\mu_t)(Z_t - \mu_t)$$

donde $T'(\mu_t)$ es la primera derivada de $T(Z_t)$ valuada en μ_t .

Ahora

$$\text{Var}(T(Z_t)) \cong (T'(\mu_t))^2 \text{Var}(Z_t) \cong c(T'(\mu_t))^2 f(\mu_t)$$

y ya que se desea que $T(Z_t)$ tenga varianza constante, entonces en la aproximación anterior se debe de tener $\text{Var}[T(Z_t)] \cong c$, para ello:

$$T(\mu_t) = \int \frac{1}{f(\mu_t)} d\mu_t \tag{5.2.1}$$

Lo cual implica que se tiene que conocer muy bien la función f para poder utilizar la transformación estabilizadora de varianza $T(Z_t)$ dada por (5.2.1). Una transformación más utilizada en la práctica, no sólo por estabilizar la varianza sino además por mejorar la aproximación a la serie a una distribución Normal, es la transformación potencia introducida por Box-Cox (1964).

Esto es, si la variable Z_t es positiva y existe una relación entre las desviaciones y las medias de la forma $\sigma_t^2 \cong \mu_t^{2(1-\lambda)}$, entonces

$$T(\mu_t) = \begin{cases} \mu_t^\lambda & \text{si } \lambda \neq 0 \\ \log(\mu_t) & \text{si } \lambda = 0 \end{cases}$$

la cual sugiere que la función T , que vuelve aproximadamente constante a la varianza de Z_t , debe ser la transformación potencia

$$T(Z_t) = \begin{cases} Z_t^\lambda & \text{si } \lambda \neq 0 \\ \log(Z_t) & \text{si } \lambda = 0 \end{cases}$$

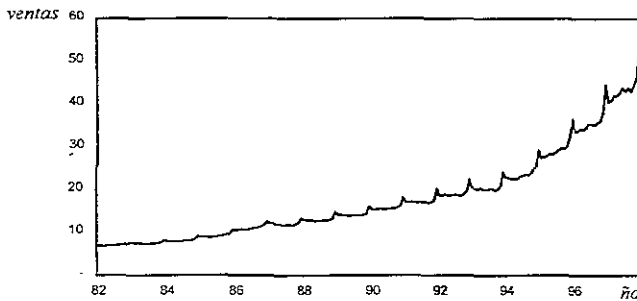
la cual es válida si $Z_t > 0$ para toda t .

5.3 Procesos *ARIMA* estacionales.

Muchas series de tiempo sobre todo en economía se distinguen por presentar un comportamiento igual o casi igual en periodos regulares de tiempo. Por ejemplo, las ventas de juguetes en diciembre tienden a aumentar de manera muy similar a las ventas de los meses de diciembre de otros años, de igual manera que las ventas de enero tendrán un comportamiento muy parecido a las ventas de enero de otros años, etc. La presencia de un fenómeno así, da como resultado una serie de tiempo no-estacionaria, ya que entre otras cosas el valor medio μ_t será distinto para cada período. A este tipo especial de no-estacionariedad se le denomina *estacionalidad* y a la unidad mínima de repetición se le llama *período estacional* s (en el ejemplo de la venta de juguetes el período estacional s es igual a 12 por ser una serie mensual).

Los periodos estacionales pueden ser determinados por distintos factores tales como el clima, el cual afecta actividades como el turismo; eventos gubernamentales como son las elecciones presidenciales; fiestas tradicionales como la Navidad, el día de las madres, año nuevo graduaciones, ... , etc.

Suponga que la *figura 5.1* muestra las ventas mensuales durante el periodo 1982-1997 de una tienda comercial; en la gráfica se pueden apreciar ondas de duración anual con una elevación perteneciente a los meses de diciembre en donde las ventas aumentan debido al festejo de las fiestas navideñas y la percepción de aguinaldos. Posteriormente en la *tabla 5.2* se puede apreciar de manera más clara la similitud que existe entre las ventas para el mismo mes.



Ventas mensuales en millones de pesos

Figura 5.1

Año	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
1982	5 782	6 840	6 751	6 855	6 891	6 835	6 908	7 021	7 045	7 089	7 253	7 600
1983	7 418	7 430	7 415	7 445	7 254	7 275	7 277	7 351	7 322	7 421	7 581	8 104
1984	7 829	7 838	7 796	7 795	7 695	7 826	7 924	8 014	7 994	8 104	8 361	9 123
1985	8 853	8 864	8 857	8 813	8 826	8 948	8 998	9 134	9 191	9 440	9 741	10 656
1986	10 435	10 539	10 724	10 555	10 623	10 693	10 599	11 079	11 102	11 357	11 642	12 438
1987	12 100	12 041	12 024	11 940	11 793	11 546	11 739	11 798	11 628	11 854	12 168	13 284
1988	12 997	12 869	12 832	12 920	12 605	12 730	12 709	12 841	12 738	12 915	13 236	14 738
1989	14 107	14 072	13 984	13 932	13 730	14 041	13 968	13 980	13 907	14 096	14 339	15 924
1990	15 213	15 340	15 335	15 479	15 337	15 317	15 431	15 555	15 513	15 756	16 142	17 996
1991	17 027	17 002	17 123	16 981	16 875	17 046	17 092	17 096	16 853	17 077	17 523	19 953
1992	18 753	18 534	18 895	18 657	18 568	18 594	18 859	18 731	18 475	18 940	19 729	22 056
1993	20 458	20 279	20 004	20 406	19 949	20 013	20 102	20 261	19 796	20 038	20 963	23 877
1994	22 878	22 700	22 524	22 471	22 472	23 026	23 158	23 565	23 358	23 990	25 129	28 948
1995	27 463	27 614	27 526	28 450	28 171	28 679	29 257	29 661	29 478	30 115	32 329	35 944
1996	33 602	33 611	34 133	34 107	34 464	35 449	35 264	35 037	35 636	35 923	37 900	43 863
1997	40 494	40 725	41 927	41 848	42 505	43 749	42 985	43 782	42 989	44 177	46 231	53 220

Ventas mensuales en millones de pesos

Tabla 5.2

Como se mencionó, el fenómeno estacional puede presentarse exactamente igual o casi igual en cada período. En el caso de ser exactamente igual se dirá que hay un componente estacional determinista inmerso en la serie que se denotará por $S_t^{(s)}$, que tiene la característica que $S_t^{(s)} = S_{t-k}^{(s)}$ para $k = \pm 1, \pm 2, \pm 3, \dots$. En este caso la serie Z_t estacional tendrá estructura:

$$Z_t = S_t^{(s)} + \eta_t \quad (5.3.1)$$

en donde η_t es un proceso estacionario. Para convertir esta serie en estacionaria, se definirá el operador de diferencia estacional

$$\nabla_s = 1 - B^s \quad (5.3.2)$$

este operador lo que hace, es tomar diferencias de observaciones separadas s periodos. En el ejemplo de la tienda comercial el operador ∇_{12} tomará la diferencia de las ventas obtenidas en enero del año t con las obtenidas en enero del año $t-1$, las de febrero con las de febrero del año anterior, etc.

Ahora si se aplica el operador ∇_s a (5.3.1), se tiene

$$\nabla_s Z_t = \nabla_s S_t^{(s)} + \nabla_s \eta_t = (1 - B^s) S_t^{(s)} + \nabla_s \eta_t = S_t^{(s)} - S_{t-s}^{(s)} + \nabla_s \eta_t$$

debido a que $S_t^{(s)}$ tiene la característica de repetirse de igual manera cada período s , entonces:

$$\nabla_s Z_t = \nabla_s \eta_t$$

la cual es una serie estacionaria.

En el caso de que el fenómeno estacional no sea exactamente constante, como en la mayoría de los casos, se dice que la serie tiene un componente estacional no determinista del siguiente estilo:

$$S_t^{(s)} = S_{t-k}^{(s)} + v_t^{(s)} + \eta_t$$

en donde $v_t^{(s)}$ es un proceso estacionario.

Al aplicar el operador ∇_s , se tiene:

$$\nabla_s Z_t = \nabla_s S_t^{(s)} + \nabla_s \eta_t = \nabla_s S_{t-k}^{(s)} + \nabla_s v_t^{(s)} + \nabla_s \eta_t = \nabla_s v_t^{(s)} + \nabla_s \eta_t$$

el cual es un proceso estacionario. Por lo tanto el operador ∇_s elimina componentes estacionales tanto deterministas como no deterministas.

En seguida se muestra la gráfica de las observaciones de la **Tabla 5.2** utilizando el operador ∇_s . Obsérvese, que aún el comportamiento de los datos no es estable en la media.

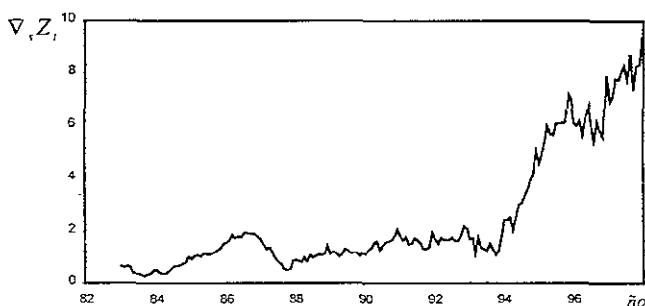


Figura 5.2

Sin pérdida de generalidad se supondrá que el período que se observa es anual, que las observaciones son mensuales ($s=12$) y que se tienen h años de información. Si se

consideran los datos cada 12 observaciones, por ejemplo todos los enero o todos los septiembres, se tendrá una nueva serie que se denotará por $Y_\tau^{(j)}$, $j = 1, \dots, 12$, $\tau = 1, 2, \dots, h$, la cual estará relacionada con la serie original Z_t , mediante

$$Y_\tau^{(j)} = Z_{j+\tau-1} \quad \tau = 1, \dots, h \quad (5.3.4)$$

Se nota de inmediato que si se aplica el operador ∇ a $Y_\tau^{(j)}$ será lo mismo que aplicar ∇ , a la serie original Z_t :

$$\begin{aligned} \nabla Y_\tau^{(j)} &= Y_\tau^{(j)} - Y_{\tau-1}^{(j)} = Z_{j+\tau-1} - Z_{j+\tau-2} = Z_{j+\tau-1} - Z_{j+\tau-2} = \\ &= Z_{j+\tau-1} - B^1 Z_{j+\tau-1} = (1 - B^1) (Z_{j+\tau-1} - Z_{j+\tau-1}) = \nabla_1 Z_{j+\tau-1} \end{aligned}$$

De igual manera que construir un modelo $ARIMA(p, d, q)$ para Y_τ :

$$\phi_p(B) \nabla^d Y_\tau = \theta_0 + \theta_q(B) u_\tau \quad (5.3.5)$$

que un modelo $ARIMA(P, D, Q)$ (modelo autorregresivo de media móvil estacional) para los datos originales, con estructura

$$\Phi_P(B^1) \nabla^D Z_t = \Theta_Q(B^1) \alpha_t \quad (5.3.6)$$

en donde $\Phi_P(B^1)$, $\Theta_Q(B^1)$ son polinomios de retraso de orden P , Q respectivamente y α_t es un proceso de ruido blanco. La relación entre $u_\tau^{(j)}$ y α_t es por supuesto

$$u_\tau^{(j)} = \alpha_{j+\tau-1} \quad j = 1, \dots, h$$

El suponer α , un proceso de ruido blanco implica que es nula la autocorrelación entre α_t y α_{t-k} para cualquier $k = \pm 1, \pm 2, \dots$ y por lo tanto la serie u_t también lo será.

5.4 Modelo multiplicativo estacional

No en todos los casos las series se comportan tan bien como se quisiera, generalmente la serie α_t no es un proceso de ruido blanco, debido a que generalmente las series que por alguna causa quedaron excluidos del modelo tienen también estructura temporal. Por lo cual hay que suponer que la serie α_t sigue un modelo $ARIMA(p, d, q)$

$$\phi_p(B)\nabla^d\alpha_t = \theta_q(B)\alpha_t \tag{5.4.1}$$

en donde α_t si es un proceso de ruido blanco. Sustituyendo (5.4.1) en (5.3.6), se obtiene el modelo general:

$$\Phi_p(B^s)\phi_p(B)\nabla^D\nabla^dZ_t = \Theta_q(B^s)\theta_q(B)\alpha_t$$

que se denominará *modelo multiplicativo estacional* y se denotará por $ARIMA(P, D, Q) \times (p, d, q)^{\otimes}$.

¹ Estos modelos fueron introducidos por Box y Jenkins en 1976.

Capítulo 2

Regresión Dinámica

1. Introducción

En muchas ocasiones se pueden presentar situaciones en donde el impacto que provoca una variable sobre otra no se deja notar simultáneamente, sino que se extiende o distribuye en varios periodos del tiempo. Estos periodos se conocen como rezagos o retardos. Ejemplos de esta situación, se pueden encontrar en muchos campos. En economía, hay una multitud de variables en donde su comportamiento no se debe únicamente a la situación actual que presenten los factores de los cuales depende, sino de la trayectoria que estos han venido presentando durante algunos periodos de tiempo; por ejemplo, ante una devaluación de la moneda, variables como tasas de desempleo, de delincuencia, de emigración, alza en los precios, etc. no se modifican drásticamente, sino que con el paso del tiempo se van viendo afectadas. En medicina, un paciente responde ante un tratamiento o una intervención quirúrgica después de transcurrido un cierto período de tiempo. La inversión en un negocio, por ejemplo, no se refleja de manera inmediata en las utilidades, hay que dejar transcurrir un tiempo para que se empiecen a notar los resultados. En agricultura, la calidad de los productos que se cosechan en un determinado tiempo no depende del clima o del estado de la tierra en ese momento, sino del que tuvieron anteriormente incluso desde antes de

sembrarse. Y así como estos ejemplos, hay una multitud en donde se requiere de un tipo de modelos de regresión que no involucren sólo variables estáticas sino series de tiempo, estos modelos se conocen como *modelos de regresión dinámica*.

Como se verá posteriormente los modelos de regresión dinámicos sólo son una extensión de los modelos de regresión clásica para variables cuyo comportamiento presenta una dependencia con respecto al tiempo.

En la segunda sección de este capítulo se presentan los problemas y alternativas que surgen para plantear un modelo de este estilo así como algunos conceptos básicos, como son medidas y funciones, que sirven para poder darse una idea del grado de dependencia que tiene una variable con respecto a otra y que tanto influye el comportamiento histórico de una sobre otra; posteriormente, en la tercera sección se explican las etapas a seguir para la construcción de un modelo de regresión dinámica, las cuales son: identificación (ver cuarta sección), estimación de parámetros (quinta sección), verificación de los supuestos (sexta sección) y predicción de la variable a partir del modelo construido (séptima sección), después de todo, en la mayoría de los casos, éste es el objetivo principal que lleva a la construcción de un modelo.

En esta última etapa, la predicción, la ventaja de utilizar un modelo dinámico cuando las observaciones sucesivas no son independientes dependerá de sí la relación entre la variable explicativa y la respuesta es o no instantánea. Cuando lo es, para prever Y_{t+k} es necesario conocer X_{t+k} , lo cual implica construir un modelo univariante para X_t , en este caso la mejora en la predicción al utilizar un modelo dinámico quizás no sea notoria. En cambio, cuando el efecto entre la variable explicativa y la respuesta no es instantánea sino que sólo se deja notar transcurridos k periodos, el modelo dinámico será más adecuado, ya que una variación en X_t afectará a Y hasta el instante $t+k$; en este caso, se dice que X_t es un indicador avanzado de la respuesta, y la utilización de un modelo dinámico presentará grandes ventajas.

2. Formulación de modelos de regresión dinámica.

Los modelos de regresión dinámica consideran que el comportamiento de la serie temporal Y_t puede ser aproximado por una relación que involucre de alguna manera la historia de las variables independientes que afecten su comportamiento. Sin embargo, al igual que en un modelo de regresión clásica, no todas las variables independientes pueden ser incluidas en el modelo, ya sea porque algunas de ellas no son conocidas u observables para el investigador o porque el número de variables que intervienen en el modelo sea tan grande que se vuelva muy complicado su manejo adecuado. Por lo cual, la serie temporal Y_t estará descompuesta básicamente en dos componentes; el primero, describirá a través de una función como una o más series explicativas X_t , así como sus valores rezagados, afectan a la variable respuesta Y_t . El segundo, englobará el efecto de todas las demás variables explicativas excluidas del modelo. De esta manera, se tendrá una relación del siguiente estilo:

$$Y_t = Y_t^* + N_t$$

en donde Y_t^* es la parte de la respuesta Y_t explicada por las series X_t y N_t , la no explicada.

A partir de este planteamiento surgen principalmente dos problemas:

- i) La forma analítica de Y_t^* puede ser muy complicada o incluso desconocida, y
- ii) El proceso N_t , denominado proceso de inercia, no puede considerarse un proceso de ruido blanco como en un modelo de regresión clásica debido a que posiblemente algunas de las variables excluidas del modelo tengan estructura temporal que afecten de modo sistemático a la variable respuesta.

Una alternativa, para resolver estos problemas es asumir algunas hipótesis acerca del comportamiento de las variables así como de la relación que hay entre ellas; estas hipótesis se presentarán a continuación.

2.1 Hipótesis básicas

A partir de este momento se considerará una sola serie explicativa X_t , aunque los conceptos se pueden extender para más variables.

Como se mencionó en la sección anterior la función de transferencia, Y_t^* , y el proceso de inercia, N_t , requieren de algunas hipótesis para que su estructura sea manejable, estas hipótesis ó supuestos son las siguientes:

- En primer lugar, se admitirá que entre las variables no hay realimentación; es decir que el comportamiento que tenga X_t puede influir en Y_t , así como en sus valores avanzados Y_{t+k} , $k > 0$, pero no al revés.
- La relación entre X_t y la respuesta Y_t es constante en el período muestral que se está analizando.
- La función Y_t^* que representa el efecto de toda la historia de la variable explicativa X_t puede ser aproximada de forma lineal, de la siguiente manera:

$$Y_t^* = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + v_p X_{t-p}$$

en donde $\{v_i\}$, conocidos como coeficientes de rezago, son constantes desconocidas a determinar y p representa el orden del rezago. En la práctica, usualmente no se tiene información suficiente *a priori* como para saber la longitud máxima del rezago, por lo que es más conveniente ponerlo igual a infinito; esto, en general, no implica alguna distorsión seria ya que los coeficientes de rezago implícitos en el modelo tenderán a ser muy pequeños para valores grandes de rezago. De esta manera el efecto retardado de la variable independiente quedará relacionado con la respuesta de la siguiente manera:

$$Y_t = v_0 X_t + v_1 X_{t-1} + \dots + N_t \quad (2.1.1)$$

➤ El proceso de inercia, N_t , a su vez podrá descomponerse en dos partes como sigue:

$$N_t = f_t + a_t \quad (2.1.2)$$

donde f_t recoge el efecto de todas aquellas variables explicativas que han sido excluidas del modelo pero que sin embargo, su comportamiento afecta de manera sistemático a la variable respuesta, y a_t que englobará una multitud de pequeñas causas sin estructura temporal que causan fluctuaciones aleatorias en la respuesta y por lo que se considerará un proceso de ruido blanco.

De acuerdo a lo anterior, las funciones (2.1.1) y (2.1.2) permiten descomponer a la respuesta en tres términos independientes:

$$Y_t = Y_t^* + f_t + a_t \quad (2.1.3)$$

El primer término de la derecha de la igualdad (2.1.3) representa a las variables incluidas en el modelo; el segundo, el efecto sistemático debido a la estructura temporal de las variables excluidas del modelo y el tercer término, denominado término de perturbación, representa a todas aquellas variables sin estructura temporal que afecten el comportamiento de Y_t .

2.2 Conceptos fundamentales

Función de transferencia y función de respuesta a impulsos

La función que representa el efecto lineal de toda la historia de la serie explicativa X_t sobre la serie respuesta Y_t en estadística se conoce como *función de transferencia*. En términos generales esta función se puede escribir como:

A la representación de los coeficientes de rezago $\{v_i\}$ en función del retardo se le denomina *función de respuesta a impulsos* (ver fig. 2.2) y como se verá más adelante, esta función jugará un papel importante en la identificación del modelo.



Figura 2.2

Función de respuesta a escalones

Ahora, si X se mantiene al mismo nivel a partir del instante t en adelante; esto es, hasta X_{t-1} la variable es cero, en X_t sufre un cambio unitario y a partir de ese momento todas las variables X_{t+k} , $k > 0$ toman el valor igual a uno, entonces Y_t responderá de la siguiente manera:

$$\begin{aligned} Y_t^* &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots = v_0 \\ Y_{t+1}^* &= v_0 \cdot 1 + v_1 \cdot 1 + v_2 \cdot 0 + \dots = v_0 + v_1 \\ Y_{t+2}^* &= v_0 \cdot 1 + v_1 \cdot 1 + v_2 \cdot 1 + \dots = v_0 + v_1 + v_2 \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \end{aligned}$$

En el mismo período de tiempo, v_0 corresponde al cambio en Y ocasionado por X , $v_0 + v_1$ al cambio ocasionado en el siguiente período, $v_0 + v_1 + v_2$ al siguiente período, y así sucesivamente. Este tipo de incremento unitario sostenido se conoce como *aumento escalón* y a la función que describe la variación de Y_t^* cuando X aumenta en escalón en t se le denomina función de respuesta a escalones (*fre*).

$$fre(t+k) = \sum_{i=0}^k v_i$$

Ganancia

Un concepto importante es el de la ganancia, ésta se define como el valor límite de la función de respuesta a escalones, es decir

$$G = \text{Ganancia} = \sum_{i=0}^{\infty} v_i$$

Se dice que la función de transferencia es *estable*, si un aumento finito en X , produce un efecto finito en Y ; lo cual implica que la ganancia sea finita o bien que $\sum_0^{\infty} v_i < \infty$.

La importancia de que una función de transferencia sea estable tiene los mismos motivos que se vieron en el capítulo anterior, del porque una serie de tiempo debe de ser estacionaria.

Retardo Medio

Cuando los coeficientes de rezago $\{v_j\}$ son todos positivos, la rapidez de la transmisión de los efectos suele describirse mediante el retardo medio. El retardo medio, τ , no es más que un promedio ponderado de los rezagos por el efecto que transmiten, por lo que se obtiene de las sumas parciales de las v_j estandarizadas.

Es decir

$$\tau = \sum_{j=1}^{\infty} \frac{v_j}{G} \cdot j$$

Por ejemplo, si $\tau = 1$, lo que sugiere este tipo de medida “resumen” es que la mitad del impacto de un cambio en la variable X se sentirá en la variable dependiente Y durante el primer período, si en cambio $\tau = 10$, el efecto de cualquier cambio en X tomará bastante tiempo para impacte a Y , ya que tan sólo la mitad de éste se sentirá después de transcurridos 10 periodos de tiempo.

2.3 Modelización de la función de transferencia

Utilizando el operador del retardo B , definido como en el capítulo anterior, se tiene

$$Y_t^* = (v_0 + v_1 B + v_2 B^2 + \dots) X_t = v(B) X_t$$

a $v(B) = v_0 + v_1 B + v_2 B^2 + \dots$ se le conoce como polinomio de rezago de orden infinito. Un polinomio de orden infinito, desde un punto de vista técnico, está lejos de ser fácilmente manejable. Sin embargo, se puede hacer una analogía con lo estudiado en el capítulo anterior en donde, se utilizaron aproximaciones de orden finito para la representación general, $\psi(B)$, de un proceso estacionario. Estas aproximaciones son: las de media móvil y las del tipo autorregresivo. En donde, en las primeras se tiene un polinomio de rezago $\theta_q(B)$ de orden finito cuyos componentes tienen igual estructura que $\psi(B)$, y en las segundas un polinomio autorregresivo $\phi_p(B)$ cuyas raíces se encuentran fuera del círculo unitario. De igual manera el polinomio de rezago $v(B)$ puede aproximarse por un polinomio con igual estructura pero con un número finito de componentes, es decir:

$$w_0 + w_1 B + w_2 B^2 + \dots + w_m B^m = w_m(B) \cong v(B)$$

utilizando las letras w para indicar un número finito de parámetros.

Asimismo, se puede admitir una representación del tipo autorregresivo de orden infinito pero con estructura simple de amortiguamiento entre los coeficientes, de la forma siguiente:

$$v(B) \cong w_0 (1 + \delta B + \delta^2 B^2 + \dots) = w_0 \lim_{n \rightarrow \infty} \left[\frac{1 - \delta^n B}{1 - \delta B} \right] = \frac{w_0}{(1 - \delta B)}$$

En este último caso se ha utilizado la estructura más simple de amortiguamiento (decrecimiento de los coeficientes en forma geométrica).

De manera más general, así como el polinomio de orden infinito $\psi(B)$, utilizado en el análisis de series temporales, puede ser aproximado mediante el cociente de dos polinomios de orden finito como:

$$\psi(B) \cong \frac{\theta_q(B)}{\phi_p(B)},$$

de manera análoga $v(B)$ puede aproximarse como:

$$v(B) \cong \frac{w_m(B)}{\delta_a(B)}$$

en donde $w_m(B) = w_0 + w_1B + w_2B^2 + \dots + w_mB^m$ y $\delta_a(B) = 1 - \delta_1B - \delta_2B^2 - \dots - \delta_aB^a$ son dos polinomios de rezago de orden finito.

Para que el sistema sea estable se requerirá de que las raíces del polinomio $\delta_a(B) = 0$ estén fuera del círculo unitario. Aunque a diferencia de un proceso *ARMA* para series de tiempo, en estos modelos el polinomio $w_m(B) = w_0 + w_1B + w_2B^2 + \dots + w_mB^m$ comienza con w_0 en lugar de uno, además de que los términos se presentan con signo positivo, como es usual en los modelos de regresión clásicos, y no hay restricciones acerca de sus raíces.

Debido a que en algunas situaciones la respuesta Y no responde inmediatamente ante un cambio en X , se anexará un tercer término, B^b , en donde b indica el número de coeficientes v_t que son cero antes de que el impacto sobre la variable respuesta se deje sentir, quedando:

$$v(B) = \frac{w_m(B)}{\delta_a(B)} B^b \tag{2.3.1}$$

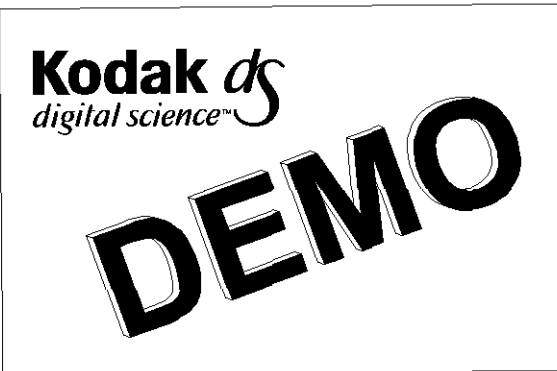
De esta manera, la función de transferencia queda caracterizada por tres parámetros: m , que indica el orden de polinomio de media móvil, a el orden del polinomio autorregresivo y b que indica el número de periodos de rezago que tarda en empezar a producir el efecto de un cambio en la variable dependiente; si $b = 0$ entonces la respuesta será inmediata.

2.4 Función de respuesta a impulsos.

Ya se pudo observar en (2.3.1) que existe una gran similitud entre una función de transferencia (m, a, b) y un proceso $ARMA(p, q)$ cuando $m = q$, $a = p$, y sobre todo si $b = 0$. Ahora, se demostrará que también la función de autocorrelación para un proceso $ARMA(p, q)$ tiene su análogo en el análisis de la función de transferencia, se está hablando de la función de respuesta a impulsos, la cual ya fue introducida al principio de este capítulo.

Si se recuerda, la FAC° de un modelo $ARMA(p, q)$, cuando $q > p$, tiene los primeros q componentes sin estructura y a partir de $q + 1$ decrece en forma exponencial y/o senoidal,

debido
observ
ejemp
repres



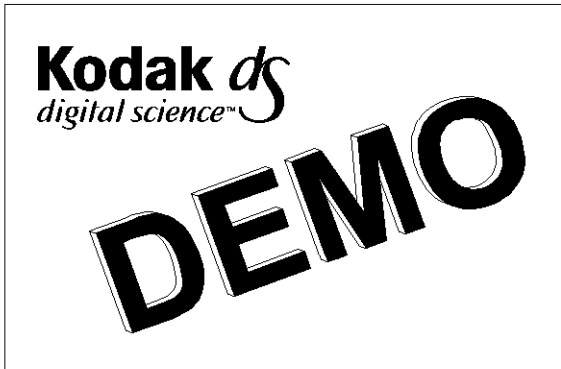
esta e:

va. A continuación se podrá
impulsos es muy similar. Por
función de transferencia está
o, es decir

$$\delta(B)v(B)X_t = w_0 B^b X_t$$

$$(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_a B^a)(v_0 + v_1 B + v_2 B^2 + \dots) = w_0 B^b$$

es decir.



$$+ v_1 B(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_a B^a) + \dots$$

$$\delta^2 - \dots - \delta_a B^a) + \dots = w_0 B^b$$

$$v_2 = \delta_1 v_1 + \delta_2 v_0$$

$$\vdots$$

$$v_h = \delta_1 v_{h-1} + \delta_2 v_{h-2} + \dots + \delta_h v_{h-h} + w_0 \quad h = \max\{b, a\}$$

$$\vdots$$

De manera que, la función de respuesta a impulsos tendrá la siguiente estructura:

$$v_j = 0 \quad j < b$$

$$v_j = w_0 \quad j = b$$

$$v_j = \delta_1 v_{j-1} + \dots + \delta_h v_{j-h} \quad j > b, \quad h = \max\{j, a\}$$

en donde, se puede observar fácilmente que la función de respuesta a impulsos tendrá una estructura que decrecerá a partir del retardo $j = b + 1$ en forma exponencial y/o senoidal, no dejando duda de que su comportamiento es muy similar al de una FAC para un proceso $AR(p)$.

Si la función de transferencia tiene ahora, sólo componentes de media móvil:

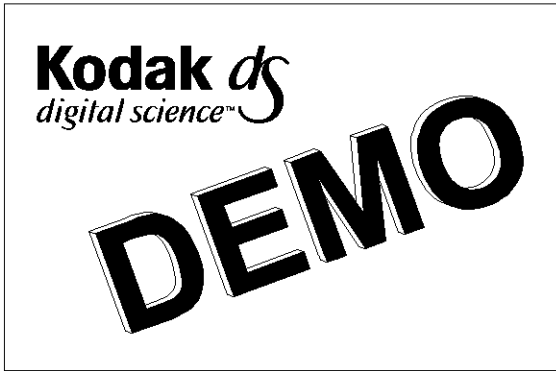
¹ Función de autocorrelación

$$Y_t^* = v(B)X_t = w(B)B^h X_t$$

entonces

$$(v_0 + v_1 B + v_2 B^2 + \dots) = (w_0 B^h + w_1 B^{h+1} + w_2 B^{h+2} + \dots + w_m B^{h+m})$$

sigui
llega
sólo



B en la ecuación anterior se
ción de transferencia que tenga

$$v_j = 0$$

$$j > b + m$$

lo cual era de esperarse que fuera finita debido a la analogía que se ha venido observando con una *FAC*.

Ahora, si se considera a la función de respuesta a impulsos de manera más general, es decir, con componentes autorregresivos y de media móvil, se tiene

$$v(B)X_t = \frac{w(B)}{\delta(B)} B^h X_t$$

$$\delta(B)v(B)X_t = w(B)B^h X_t$$

es decir:

$$(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_a B^a)(v_0 + v_1 B + v_2 B^2 + \dots) = (w_0 + w_1 B + \dots + w_m B^m) B^b$$

igualando coeficientes en B se obtiene la solución general:

$$v_j = 0 \quad j < b \quad (2.4.1)$$

$$v_j = \delta_1 v_{j-1} + \dots + \delta_h v_{j-h} + w_{j-b} \quad j = b, b+1, \dots, b+m, \quad h = \max\{j, a\} \quad (2.4.2)$$

$$v_j = \delta_1 v_{j-1} + \dots + \delta_h v_{j-h} \quad j > b+m, \quad h = \max\{j, a\} \quad (2.4.3)$$

En conclusión, la función de respuesta a impulsos tiene la siguiente estructura:

- i) los primeros b coeficientes son iguales a cero,
- ii) a partir de b hasta $b+m$, no presenta ningún patrón de comportamiento, y
- iii) a partir de $b+m+1$ seguirá un comportamiento determinado únicamente por la parte autorregresiva de la función de transferencia, ya sea que decaiga en forma exponencial y/o en ondas senoidales.

Al igual que la *FAC* es una herramienta importante en la identificación del orden de los polinomios de rezago involucrados en un modelo *ARMA*, también la función de respuesta a impulsos jugará un papel importante en la identificación de la función de transferencia.

2.5 Modelización del Proceso de Inercia.

El proceso de inercia, como ya se había mencionado, representa a todas aquellas variables que por diversas circunstancias no han sido incluidas en el modelo, está dividido en dos componentes:

- el primero, f_t , representa a todas las variables excluidas que afecten significativamente a la variable respuesta Y_t (en los modelos de regresión clásica, se supone que todas estas variables han sido incluidas en el modelo por lo tanto éste componente se supone nulo),
- el segundo término, a_t , denominado *término de perturbancia* encierra a todos aquellos factores sin estructura que de alguna manera afectan el comportamiento de Y_t .

De acuerdo con lo anterior, es claro que el proceso de inercia no se puede suponer ruido blanco ya que contiene variables que al igual que la variable explicativa incluida en el modelo puede tener comportamientos rezagados que afecten temporalmente a la respuesta, por lo que se supondrá que este modelo puede ser modelado por un proceso $ARIMA(p, d, q)$:

$$\phi_p(B)\nabla^d N_t = \theta_q(B)\alpha_t \tag{2.5.1}$$

Es así como, la serie Y_t puede representarse por medio de la siguiente expresión:

$$Y_t = Y_t^* + N_t = \frac{w_m(B)B^b}{\delta_a(B)} X_t + \frac{\theta_q(B)}{\phi_p(B)\nabla^d} \alpha_t \tag{2.5.2}$$

2.6 Cointegración

Hasta este momento, para la construcción del modelo de regresión dinámica, se ha supuesto que la serie explicativa X_t es estacionaria, por lo cual el efecto que provoca sobre la serie respuesta ha podido ser representado por un modelo $ARMA$; sin embargo, como se vio en el capítulo anterior no todas las series pueden ser estacionarias de primer instancia, hay algunas que tienen que ser integradas para que puedan cumplir con las características de estacionariedad, o lo que es igual, se deben aceptar algunas raíces unitarias en $\delta_a(B)=0$, teniendo así, un proceso que no es estacionario, (o integrado de orden cero, $I(0)$), pero que tampoco crece de manera explosiva. En estos casos, ¿qué pasa con la serie respuesta, Y_t ?, hereda las propiedades de la función de transferencia, del proceso de inercia, o una combinación de ambos.

Cuando la función de transferencia es estacionaria y el proceso de inercia también lo es, se respetan las reglas de adición de procesos estocásticos; debido a que la media de un proceso que es el resultado de una combinación lineal de procesos estacionarios, seguirá siendo constante, la varianza constante y la correlación seguirá dependiendo únicamente del retardo. Así, si $I(0)$ representa a un proceso estacionario, se verifica que

$$\alpha_1 I(0) + \alpha_2 I(0) = I(0) \quad (2.6.1)$$

lo cual implica que, si se tienen dos procesos que para ser estacionarios han tenido que ser diferenciados h veces, cualquier combinación lineal de ellos tendrá que ser diferenciada un número igual de veces para que también lo sea, en algunos casos la combinación lineal tiene que ser diferenciada un número menor de veces. Así, cuando se da la siguiente situación:

$$\alpha_1 I(h) + \alpha_2 I(h) = I(l \leq h) \quad (2.6.2)$$

se dice que el proceso uno y el proceso dos son *cointegrados*.

De esta manera, si el proceso de inercia se ha supuesto estacionario y la serie explicativa también, entonces la serie respuesta tendrá que serlo también. En caso de no ser así hay que revisar que el proceso de inercia cumpla verdaderamente con las características de estacionariedad.

El concepto de cointegración tiene mucha importancia en el estudio de series de tiempo, sobre todo cuando se quiere determinar la relación que existe entre dos series de tiempo a largo plazo, ya que con el paso del tiempo puede suceder que dos series que en un momento dado estuvieron muy relacionadas en un futuro ninguna de ellas tenga que ver nada con la otra.

Supóngase que $y_t \sim I(1)$ y $x_t \sim I(1)$. Entonces, se dice que y_t y x_t son cointegradas si existe un α tal que la serie $\{y_t - \alpha x_t\}$ sea $I(0)$. Esto, se denota diciendo que y_t y x_t son $C(1,1)$. Lo cual significa que la siguiente ecuación de regresión

$$y_t = \alpha x_t + \eta_t \quad (2.6.3)$$

tiene sentido, pues y_t y x_t no se desvían mucho entre sí con el paso del tiempo, es decir, que, existe una relación de equilibrio a largo plazo entre ellas. En el caso contrario, es decir, si $\{y_t - \alpha x_t\}$ no es estacionario, el comportamiento de cada una de ellas estará cada vez más desviado el uno del otro al paso del tiempo, no existiendo ninguna relación de equilibrio a largo plazo entre ellas.

En el capítulo anterior, en donde se analizó a las series de tiempo univariantes, se vio que si la serie de tiempo es no estacionaria, había que diferenciarla hasta lograr estacionariedad y después poder ajustar algún modelo *ARMA*. Cuando se consideran dos series de tiempo, y_t y x_t , se hace lo mismo, sin embargo esta operación, de diferenciación, elimina la tendencia o movimiento a largo plazo. Por lo cual, antes de cualquier diferenciación se debe de ver cual es la relación a largo plazo entre las tendencias y_t y x_t , es decir, si y_t y x_t están cointegradas.

El caso con el ajuste estacional es similar. En lugar de eliminar las componentes estacionales de y_t y de x_t , y después analizar los datos desestacionalizados, se preguntará primero si existe una relación estacional entre y_t y x_t .

3. Construcción de un modelo de regresión dinámica

El enfoque que aquí se estudiará para la construcción de un modelo de regresión dinámica es el sugerido por Box y Jenkins, el cual está basado en las siguientes cuatro etapas:

- i) Identificación del modelo
- ii) Estimación de parámetros
- iii) Verificación de los supuestos
- iv) Uso del modelo

Estas cuatro etapas básicas para la construcción de un modelo se pueden apreciar de manera más clara en el diagrama que se muestra en la *figura 3.1*

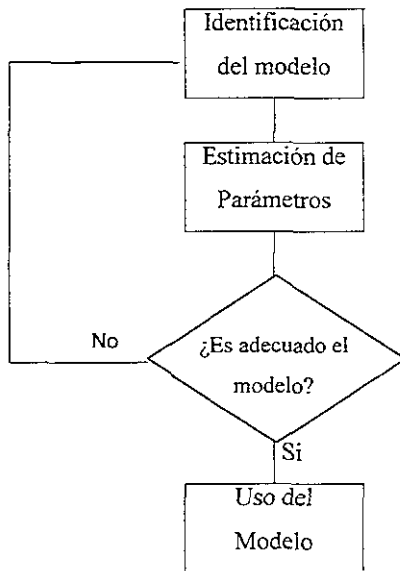


Figura 3.1

Los modelos $ARMA(p, q)$:

$$\phi_p(B)Z_t = \theta_q(B)\alpha_t \tag{3.1.1}$$

para series de tiempo univariadas Z_t , estudiados en el capítulo anterior, son casos especiales de modelos de regresión dinámica si se considera al proceso Z_t como a la serie respuesta y al proceso de ruido blanco, a_t , como a la serie explicativa.

De tal manera que el modelo (3.1.1) puede ser reescrito como:

$$\tilde{Z}_t = \frac{\theta_q(B)}{\phi_p(B)} a_t \quad (3.1.2)$$

En donde si se utiliza la formulación de modelos de regresión dinámica, la función de transferencia estará dada por:

$$v(B) = \frac{\theta_q(B)}{\phi_p(B)}$$

y el proceso de inercia será nulo.

Debido a esto, se comenzará a explicar cada una de las etapas para este tipo de modelos y posteriormente una vez comprendidos los procedimientos, se hará la generalización a los modelos de regresión dinámicos.

4. Identificación del modelo

Esta primera etapa se puede considerar la más importante ya que de ella depende en gran parte el éxito que se logre en las siguientes, pero también la más difícil ya que se requiere de una buena comprensión de los modelos que se han venido discutiendo y de la práctica que se haya adquirido en su aplicación.

4.1 Identificación del modelo

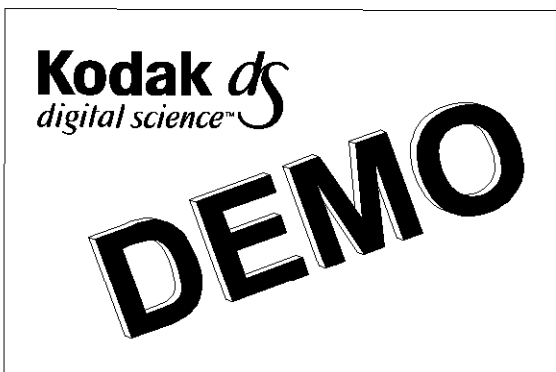
La identificación de un modelo $ARMA(p, q)$ para una serie de tiempo univariante y_t , consistirá básicamente de los siguientes dos pasos:

1. Escoger, si es necesario, las transformaciones adecuadas para obtener una serie estacionaria a partir de la original.

Una herramienta importante para tomar una decisión acerca de las transformaciones que se requieren aplicar para estacionarizar a la serie es la gráfica de los datos originales, ya que a través de ella se puede detectar o tener alguna idea sobre si la serie presenta alguna

fenómeno no-estacionario. Una transformación y decir que sean la

Otra herramienta importante que como se vio en el capítulo anterior, puede ser utilizada rápidamente a caso de que se presente el orden de difer



o algún otro fenómeno no-estacionario. Esto no quiere decir que sea necesario

le la serie, ya que la función de autocorrelación decaerá rápidamente a cero. En el caso de que se presente un patrón de autocorrelación no estacionario para encontrar el orden de integración

la serie original y de sus diferencias $\nabla Z_t, \nabla^2 Z_t, \dots$, hasta que la gráfica muestre un decaimiento rápido a cero. En caso dudoso es recomendable diferenciar una vez más, ya que las consecuencias de una diferenciación innecesaria se considera menos grave que la subdiferenciación.

2. Determinar los parámetros p y q del modelo $ARMA$ para la serie estacionaria obtenida.

La FAC^{\circledast} y la $FACP^{\circledast}$ jugarán un papel importante en la identificación de p y q , ya que como se discutió anteriormente la $FACP$ de un modelo $AR(p)$ es cero después de p rezagos y la FAC de un modelo $MA(q)$ después de q rezagos.

¹ Función de autocorrelación

² Función de autocorrelación parcial

Desgraciadamente en la práctica estas funciones son desconocidas, por lo tanto el procedimiento para identificar los valores p y q consistirá en graficar la FAC muestral y la $FACP$ muestral de la serie y compararlas con las funciones teóricas. El cálculo de estas funciones muestrales se explicará posteriormente en la siguiente sección.

La identificación de un modelo con estructura $ARMA$ es un poco más complicada, por lo cual se recomienda empezar a suponer modelos AR y MA de orden bajo e ir modificándolos de acuerdo a las necesidades que se tengan que ir cubriendo.

En conclusión para la identificación de modelos hay que tener muy claras las características de la FAC y de la $FACP$, las cuales se resumen en el cuadro (4.1)

Cuadro 4.1

Comportamiento de la FAC y de la $FACP$ para procesos AR , MA y $ARMA$.

Proceso	FAC	$FACP$
$AR(p)$	Convergencia a cero en forma exponencial y/o senoidal dependiendo de la ecuación $\phi(B)\rho_k = 0$, para $k \geq p$.	Primeras p autocorrelaciones distintas de cero y el resto nulos.
$MA(q)$	Primeras q autocorrelaciones distintas de cero y el resto nulos.	Coefficientes no nulos que decrecen a cero conforme el retardo en una mezcla de exponenciales y senoides.
$ARMA(p, q)$	Comportamiento irregular de las primeras q autocorrelaciones y después convergencia a cero con un comportamiento dictado por $\phi(B)\rho_k = 0$, para $k > q$.	Sucesión infinita convergente a cero.

Es importante, además tener en cuenta que estos métodos de identificación sólo ayudan a reducir el número de modelos factibles, ya que el mejor modelo se logra a través de la experiencia, la intuición y la constancia.

4.1.1 Función de autocorrelación muestral y función de autocorrelación parcial muestral.

Para obtener la *FAC* muestral se calculan primero las autocovarianzas muestrales. En cada instante del tiempo, en la mayoría de los casos en cada instante de tiempo sólo se cuenta con una observación; por lo tanto para estimar γ_k se utilizará el promedio de la autocovarianza entre Z_t y Z_{t+k} para $t = 1, \dots, N - k$, quedando :

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) \tag{4.1.1.1}$$

de ahí que

$$r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}, \quad k = 0, 1, 2, \dots \tag{4.1.1.2}$$

en donde $\bar{y} = \sum_{t=1}^N \frac{y_t}{N}$ representa a la media muestral de la serie.

Debido a que la estimación está expuesta a errores muestrales habrá que buscar un criterio para decidir si r_k puede o no ser considerado como cero. Un criterio lo proporcionó Bartlett (1946)⁹ al demostrar que en el caso de que el proceso tenga una distribución normal y suponiendo que sólo las primeras h autocorrelaciones teóricas son distintas de cero, se tiene la siguiente aproximación

⁹ Bartlett, M. S. (1946). *On the theoretical specification of sampling properties of autocorrelated time series* Journal of Royal Statistical Society (Suppl.), 7, 211.

$$Var(r_k) \cong \frac{1}{N} (1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_h^2) \quad k > h \quad (4.1.1.3)$$

en la práctica, esta fórmula se utiliza sustituyendo ρ_i por r_i , $i = 1, \dots, h$.

Entonces, la distribución de la función de autocorrelación r_k puede aproximarse por una distribución normal con media cero y varianza (4.1.1.3); por lo cual los valores ± 1.96 (usualmente se tomará ± 2) multiplicados por la desviación estándar constituyen los límites para determinar observaciones significativamente distintas de cero al nivel de significancia del 5%.

Es así, como se considera a r_k igual a cero si

$$|r_k| \leq 2 \sqrt{\frac{1}{N} \left(1 + 2 \sum_{j=1}^q r_j^2 \right)} \quad \text{para } k > q.$$

Los coeficientes de autocorrelación parcial muestrales se obtienen de sustituir r_i en:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}} \quad \textcircled{*}$$

quedando el estimador de la función de autocorrelación parcial, *FACP*, como:

* La expresión que relaciona a ρ_k con ϕ_{kk} se obtuvo en la sección 4.1.1 del primer capítulo.

$$\hat{\phi}_{kk} = \frac{\begin{matrix} 1 & r_1 & r_2 & \dots & r_{k-2} & r_1 \\ r_1 & 1 & r_1 & \dots & r_{k-3} & r_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \end{matrix}}{\begin{matrix} r_{k-1} & r_{k-2} & r_{k-3} & \dots & r_1 & r_k \\ 1 & r_1 & r_2 & \dots & r_{k-2} & r_{k-1} \\ r_1 & 1 & r_1 & \dots & r_{k-3} & r_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \dots & r_1 & 1 \end{matrix}} \quad (4.1.1.4)$$

4.2 Identificación de un modelo de regresión dinámica.

Para la identificación de modelos de regresión dinámica se seguirán básicamente los mismos pasos, descritos con anterioridad. Sin embargo, como es de esperarse, las herramientas que se han venido utilizando tendrán que ser generalizadas para poder ser utilizadas en estos modelos de regresión.

4.2.1 Función de correlación cruzada

Hasta ahora, se ha supuesto que a partir del conocimiento de la serie X_t , puede predecirse el comportamiento de la serie Y_t , pero ¿qué tan relacionadas están estas dos series cuando a partir de una puede predecirse la otra?. Una herramienta importante en el análisis de regresión para poder determinar que tanta relación existe entre una variable y otra, es el *coeficiente de correlación*. En regresión clásica, este coeficiente sólo se utiliza entre variables observadas en el mismo instante de tiempo, pero ahora, en el análisis de regresión dinámica, se utilizará entre variables rezagas k periodos de tiempo para poder conocer la relación entre las variables no sólo en el mismo instante sino separadas por k periodos.

Esta extensión del coeficiente de correlación para diferentes valores de k se denomina *función de correlación cruzada (FCC)*, y se define como:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \quad (4.2.1.1)$$

en donde $\gamma_{xy}(k)$ es la *función de covarianza cruzada* y, σ_x y σ_y denotan la desviación estándar de X_t y Y_t , respectivamente.

La función de covarianzas cruzadas a su vez, está definida como:

$$\gamma_{xy}(k) = E\{(x_t - \mu_x)(y_{t+k} - \mu_y)\} \quad k = 0, \pm 1, \pm 2, \dots$$

Como puede observarse, $\gamma_{xy}(k)$ es la que en realidad mide la relación que hay entre X_t y Y_t , pero en la práctica es más común utilizar su estandarización (4.2.1.1) debido a que su interpretación es más fácil.

Aplicando la función de covarianzas cruzada (FCC) entre la misma serie y utilizando $k = 0$ se obtienen las desviaciones estándares de las series X_t y Y_t , respectivamente:

$$\gamma_{xx}(0) = E\{(x_t - \mu_x)^2\} = \sigma_x^2 \tag{4.2.1.2}$$

$$\gamma_{yy}(0) = E\{(y_t - \mu_y)^2\} = \sigma_y^2$$

A partir de esto, se dice que dos procesos (o series) son conjuntamente estacionarios si:

- i) cada proceso puede transformarse de tal manera que sean estacionarios, independientemente uno del otro, y
- ii) la correlación cruzada, $\rho_{xy}(k)$, depende únicamente del rezago k .

Si esto se cumple, $\rho_{xy}(k)$ tendrá además las siguientes propiedades:

- i) $\rho_{xy}(k) = \rho_{yx}(-k)$.

Esta propiedad puede observarse claramente si se escribe

$$\begin{aligned} \rho_{xy}(k) &= \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} = \frac{E\{(x_t - \mu_x)(y_{t+k} - \mu_y)\}}{\sigma_x \sigma_y} = \\ &= \frac{E\{(y_{t+k} - \mu_y)(x_t - \mu_x)\}}{\sigma_y \sigma_x} = \frac{\gamma_{yx}(-k)}{\sigma_y \sigma_x} = \rho_{yx}(-k) \end{aligned}$$

- ii) para $k > 0$, $\rho_{xy}(k)$ representa la influencia que tiene x sobre y .
- iii) para $k < 0$, $\rho_{xy}(k)$ representa la influencia que tiene y sobre x .
- iv) la función $\rho_{xy}(k)$ en general no será simétrica, es decir, $\rho_{xy}(k) \neq \rho_{xy}(-k)$

Estas últimas tres propiedades (ii, iii, iv) son consecuencia de la propiedad (i), y de ellas se puede concluir que la *función de correlación cruzada* no sólo mide la fuerza sino también la dirección con que una variable puede influir sobre otra.

Debido a que una de las hipótesis sobre las que se está basando la construcción de los modelos de regresión dinámica, es que entre las variables no hay realimentación, es decir, que el comportamiento de X_t puede influir sobre Y_{t+k} para todo $k > 0$ pero no al revés, se espera que para $k > 0$ la función ρ_{xy} sea mayor que cero y que para $k < 0$, $\rho_{xy}(k)$ sea idénticamente cero. Por lo cual es importante examinar la FCC, $\rho_{xy}(k)$, para ambos valores de rezagos positivos $k > 0$ y negativos $k < 0$.

4.2.2 La relación entre la función de correlación cruzada y la función de respuesta a impulsos.

Es lógico pensar que hay una fuerte relación entre la función de correlación cruzada (FCC) y la función de respuesta a impulsos debido a que ambas miden la dependencia lineal que hay entre una serie y otra. Para ver cual es la relación entre ellas, supóngase que dos procesos estacionarios están relacionados de la siguiente manera:

$$Y_{t+k} = v(B)X_{t+k} + \eta_{t+k} \tag{4.2.2.1}$$

en donde η_t también es un proceso estacionario, y sin pérdida de generalidad supóngase, además, que $\mu_x = 0$ y $\mu_y = 0$.

Multiplicar de ambos lados de la ecuación (4.2.2.1) por X_t ,

$$X_t Y_{t-k} = v(B)X_t X_{t+k} + X_t \eta_{t-k}$$

expandir el polinomio $v(B)$

$$X_t Y_{t-k} = (v_0 + v_1 B + v_2 B^2 + \dots)X_t X_{t+k} + X_t \eta_{t-k}$$

desarrollar el operador de rezago

$$X_t Y_{t-k} = v_0 X_t X_{t+k} + v_1 X_t X_{t+k-1} + v_2 X_t X_{t+k-2} + \dots + X_t \eta_{t-k}$$

y al aplicar el operador esperanza, se tiene

$$E(X_t Y_{t+k}) = v_0 E(X_t X_{t+k}) + v_1 E(X_t X_{t+k-1}) + v_2 E(X_t X_{t+k-2}) + \dots + E(X_t \eta_{t+k})$$

obsérvese que:

$$\gamma_{XY}(k) = E[(X_t - \mu_X)(Y_{t+k} - \mu_Y)] = E(X_t Y_{t+k}) - \mu_X \mu_Y, \text{ como } \mu_X = 0 \text{ y } \mu_Y = 0, \\ \text{entonces } \gamma_{XY}(k) = E(X_t Y_{t+k}).$$

así mismo:

$$\gamma_{XX}(k) = E[(X_t - \mu_X)(X_{t+k} - \mu_X)] = E(X_t X_{t+k}) - \mu_X^2 \\ \gamma_{XX}(k-1) = E[(X_t - \mu_X)(X_{t+k-1} - \mu_X)] = E(X_t X_{t+k-1}) - \mu_X^2 \\ \vdots$$

como $\mu_1 = 0$ entonces:

$$\gamma_{XX}(k) = E[X_t X_{t+k}]$$

$$\gamma_{XX}(k-1) = E[X_t X_{t-1}]$$

⋮

se puede ver entonces, que:

$$\gamma_{XY}(k) = v_0 \gamma_{XY}(k) + v_1 \gamma_{XY}(k-1) + v_2 \gamma_{XY}(k-2) + \dots \quad (4.2.2.2)$$

y, el coeficiente de correlación cruzada

$$\begin{aligned} \rho_{XY}(k) &= \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y} = \frac{v_0 \gamma_{XY}(k) + v_1 \gamma_{XY}(k-1) + v_2 \gamma_{XY}(k-2) + \dots}{\sigma_X \sigma_Y} = \\ &= \frac{v_0 \sigma_X \rho_{XX}(k) + v_1 \sigma_X \rho_{XX}(k-1) + v_2 \sigma_X \rho_{XX}(k-2) + \dots}{\sigma_Y} \end{aligned}$$

por lo tanto

$$\rho_{XY}(k) = \frac{\sigma_X}{\sigma_Y} [v_0 \rho_{XX}(k) + v_1 \rho_{XX}(k-1) + v_2 \rho_{XX}(k-2) + \dots] \quad (4.2.2.3)$$

Como puede observarse la relación entre la función de correlación cruzada y los coeficientes de rezago v_j , está contaminada por la estructura de autocorrelación de la serie X_t . Sin embargo, si el número de coeficientes de rezago v_j , en (4.2.2.1) es finito, los coeficientes v_j , se pueden estimar a partir de un sistema de ecuaciones, ya sea de (4.2.2.2) o de (4.2.2.3).

Por ejemplo, si la función de transferencia está representada por la forma racional:

$$v(B) = \frac{w_m(B)}{\delta_a(B)} B^b \quad (4.2.2.4)$$

con $a = 0$ y $b = 0$, entonces $v(B)$ es una función con un número de coeficientes de rezago finito. Utilizando (4.2.2.2) y particularizando para $k = 0, 1, \dots, m-1$ se obtiene el siguiente sistema de ecuaciones

$$\begin{aligned} \gamma_{xy}(0) &= v_0 \gamma_{xx}(0) + v_1 \gamma_{xx}(1) + \dots + v_m \gamma_{xx}(m) \\ \gamma_{xy}(1) &= v_0 \gamma_{xx}(1) + v_1 \gamma_{xx}(0) + \dots + v_m \gamma_{xx}(m-1) \\ &\vdots \\ \gamma_{xy}(m) &= v_0 \gamma_{xx}(m) + v_1 \gamma_{xx}(m-1) + \dots + v_m \gamma_{xx}(0) \end{aligned}$$

el cual puede escribirse matricialmente como

$$\begin{pmatrix} \gamma_{xy}(0) \\ \gamma_{xy}(1) \\ \vdots \\ \gamma_{xy}(m) \end{pmatrix} = \begin{pmatrix} \gamma_{xx}(0) & \gamma_{xx}(1) & \dots & \gamma_{xx}(m) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \dots & \gamma_{xx}(m-1) \\ \vdots & \vdots & & \vdots \\ \gamma_{xx}(m) & \gamma_{xx}(m-1) & & \gamma_{xx}(0) \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_m \end{pmatrix}$$

es decir,

$$\underline{\gamma}_{xy} = \Gamma_{xx} V$$

en donde

$$\gamma_{xy} = \begin{pmatrix} \gamma_{xy}(0) \\ \gamma_{xy}(1) \\ \vdots \\ \gamma_{xy}(m) \end{pmatrix}, \Gamma_{xx} = \begin{pmatrix} \gamma_{xx}(0) & \gamma_{xx}(1) & \dots & \gamma_{xx}(m) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \dots & \gamma_{xx}(m-1) \\ \vdots & \vdots & & \vdots \\ \gamma_{xx}(m) & \gamma_{xx}(m-1) & & \gamma_{xx}(0) \end{pmatrix}, V = \begin{pmatrix} v_0 & v_1 & \dots & v_m \\ v_0 & v_1 & \dots & v_m \\ \vdots & \vdots & & \vdots \\ v_0 & v_1 & \dots & v_m \end{pmatrix}$$

y cuya solución es:

$$V = \Gamma_{xx}^{-1} \gamma_{xy}$$

Siempre y cuando Γ_{xx} sea una matriz de rango completo, lo cual implica que $X_t, X_{t-1}, X_{t-2}, \dots, X_1$ sean variables dependientes.

Si ahora, X_t se supone que es un proceso de ruido blanco, es decir $\rho_{xx}(k) = 0$ para todo $k \neq 0$, entonces la ecuación (4.2.2.3) se reduce a

$$\rho_{xy}(k) = \frac{\sigma_y}{\sigma_x} v_k \tag{4.2.2.5}$$

es decir,

$$v_k = \frac{\sigma_x}{\sigma_y} \rho_{xy}(k)$$

de esta manera, la función de respuesta a impulsos, v_k , es proporcional a la función de correlación cruzada $\rho_{xy}(k)$ únicamente cuando la variable explicativa es ruido blanco.

Sin embargo, no todas las series explicativas pueden ser ruido blanco por lo cual se utiliza un procedimiento denominado *método de preblanqueado de series*, que permite obtener v_k a partir $\rho_{xy}(j)$; sin importar que la serie explicativa sea o no ruido blanco.

Para explicar este método, considérese el modelo general

$$Y_t = v(B)X_t + \eta_t \quad (4.2.6)$$

si se supone que X_t sigue un modelo *ARMA*, entonces:

$$\phi_x(B)X_t = \theta_x(B)\alpha_t$$

en donde α_t es un proceso de ruido blanco, el cual puede ser reescrito como;

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)}X_t \quad (4.2.7)$$

Usualmente, a la serie α_t se le llama *serie explicativa preblanqueada*.

Supóngase ahora, que la serie Y_t sigue el mismo modelo que X_t . Para ello aplíquese la misma transformación a Y_t ,

$$\phi_x(B)Y_t = \theta_x(B)\beta_t$$

de donde se obtiene la serie de error, β_t , que no necesariamente tiene que ser ruido blanco

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)}Y_t \quad (4.2.8)$$

Utilizando (4.2.7) y (4.2.6) en (4.2.8) se obtiene

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)}[v(B)X_t + \eta_t] = \frac{\phi_x(B)}{\theta_x(B)}v(B)X_t + \frac{\phi_x(B)}{\theta_x(B)}\eta_t =$$

$$= v(B)\alpha_t + \frac{\phi_x(B)}{\theta_x(B)}\eta_t$$

sea $\varepsilon_t = \frac{\phi_x(B)}{\theta_x(B)}\eta_t$, entonces

$$\beta_t = v(B)\alpha_t + \varepsilon_t \tag{4.2.9}$$

éste último modelo tiene la misma función de transferencia $v(B)$ pero en esta ocasión se encuentra relacionando dos series en donde α_t , sí es un proceso de ruido blanco.

De esta manera, los coeficientes de rezago v_t pueden ser encontrados a partir de $\rho_{\alpha\beta}(k)$ utilizando la siguiente relación:

$$v_k = \frac{\sigma_\beta}{\sigma_\alpha} \rho_{\alpha\beta}(k)$$

obtenida para el caso en donde la serie explicativa, X_t , se suponía ruido blanco.

4.3 Estimador más usual de la función de correlación cruzada.

La función de correlación cruzada ρ_{xy} , como se vio anteriormente, en el rezago k está definida como:

$$\rho_{xy}(k) = \frac{E\{(x_t - \mu_x)(y_{t+k} - \mu_y)\}}{\sigma_x \sigma_y}$$

donde μ y σ denotan media y desviación estándar respectivamente. En la práctica, $\rho_{xy}(k)$ puede ser estimado por:

$$r_{xy}(k) = \frac{S_{xy}(k)}{\sqrt{S_{xx}(0)}\sqrt{S_{yy}(0)}}$$

en donde

$$S_{xy}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad k \geq 0 \tag{4.3.1}$$

$$S_{xy}(k) = \frac{1}{N} \sum_{t=1-k}^N (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad k < 0 \tag{4.3.2}$$

$$S_{xx}(0) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})^2,$$

$$S_{yy}(0) = \frac{1}{N} \sum_{t=1}^{N-k} (y_t - \bar{y})^2.$$

y \bar{x} y \bar{y} son las medias muestrales de las series X_t y Y_t respectivamente.

Debido a que una estimación está expuesta a errores muestrales habrá que buscar un criterio para decidir si $r_{xy}(k)$ puede o no ser considerado cero.

Bajo la suposición de que x_t y y_t son conjuntamente estacionarios y normalmente distribuidos, Bartlett (1955)⁶ derivó la varianza y covarianza aproximada entre dos correlaciones cruzadas estimadas $r_{xy}(k)$ y $r_{xy}(k+j)$. La covarianza aproximada está dada por:

$$\begin{aligned} & \text{Cov}\{r_{xy}(k), r_{xy}(k+j)\} \\ & \cong (N-k)^{-1} \sum_{j=-\infty}^{\infty} \{\rho_{xx}(i)\rho_{yy}(i+j) + \rho_{xy}(i+k+j)\rho_{yx}(k-i)\} \end{aligned}$$

⁶ Bartlett, M.S. (1950). *Stochastic Processes*, Cambridge University Press, Cambridge.

$$\begin{aligned}
 & + \rho_{xy}(k)\rho_{yx}(k+j) \left[\rho_{xx}^2(i) + \frac{1}{2}\rho_{xx}^2(i) + \frac{1}{2}\rho_{yy}^2(i) \right] \\
 & - \rho_{xy}(k) \left[\rho_{xx}(i)\rho_{yy}(i+k+j) + \rho_{xy}(-i)\rho_{yy}(i+k+j) \right] \\
 & - \rho_{yx}(k+j) \left[\rho_{xx}(i)\rho_{xy}(i+k) + \rho_{xy}(-i)\rho_{yy}(i+k) \right]
 \end{aligned} \tag{4.3.3}$$

De aquí, la varianza

$$\begin{aligned}
 & Var[r_{xy}(k)] \\
 & \cong (N-k)^{-1} \sum_{i=-\infty}^{\infty} \left\{ \rho_{xx}(i)\rho_{yy}(i) + \rho_{xy}(k+i)\rho_{xy}(k-i) + \right. \\
 & \left. + \rho_{xy}^2(k) \left[\rho_{yy}^2(i) + \frac{1}{2}\rho_{xx}^2(i) + \frac{1}{2}\rho_{yy}^2(i) \right] + \right. \\
 & \left. - 2\rho_{xy}(k) \left[\rho_{xx}(i)\rho_{xy}(i+k) + \rho_{xy}(-i)\rho_{yy}(i+k) \right] \right\}
 \end{aligned} \tag{4.3.4}$$

Cuando las series son independientes (4.3.3) se simplifica a:

$$\begin{aligned}
 & Cov\{r_{xy}(k), r_{xy}(k+j)\} \\
 & \cong (N-k)^{-1} \sum_{i=-\infty}^{\infty} \rho_{xx}(i)\rho_{yy}(i+j)
 \end{aligned} \tag{4.3.5}$$

Si ahora se supone que la serie x_t es ruido blanco, entonces $\rho_{xx}(i)$ es diferente de cero únicamente para $i=0$, por lo cual (4.3.5) se reduce a

$$Cov\{r_{xy}(k), r_{xy}(k+j)\} \cong (N-k)^{-1} \rho_{yy}(j)$$

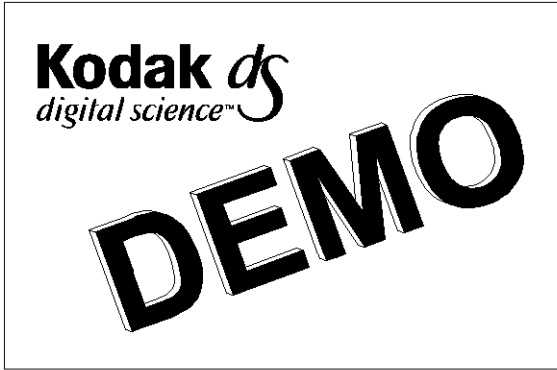
en el caso de la varianza, al suponer que las serie x_t y y_t son independientes, (4.3.4) se reduce a:

$$Var[r_{xy}(k)] = (N - k)^{-1} \sum_{i=-\infty}^{\infty} \{\rho_{xx}(i)\rho_{yy}(i)\}$$

con x_t ruido blanco

$$Var[r_{xy}(k)] \cong (N - k)^{-1}$$

De esta manera, las dos series estimadas en la práctica (sección 4.4) son ruido blanco.



probar la hipótesis de que no hay correlación cruzada entre las series, explicado en la sección 4.4 que un proceso de ruido blanco.

4.4 Identificación de la función de transferencia.

De acuerdo con lo anterior, y en forma de resumen se darán a continuación los pasos que se deben seguir para la identificación de la función de transferencia $v(B)$:

1. En el caso de que la serie explicativa, x_t , no siga un proceso de ruido blanco, entonces se tiene que aplicar el procedimiento de preblanqueado de series. Este procedimiento consiste en suponer que x_t sigue el siguiente proceso $ARMA(p, q)$:

$$\phi_x(B)X_t = \theta_x(B)\alpha_t \tag{4.4.1}$$

es decir

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)} X_t \tag{4.4.2}$$

donde α_t es un proceso de ruido blanco con media cero y varianza σ_a^2 .

2. Transformar la serie respuesta Y_t usando el mismo modelo de preblanqueado que se utilizó en (4.4.1), generando la serie

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)} Y_t \quad (4.4.3)$$

Logrando con esto, una relación equivalente entre dos series, en donde la función de transferencia es la original y la serie que entra como explicativa si sigue un proceso de ruido blanco.

3. El preblanqueado de las series realizado entre los dos pasos anteriores ayudará a obtener los estimadores de v_k a partir de la relación que hay entre la función de respuesta a impulsos y la función de correlación cruzada entre las series α_t y β_t ,

$$\hat{v}_k = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha} r_{\alpha\beta}(k)$$

Para determinar si $r_{\alpha\beta}(k)$ es significativamente distinto de cero, se puede comparar con su error estándar $\frac{1}{\sqrt{N-k}}$.

4. Para identificar el parámetro b , el orden del polinomio $w_m(B)$ y el del polinomio $\delta_a(B)$, se deben de tomar los estimadores preliminares \hat{w}_t y $\hat{\delta}_t$ encontrados a partir de las relaciones que w_t y δ_t tienen con v_k , como se vio en la sección (2.3)

$$\hat{v}(B) = \frac{\hat{w}_m(B)}{\hat{\delta}_a(B)} B^{\hat{h}}$$

4.5 Identificación del proceso de inercia

Una vez identificada la función de transferencia, el proceso de inercia puede ser identificado si se aplican los métodos de identificación para un modelo de series univariantes a:

$$\begin{aligned} N_t &= Y_t - \hat{v}_0 X_t - \hat{v}_1 X_{t-1} - \dots - \hat{v}_k X_{t-k} \\ &= Y_t - \frac{\hat{w}_m(B)}{\hat{\delta}_a(B)} X_{t-h} \end{aligned} \tag{4.5.1}$$

Otra manera, que resulta más sencilla y además que en la práctica ha resultado efectiva, es suponer que entre la serie Y_t y la serie X_t no hay ninguna relación, es decir, que todo el comportamiento de Y_t puede ser explicado a partir del proceso de inercia, N_t ,

$$Y_t = N_t \tag{4.5.2}$$

Así, se pueden aplicar los métodos de identificación univariante a N_t , como son la identificación de la función de autocorrelación (*FAC*) y de la función de autocorrelación parcial (*FACP*), para encontrar el modelo más apropiado. Posteriormente, si Y_t no es del todo indiferente al comportamiento de X_t , entonces el modelo para N_t se simplificará.

5. Estimación

Una vez que se ha identificado un modelo tentativo, el siguiente paso es encontrar los mejores valores de los parámetros para que el modelo represente adecuadamente a las series en consideración. Hay varios métodos estadísticos mediante los cuales se pueden estimar

los parámetros $\delta = (\delta_1, \delta_2, \dots, \delta_a)'$, $w = (w_0, w_1, \dots, w_m)'$, $\phi = (\phi_1, \phi_2, \dots, \phi_p)'$,
 $\theta = (\theta_1, \theta_2, \dots, \theta_q)$, y σ_a^2 del modelo

$$Y_t = \frac{w(B)}{\delta(B)} X_{t-b} + \frac{\theta(B)}{\phi(B)} a_t.$$

Sin embargo, el método que aquí se estudiará es el de *máxima verosimilitud condicionado*.

5.1 Procedimientos básicos para la estimación.

Estimación de modelos AR(p)

Supóngase el modelo:

$$\phi_p(B)y_t = \theta_0 + a_t \tag{5.1.1}$$

equivalente a

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_0 + a_t \tag{5.1.2}$$

en donde y_t representa a la serie estacionaria originada por aplicar las respectivas transformaciones a la serie Y_t , en caso de que ésta originalmente no sea estacionaria; y a_t es un proceso de ruido blanco con distribución normal, media cero y varianza σ_a^2 .

Supóngase que se dispone de T observaciones de la serie $\{y_t\}$, $t = 1, \dots, T$. Para utilizar el método de máxima verosimilitud en la estimación de parámetros, se requiere obtener la función de densidad conjunta $f(y_1, y_2, \dots, y_T)$, la cual puede descomponerse de la siguiente manera

$$f(y_1, y_2, \dots, y_T) = f(y_1) f(y_2 / y_1) f(y_3 / y_2, y_1) \dots f(y_T / y_{T-1}, y_{T-2}, \dots, y_1)$$

originada de utilizar la siguiente definición

$$f(xy) = f(x)f(y/x) \tag{5.1.3}$$

en donde x y y no sólo representan variables sino también vectores de variables.

De esta manera, si el vector de observaciones $y_T = (y_1, y_2, \dots, y_T)'$ se divide en dos subvectores como a continuación se muestra

$$y_T = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \\ y_{p+1} \\ y_{p+2} \\ \vdots \\ y_T \end{pmatrix}$$

entonces $f(y_1, y_2, \dots, y_T)$ puede expresarse de la siguiente manera:

$$f(y_1, y_2, \dots, y_T) = f(y_1, y_2, \dots, y_p) f(y_{p+1}, y_{p+2}, \dots, y_T / y_1, \dots, y_p) \tag{5.1.4}$$

tomando logaritmos de ambos lados y aplicando al segundo término de la derecha sucesivamente la definición (5.1.3), se tiene

$$\ln f(y_T) = \ln f(y_1, y_2, \dots, y_p) + \sum_{t=p+1}^T \ln f(y_t / y_{t-1}, \dots, y_1) \tag{5.1.5}$$

Posteriormente se ilustrará con un ejemplo, como el considerar la distribución conjunta de y_1, y_2, \dots, y_p puede complicar la estimación de los parámetros, por lo cual se asumirá que este término es conocido.

Dado que a_t tiene una distribución normal y que se conocen los p valores inmediatos anteriores de y_t , entonces

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_0 + a_t$$

se distribuye normal, con media

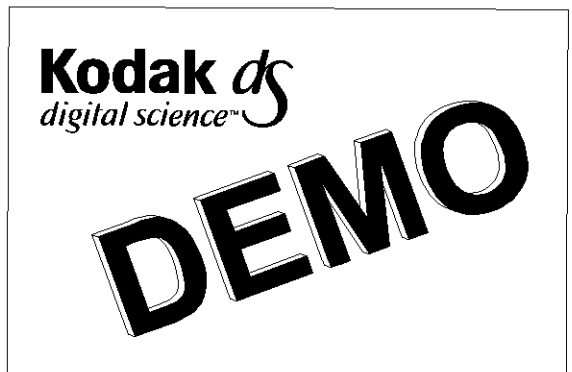
$$\begin{aligned} E(y_t / y_{t-1}, y_{t-2}, \dots, y_1) &= E(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_0) + E(a_t) \\ &= \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} = \theta_0 + \sum_{i=1}^p \phi_i y_{t-i} \end{aligned}$$

en donde $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$; y varianza,

$$Var(y_t / y_{t-1}, y_{t-2}, \dots, y_1) = \sigma_a^2.$$

Es así como:

$$\begin{aligned} \sum_{t=p+1}^T \ln f(y_t / y_{t-1}, \dots, y_1) &= \\ &= \sum_{t=p+1}^T \left[-\frac{1}{2} \ln(2\pi) \right. \end{aligned}$$



$$\begin{aligned}
 &= -\frac{1}{2} \sum_{t=p+1}^T \ln 2\pi - \frac{1}{2} \sum_{t=p+1}^T \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_{t=p+1}^T \left(y_t - \theta_0 - \sum_{i=1}^p \phi_i y_{t-i} \right)^2 = \\
 &= -\frac{(T-p)}{2} \ln 2\pi - \frac{(T-p)}{2} \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_{t=p+1}^T \left(y_t - \theta_0 - \sum_{i=1}^p \phi_i y_{t-i} \right)^2 \quad (5.1.6)
 \end{aligned}$$

La función anterior, (5.1.6), se puede considerar una función de verosimilitud ya que a través de ella se puede obtener probabilidades una vez que se conocen los parámetros $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ y θ_0 . Eliminando el primer término de (5.1.6), por simplificación, se llega a que la función de verosimilitud es:

$$L(\phi, \sigma_a^2 / y_t) = -\frac{(T-p)}{2} \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} S(\phi)$$

con

$$S(\phi) = \sum_{t=p+1}^T (y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} - \theta_0)^2 = \sum_{t=p+1}^T a_t^2 \quad (5.1.7)$$

esta función de verosimilitud $L(\phi, \sigma_a^2 / y_t)$ debe maximizarse con respecto a los parámetros para encontrar los estimadores $\hat{\phi}$ y $\hat{\sigma}_a^2$, denominados *estimadores máximo verosímiles*, los cuales ayudarán a representar a la serie y_t de la mejor manera posible.

Maximizar $L(\phi, \sigma_a^2 / y_t)$ se reduce a minimizar $S(\phi)$ y posteriormente determinar el estimador de σ_a^2 .

Si se observa, la estimación de un modelo $AR(p)$ es equivalente a estimar los parámetros de un modelo de regresión clásico multivariado en donde las variables explicativas están

representadas por $X_i' = (y_{i-1}, y_{i-2}, \dots, y_{i-p})$ y la variable dependiente por y_i . Si se denota al vector de parámetros por $\beta = (\theta_0, \phi_1, \phi_2, \dots, \phi_p)'$ entonces los estimadores mínimos cuadráticos podrán obtenerse mediante la siguiente ecuación, del modelo de regresión lineal multivariado:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

siempre y cuando exista la inversa de $X'X$.

Estimación de modelos MA(q)

Hasta este momento, la estimación de los parámetros no ha presentado ninguna complicación; sin embargo cuando se involucran en el modelo operadores de media móvil, ésta ya no resulta ser tan sencilla. Como ejemplo de las dificultades con que puede uno encontrarse considere al modelo $MA(1)$:

$$\bar{y}_i = a_i - \theta a_{i-1}$$

Sin pérdida de generalidad supóngase que la media de la serie y_i es cero. De esta manera, obsérvese que a_i puede expresarse como

$$\begin{aligned} a_i &= y_i + \theta a_{i-1} \\ &= y_i + \theta y_{i-1} + \theta^2 a_{i-2} \\ &\dots \\ &= y_i + \theta y_{i-1} + \theta^2 y_{i-2} + \dots + \theta^{i-1} y_1 + \theta^i a_0 \end{aligned} \tag{5.1.8}$$

es verdad que a_0 se desconoce, pero ya que se ha supuesto que a_i tiene una distribución $N(0, \sigma_a^2)$ suele suponerse que $a_0 = E(a_i) = 0$.

La distribución conjunta de a_1, a_2, \dots, a_r está dada por

$$f(a_1, a_2, \dots, a_r) = \frac{1}{(2\pi\sigma_a^2)^{\frac{r}{2}}} \exp\left\{-\sum_{i=1}^r \frac{a_i^2}{2\sigma_a^2}\right\} \quad (5.1.9)$$

de ahí que, la distribución conjunta de y_1, y_2, \dots, y_r puede obtenerse mediante la transformación:

$$f(y_1, y_2, \dots, y_r) = f(a_1, a_2, \dots, a_r) \cdot J \quad (5.1.10)$$

en donde J denota el jacobiano de la transformación definido como:

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial a_1} & \frac{\partial y_1}{\partial a_2} & \dots & \frac{\partial y_1}{\partial a_r} \\ \frac{\partial y_2}{\partial a_1} & \frac{\partial y_2}{\partial a_2} & \dots & \frac{\partial y_2}{\partial a_r} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_r}{\partial a_1} & \frac{\partial y_r}{\partial a_2} & \dots & \frac{\partial y_r}{\partial a_r} \end{vmatrix} \quad (5.1.14)$$

Para este caso, se tiene:

$$J = \begin{vmatrix} 1 & 0 & \dots & 0 & 0 \\ -\theta & 1 & \dots & 0 & 0 \\ 0 & -\theta & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -\theta & 1 \end{vmatrix}$$

Obsérvese que $\frac{\partial y_i}{\partial a_j} = 0$ para $i < j$ por lo que se tiene una matriz triangular cuyo

determinante es el producto de los valores de la diagonal. Es así como

$$J = \prod_{t=1}^T \frac{\partial y_t}{\partial a_t} = 1 \quad (5.1.12)$$

y de (5.1.10) se obtiene así la siguiente función de verosimilitud:

$$L(\theta, \sigma_a^2 / y_1, \dots, y_T, a_0) = f(y_1, y_2, \dots, y_T) = \frac{1}{(2\pi\sigma_a^2)^{\frac{T}{2}}} \exp\left\{-\frac{S(\theta)}{2\sigma_a^2}\right\} \quad (5.1.13)$$

con

$$S(\theta) = \sum_{t=1}^T a_t^2 = \sum_{t=1}^T (y_t + \theta y_{t-1} + \theta^2 y_{t-2} + \dots + \theta^{t-1} y_1 + \theta^t a_0)^2$$

puede pensarse en minimizar $S(\theta)$, como se había hecho anteriormente, para después maximizar $L(\sigma_a^2 / \hat{\theta})$, sin embargo la ecuación

$$\frac{\partial S(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} = 0$$

no tiene solución analítica debido a que no es lineal en θ . Puede utilizarse algún procedimiento numérico para encontrar el mínimo de $S(\theta)$; como es una búsqueda en rejilla que consiste en evaluar $S(\theta)$ en diversos puntos de θ tomando en consideración que $-1 < \theta < 1$, obteniéndose una gráfica como la de la *fig.(4.1)* en donde se puede observar en que valor $S(\theta)$ se hace mínimo.

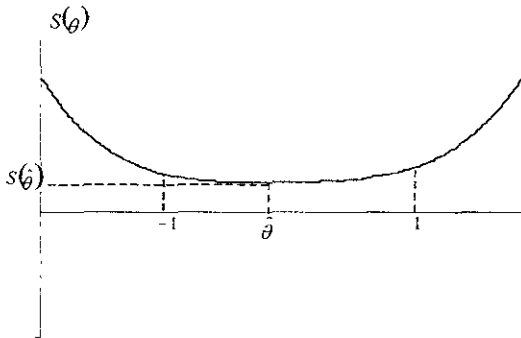


Figura 4.1

Otro procedimiento numérico de estimación no-lineal es el algoritmo de Marquard, sugerido por Box y Jenkins, el cual se explicará en la sección (5.2).

Estimación de modelos ARMA(p, q)

Considérese, ahora, un proceso ARMA(p, q):

$$\phi_p(B)y_t = \theta_0 + \theta_q(B)a_t \tag{5.1.14}$$

la función de verosimilitud para estimar $\theta_0, \theta, \phi, \sigma_a^2$, se obtendrá de una combinación de los procedimientos utilizados anteriormente.

Así, para obtener la función conjunta $f(y_1, y_2, \dots, y_T)$ de las T observaciones de la serie y_t , se supondrá que se conoce la distribución conjunta de y_1, y_2, \dots, y_p , en la ecuación (5.1.5).

Ahora, a_t esta dada por:

$$a_t = y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} - \theta_0 + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

dado que el proceso a_t sigue un proceso de ruido blanco. La función conjunta de $y_{p+1}, y_{p+2}, \dots, y_T$ se obtiene a partir de la transformación:

$$f(y_{p+1}, y_{p+2}, y_{p+3}, \dots, y_T) = f(a_{p+1}, a_{p+2}, a_{p+3}, \dots, a_T) \cdot J$$

en donde el jacobiano es:

$$J = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ -\theta_1 & 1 & \dots & 0 & 0 \\ -\theta_2 & -\theta_1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ -\theta_T & -\theta_{T-1} & \dots & -\theta_1 & 1 \end{pmatrix}$$

$$L(\phi, \theta_0, \theta, \sigma_a^2 / y_T, a_0) = -\frac{(T-p)}{2} \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} S(\phi, \theta_0, \theta) \quad (5.1.15)$$

con:

$$S(\phi, \theta_0, \theta) = \sum_{p+1}^T a_t^2$$

en donde a_t deberá ser expresada sólo en términos de la serie y , y los valores iniciales $a_0 = (a_0, a_1, \dots, a_q)$.

Con base en la suposición de que $\{a_t\}$ tiene una distribución $N(0, \sigma_a^2)$, se puede suponer que los primeros residuales son igual a su valor esperado cero.

Al igual que en los casos anteriores, la estimación de los parámetros, se reduce a la minimización de $S(\phi, \theta_0, \theta)$, la cual se realizará mediante la aplicación de un algoritmo no lineal, explicado en la siguiente sección.

Por el momento supóngase que se han encontrado los estimadores máximo verosímiles, $\hat{\phi}$, $\hat{\theta}_0$ y $\hat{\theta}$, por lo cual hay que determinar ahora el estimador de σ_a^2 . Éste se obtiene de maximizar la función de verosimilitud, para ello dérivese $L(\sigma_a^2 / Y_t; \hat{\phi}, \hat{\theta}_0, \hat{\theta}_1)$ respecto a σ_a^2 :

$$\frac{\partial L}{\partial \sigma_a^2} \Big|_{\sigma_a^2 = \hat{\sigma}_a^2} = -\frac{(T-p)}{2\hat{\sigma}_a^2} + \frac{S(\hat{\phi}, \hat{\theta}_0, \hat{\theta})}{2\hat{\sigma}_a^4} = 0 \tag{5.1.16}$$

de donde se obtiene

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}, \hat{\theta}_0, \hat{\theta})}{(T-p)} \tag{5.1.17}$$

5.2 Estimación no lineal

Para la minimización de $S(\phi, \theta_0, \theta)$ se requiere de un algoritmo no lineal por las razones explicadas en la sección anterior, es por eso que Box y Jenkins sugieren la utilización de un método numérico para encontrar los estimadores $\hat{\phi}, \hat{\theta}_0, \hat{\theta}$ que hagan mínimo $S(\phi, \theta_0, \theta)$. Este algoritmo es una modificación del algoritmo de Gauss-Newton debida a Marquardt.

Debido a que los modelos lineales son casos especiales de los no lineales, se pueden ilustrar los pasos de que consta el algoritmo utilizando el siguiente modelo $AR(p)$, cuya función a minimizar, $S(\phi)$, es lineal

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t \quad (5.2.1)$$

si se denota a y_{t-1} por x_{t1} , a y_{t-2} por x_{t2} , y así sucesivamente, se puede observar que (5.2.1) es equivalente a estimar un modelo de regresión lineal multivariado:

$$y_t = \phi_1 x_{t1} + \phi_2 x_{t2} + \dots + \phi_p x_{tp} + a_t \quad (5.2.2)$$

para $t=1,2,\dots,T$, y a_t variables independientes idénticamente distribuidas $N(0, \sigma_a^2)$ independientes de todas las x_{ti} .

Sea $\mathbf{Y} = (y_1, y_2, \dots, y_T)$, $\phi = (\phi_1, \phi_2, \dots, \phi_p)'$ y \mathbf{X} la matriz correspondiente a las variables x_{ti} , entonces

$$\hat{\phi} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (5.2.3)$$

son los estimadores, obtenidos de minimizar

$$\begin{aligned} S(\phi) &= \sum_{t=1}^T a_t^2 \\ &= \sum_{t=1}^T (y_t - \hat{\phi}_1 x_{t1} + \hat{\phi}_2 x_{t2} + \dots + \hat{\phi}_p x_{tp})^2 \end{aligned} \quad (5.2.4)$$

por el método de mínimos cuadrados utilizado en el análisis de regresión lineal multivariado.

Estos estimadores mínimo cuadráticos, también pueden ser obtenidos mediante el siguiente procedimiento.

Sea $\tilde{\phi} = (\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_p)'$ valores iniciales de $\phi = (\phi_1, \phi_2, \dots, \phi_p)$, entonces la suma de cuadrados de los residuales, $S(\phi)$, puede ser reescrita como:

$$\begin{aligned}
 S(\hat{\phi}) &= \sum_{i=1}^T \left(y_i - \hat{\phi}_1 x_{i1} - \hat{\phi}_2 x_{i2} - \dots - \hat{\phi}_p x_{ip} \right)^2 \\
 &= \sum_{i=1}^T \left[\left(y_i - \hat{\phi}_1 x_{i1} - \dots - \hat{\phi}_p x_{ip} \right) + \left(\tilde{\phi}_1 x_{i1} + \dots + \tilde{\phi}_p x_{ip} \right) - \left(\tilde{\phi}_1 x_{i1} + \dots + \tilde{\phi}_p x_{ip} \right) \right]^2 \\
 &= \sum_{i=1}^T \left[y_i - \tilde{\phi}_1 x_{i1} - \dots - \tilde{\phi}_p x_{ip} \right. \\
 &\quad \left. - (\hat{\phi}_1 - \tilde{\phi}_1) x_{i1} - \dots - (\hat{\phi}_p - \tilde{\phi}_p) x_{ip} \right]^2
 \end{aligned} \tag{5.2.5}$$

ó

$$S(\delta) = S(\hat{\phi}) = \sum \left(\tilde{\alpha}_i - \delta_1 x_{i1} - \dots - \delta_p x_{ip} \right)^2 \tag{5.2.6}$$

en donde $\tilde{\alpha}_i$ son los residuales obtenidos de utilizar los estimadores iniciales, $\tilde{\phi}$, y $\delta = (\hat{\phi} - \tilde{\phi})$.

Si se observa $S(\delta)$, en (5.2.6), tiene la misma forma que $S(\phi)$ en (5.2.4); de aquí que el valor de δ obtenido por mínimos cuadrados es

$$\delta = (X'X)^{-1} X'\tilde{\alpha} \tag{5.2.7}$$

Una vez calculado δ , entonces los estimadores por mínimos cuadrados de ϕ están dados por:

$$\hat{\phi} = \tilde{\phi} + \delta \tag{5.2.8}$$

Nótese que en (5.2.2)

$$\frac{\partial y_t}{\partial \phi_i} = x_{ti} \quad (5.2.9)$$

para $i = 1, 2, \dots, p$ y $t = 1, 2, \dots, T$. De ahí que la matriz X usada en (5.2.3) y (5.2.7) es la matriz de derivadas parciales de y_t con respecto a cada uno de los parámetros, ϕ_i .

Para este modelo que es lineal, la ecuación (5.2.8) da los estimadores mínimo cuadráticos, sin embargo, para un modelo no-lineal, el parámetro obtenido sólo será el nuevo valor inicial de una nueva iteración, realizando este procedimiento cuantas veces sea necesario para obtener convergencia, es decir, hasta que el cambio relativo en cada parámetro no sea mayor que un cierto valor pequeño, fijado de antemano.

De acuerdo con lo anterior, el algoritmo de minimización numérico puede describirse en los siguientes pasos:

- 1) Obtener valores iniciales de los parámetros ϕ, θ_0, θ , los cuales a partir de este momento se denotarán por $\hat{\beta}_0 = (\hat{\phi}, \hat{\theta}_0, \hat{\theta})'$.
- 2) Calcular los residuos del modelo utilizando los estimadores iniciales, $a_i(\hat{\beta}_0)$.
- 3) Calcular las derivadas parciales de los residuos respecto a cada uno de los parámetros:

$$\frac{\partial a_i(\hat{\beta}_0)}{\partial \beta_i} \quad i = 1, 2, \dots, p + q + 1$$

En el caso de que se quiera programar este algoritmo, las derivadas parciales pueden calcularse numéricamente mediante:

$$\frac{\partial a_i(\hat{\beta}_0)}{\partial \beta_i} \approx \frac{a_i(\hat{\beta}_0 + \mathbf{1}_i, \varepsilon) - a_i(\hat{\beta}_0)}{\varepsilon}$$

en donde $\mathbf{1}_i$ es un vector con un uno en la componente i y cero en otro caso, y ε es un número muy pequeño. Esta fórmula para la diferenciación numérica surge de manera directa de la definición, en términos de límite, de la derivada de una función.

- 4) Estimar los coeficientes de una regresión entre los residuos calculados en el inciso 2) como la variable dependiente, y el vector de $(p + q + 1)$ variables explicativas x_i , que son iguales a las derivadas parciales de los residuos respecto a cada uno de los parámetros:

$$x_i = - \frac{\partial a_t(\hat{\beta}_0)}{\partial \beta_i}$$

- 5) Sea δ el vector de parámetros obtenido, a partir de la regresión, entonces se definirá al nuevo estimador inicial como:

$$\hat{\beta}_1 = \hat{\beta}_0 + \delta$$

si aún $\hat{\beta}_{t-1} - \hat{\beta}_t < \alpha$, siendo α un valor pequeño, no se cumple entonces hay que volver al segundo paso de este algoritmo, ahora con $\hat{\beta}_t$ como estimador inicial.

5.3 Estimación de máxima verosimilitud exacta

La función de verosimilitud utilizada en el método de máxima verosimilitud condicionada, explicado anteriormente, es sólo una aproximación de la *función de verosimilitud exacta* (5.1.5) ya que, como se mencionó, el utilizar la función exacta en vez de la aproximación complica los procedimientos de estimación. Para ilustrar este hecho, considérese el modelo

$AR(1)$:

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + a_t \tag{5.3.1}$$

Como se sabe, un modelo autorregresivo estacionario tienen una representación equivalente de media móvil, en este caso el modelo MA correspondiente a (5.3.1) es :

$$\bar{y}_t = \sum_{i=0}^{\infty} \phi_i' \alpha_{t-i} \quad (5.3.2)$$

Con media igual a cero, por ser α , un proceso de ruido blanco normalmente distribuido y varianza $\sigma^2 = \left(\frac{\sigma_a^2}{1-\phi^2} \right)$ (ver capítulo 1, sec. 4.2).

Dado que se cuenta con T observaciones de la serie y_t , y de acuerdo a (5.1.1), la función de máxima verosimilitud exacta tiene la siguiente forma:

$$\ln f(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_T) = \ln f(\bar{y}_1) + \sum_{t=2}^T \ln f(\bar{y}_t / \bar{y}_{t-1}) \quad (5.3.2)$$

El primer término del lado derecho de la igualdad, que en los procedimientos de estimación anteriores se suponía conocido, ahora se puede sustituir por su verdadera distribución, es decir

$$\begin{aligned} \ln f(\bar{y}_1) &= \ln \left[\frac{1}{2\pi \frac{\sigma_a^2}{(1-\phi^2)}} \exp \left[-\frac{\bar{y}_1^2 (1-\phi^2)}{\sigma_a^2} \right] \right] = \\ &= -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \left(\frac{\sigma_a^2}{1-\phi^2} \right) - \bar{y}_1^2 \frac{(1-\phi^2)}{\sigma_a^2} \end{aligned} \quad (5.3.3)$$

El segundo término de la derecha de (5.3.2) se puede sustituir por la distribución condicional, obtenida de acuerdo a los procedimientos de estimación condicional vistos en las secciones anteriores, es decir

$$\begin{aligned} \sum_{i=2}^T \ln f(\bar{y}_i / \bar{y}_{i-1}) &= \sum_{i=2}^T \ln \left\{ \frac{1}{2\pi\sigma_u^2} \exp \left[-\frac{(\bar{y}_i - \phi_1 \bar{y}_{i-1})^2}{2\sigma_u^2} \right] \right\} = \\ &= -\frac{T-1}{2} \ln 2\pi - \frac{T-1}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_{i=2}^T (\bar{y}_i - \phi_1 \bar{y}_{i-1})^2 \end{aligned} \quad (5.3.4)$$

Por simplificación se pueden eliminar los términos constantes, es decir aquellos que no dependen de σ_u^2 ó de ϕ_1 , tanto del término (5.3.3) como de (5.3.4). y uniendo ambos términos es como se tiene la función de verosimilitud exacta del modelo $AR(1)$:

$$\begin{aligned} L(\phi, \sigma_u^2 / y_i) &= -\frac{1}{2} \ln \left(\frac{\sigma_u^2}{1-\phi_1^2} \right) - \bar{y}_1^2 \frac{(1-\phi_1^2)}{\sigma_u^2} - \frac{T-1}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_{i=2}^T (\bar{y}_i - \phi_1 \bar{y}_{i-1})^2 = \\ &= -\frac{T}{2} \ln \sigma_u^2 + \frac{1}{2} \ln(1-\phi_1^2) - \frac{(1-\phi_1^2)}{\sigma_u^2} \bar{y}_1^2 - \frac{1}{2\sigma_u^2} \sum_{i=2}^T (\bar{y}_i - \phi_1 \bar{y}_{i-1})^2 \end{aligned} \quad (5.3.5)$$

derívese a (5.3.5) con respecto a ϕ , e igualando a cero encuéntrase el valor de ϕ que maximice la función de máxima verosimilitud exacta,

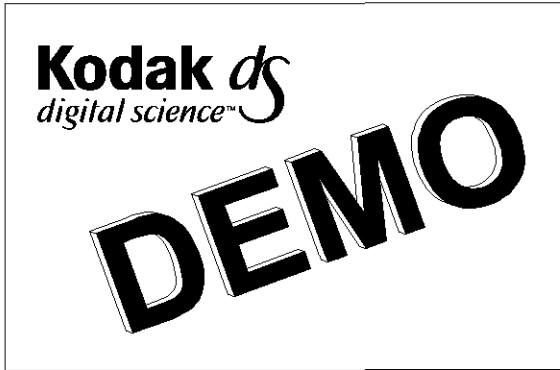
$$\frac{\partial L(\phi, \sigma_u^2 / y_i)}{\partial \phi} = -\frac{\phi_1}{(1-\phi_1^2)} + 2 \frac{\phi_1}{\sigma_u^2} \bar{y}_1^2 + \frac{1}{\sigma_u^2} \sum_{i=2}^T (\bar{y}_i - \phi_1 \bar{y}_{i-1}) \bar{y}_{i-1}$$

la cual al igualarla a cero y desarrollando se tiene una función cúbica cuya solución se puede obtener aplicando algún método numérico. Nótese que el modelo $AR(1)$ es el modelo para series de tiempo más sencillo que se puede encontrar, el manejar modelos ya mas complicados y querer estimar los valores de los parámetros involucrados mediante la función de máxima verosimilitud exacta, puede ser verdaderamente complicado.

5.4 Obtención de valores iniciales

La utilización del algoritmo de minimización no-lineal requiere de ciertos valores iniciales para comenzar a iterar sucesivamente. Estos valores podrían darse de manera arbitraria, sin embargo, para obtener convergencia rápidamente es preferible que los valores iniciales se encuentren lo más cerca posible al mínimo que se desea encontrar.

Una manera de encontrar estos valores iniciales es a partir de las relaciones que tienen los ρ_k en la relación entre las series x_t y y_{t+k} ,



series de tiempo univariantes. En este caso se estiman los coeficientes de autocorrelación mediante

$$\begin{aligned} \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned}$$

los coeficientes ρ_k , $k = 1, 2, \dots, p$, son desconocidos, pero se pueden sustituir por sus estimadores r_k definidos como:

$$r_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T-k} (y_t - \bar{y})^2}, \quad k = 0$$

con $y = \sum_{i=1}^T \frac{y_i}{T}$ como la media muestral. Así, se tendrá como únicos valores desconocidos a

$\phi_1, \phi_2, \dots, \phi_p$; y sus estimadores se pueden encontrar resolviendo el sistema:

$$\begin{aligned} r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 + \dots + \hat{\phi}_p r_{p-1} \\ r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 + \dots + \hat{\phi}_p r_{p-2} \\ &\vdots \\ r_p &= \hat{\phi}_1 r_{p-1} + \hat{\phi}_2 r_{p-2} + \dots + \hat{\phi}_p \end{aligned}$$

Por ejemplo, si se tiene un modelo $AR(1)$, entonces

$$\hat{\phi}_1 = r_1$$

y para modelos $AR(2)$, hay que resolver el sistema de ecuaciones:

$$\begin{aligned} r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 \\ r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 \end{aligned}$$

teniendo como estimadores de ϕ_1 y de ϕ_2 a:

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{(1-r_1^2)}, \quad \hat{\phi}_2 = \frac{(r_2-r_1^2)}{(1-r_1^2)}$$

Cuando el modelo es del tipo media móvil, $MA(q)$, no se cuenta con una relación lineal entre los parámetros $\theta_1, \dots, \theta_q$ y las autocorrelaciones $\rho_1, \rho_2, \dots, \rho_q$ por lo cual deberá resolverse el sistema de ecuaciones no-lineal

$$r_1 = \frac{(-\theta_1 + \theta_1 \theta_2 + \dots + \theta_{q-1} \theta_q)}{(1 + \hat{\theta}_1^2 + \hat{\theta}_2^2 + \dots + \hat{\theta}_q^2)}$$

$$r_2 = \frac{(-\theta_2 + \theta_1\theta_3 + \dots + \theta_{q-2}\theta_q)}{(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)}$$

⋮

$$r_q = \frac{-\theta_q}{(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)}$$

Como puede observarse este sistema no tendrá una única solución, sin embargo solamente una de ellas satisficera las condiciones de invertibilidad, por ejemplo en un modelo $MA(1)$ se obtienen dos soluciones:

$$\hat{\theta} = -(2r_1)^{-1} + \sqrt{(2r_1)^2 - 1} \quad \text{y} \quad \hat{\theta}' = -(2r_1)^{-1} - \sqrt{(2r_1)^2 - 1}$$

los cuales cumplen con:

$$\hat{\theta}\hat{\theta}' = (2r_1)^{-2} + (2r_1)^{-1} \sqrt{(2r_1)^2 - 1} - (2r_1)^{-1} \sqrt{(2r_1)^2 - 1} - (2r_1)^{-2} - 1 = 1$$

Y por lo tanto sólo una de ellas satisface la condición de invertibilidad ya que si $\hat{\theta} < 1$ entonces $\hat{\theta}' > 1$ y al revés.

En el caso de un proceso $ARMA(p, q)$ se presentarán tanto ecuaciones lineales como no-lineales, ya que como se vio en la sección 4.3 del primer capítulo la relación entre los parámetros y los coeficientes de autocorrelación es la siguiente:

$$\rho_k = \phi_1\rho_{k-1} + \dots + \phi_p\rho_{k-p} \quad \text{para } k > q$$

y los estimadores iniciales de los parámetros $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$ se obtienen de:

$$\begin{aligned} r_{q+1} &= \hat{\phi}_1 r_q + \hat{\phi}_2 r_{q-1} + \dots + \hat{\phi}_p r_{q-p+1} \\ r_{q+2} &= \hat{\phi}_1 r_{q+1} + \hat{\phi}_2 r_q + \dots + \hat{\phi}_p r_{q-p+2} \\ &\vdots \\ r_{q+p} &= \hat{\phi}_1 r_{q+p-1} + \hat{\phi}_2 r_{q+p-2} + \dots + \hat{\phi}_p r_q \end{aligned}$$

Posteriormente, estos estimadores, $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$, deberán sustituirse en las primeras q , autocorrelaciones muestrales, para determinar $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$.

Tómese como ejemplo al modelo $ARMA(1,1)$, a través del sistema de ecuaciones lineales se determinará que el estimador de ϕ está dado por:

$$\hat{\phi} = \frac{r_2}{r_1}$$

mientras que θ se puede obtener de

$$r_1 = \frac{(1 - \theta \hat{\phi})(\hat{\phi} - \theta)}{(1 + \theta^2 - 2\theta \hat{\phi})}$$

es decir

$$(r_1 - \hat{\phi})\theta^2 + (1 + \hat{\phi}^2 - 2\hat{\phi}r_1)\theta + (r_1 - \hat{\phi}) = 0$$

del cual se obtienen las siguientes dos soluciones

$$\theta = \frac{-(1 + \hat{\phi}^2 - 2\hat{\phi}r_1) \pm \sqrt{(1 + \hat{\phi}^2 - 2\hat{\phi}r_1)^2 - 4(r_1 - \hat{\phi})^2}}{2(r_1 - \hat{\phi})}$$

$$\hat{\theta}' = \frac{-(1 + \hat{\phi}_2 - 2\hat{\phi}_1) \pm \sqrt{(1 + \hat{\phi}_2 - 2\hat{\phi}_1)^2 - 4(r_1 - \hat{\phi})^2}}{2(r_1 - \hat{\phi})}$$

de las cuales, la que cumpla estar dentro del círculo unitario será la que se tome.

Un estimador inicial de θ_0 puede obtenerse a partir de que se sabe que:

$$\theta_0 = E[\phi(B)y_t]$$

de modo que

$$\hat{\theta}_0 = (1 - \hat{\phi}_1 - \hat{\phi}_2 - \dots - \hat{\phi}_p)E(y_t) = (1 - \hat{\phi}_1 - \hat{\phi}_2 - \dots - \hat{\phi}_p)\bar{y}$$

Sin embargo en la práctica se prefiere omitir este parámetro hasta que en la etapa de verificación se haga verdaderamente necesario.

5.5 Estimación de modelos de regresión dinámica.

Los procedimientos de las secciones anteriores serán ahora generalizados para los modelos dinámicos de regresión. Para ello supóngase el modelo

$$y_t = y_t^* + \eta_t \tag{5.5.1}$$

en donde y_t^* representa a la función de transferencia con parámetros m, a, b :

$$y_t^* = \frac{w_m(B)}{\delta_a(B)} x_{t-h} \tag{5.5.2}$$

y η_t al proceso de inercia, el cual sigue un proceso $ARMA(p, q)$

$$\eta_t = \frac{\theta_q(B)}{\phi_p(B)} a_t \quad (5.5.3)$$

A partir de (5.5.2) y (5.5.3) se tiene el modelo:

$$y_t = \frac{w_m(B)B^b}{\delta_a(B)} x_t + \frac{\theta_q(B)}{\phi_p(B)} a_t \quad (5.5.4)$$

el cual puede reescribirse de la siguiente manera:

$$\delta_a(B)\phi_p(B)y_t = w_m(B)\phi_p(B)B^b x_t + \delta_a(B)\theta_q(B)a_t \quad (5.5.5)$$

equivalente a

$$c(B)y_t = d(B)B^b x_t + e(B)a_t \quad (5.5.6)$$

en donde $c(B)$ es un polinomio de rezago de orden $a + b$ obtenido de:

$$\begin{aligned} c(B) &= \delta_a(B)\phi_p(B) = (1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_a B^a)(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \\ &= (1 - c_1 B - c_2 B^2 - \dots - c_{a+p} B^{a+p}) \end{aligned}$$

de la misma manera que:

$$\begin{aligned} d(B) &= w_m(B)\phi_p(B) = (w_0 + w_1 B + w_2 B^2 + \dots + w_m B^m)(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \\ &= (d_0 - d_1 B - d_2 B^2 - \dots - d_{m+p} B^{m+p}) \end{aligned}$$

y

$$e(B) = \delta_a(B)\theta_q(B) = (1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_a B^a)(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \\ = (1 - e_1 B - e_2 B^2 - \dots - e_{a+q} B^{a+q})$$

es así como:

$$y_t = c_1 y_{t-1} + c_2 y_{t-2} + \dots + c_{a+p} y_{t-a-p} + d_0 x_{t-h} - d_1 x_{t-h-1} - \dots - d_{m+p} x_{t-m-p-h} + \\ + a_t - e_1 a_{t-1} - \dots - e_{a+q} a_{t-a-q} \quad (5.5.7)$$

en donde se puede observar, que la función de verosimilitud estará condicionada a $a + p$ valores rezagados de y_t , a $m + p + b$ valores de la variable explicativa x_t , y a un vector $a_0 = (a_1, \dots, a_{a+q})$ de valores iniciales que se supondrán cero, por ser su valor esperado.

Por lo tanto, llamando $h = \max\{a + p, m + p + b\}$ la función de densidad conjunta de y_1, y_2, \dots, y_T podrá expresarse como:

$$f(y_1, y_2, \dots, y_T) = f(y_1, y_2, \dots, y_h) \prod_{i=h+1}^T f(y_i / y_{i-1}, y_{i-2}, \dots, y_1) \quad (5.5.8)$$

aplicando logaritmos de ambos lados de (5.5.8) se sigue que la función de verosimilitud puede ser expresada como sigue:

$$L(y_T) = L(y_h) + \sum_{i=h+1}^T L(y_i / y_{i-1}, y_{i-2}, \dots, y_1)$$

en donde $y_T = (y_1, y_2, \dots, y_T)'$ es el vector de las T observaciones y $y_h = (y_1, y_2, \dots, y_h)'$ representa a las primeras h observaciones.

A partir de (5.5.7) se obtiene

$$a_t = y_t - c_1 y_{t-1} - c_2 y_{t-2} - \dots - c_{a+p} y_{t-a-p} - d_0 x_{t-h} + d_1 x_{t-h-1} + \dots + d_{m+p} x_{t-m-p-h} + e_1 a_{t-1} + \dots + e_{a+q} a_{t-a-q} \quad (5.5.8)$$

si se supone como hasta ahora se ha hecho, que a_t tiene una distribución normal con media cero y varianza constante, σ_a , se obtiene la función de densidad conjunta de los errores aleatorios

$$f(a_{h+1}, a_{h+2}, a_{h+3}, \dots, a_T) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{T-h} \exp \left\{ -\frac{1}{2\sigma_a^2} \sum_{i=h+1}^T a_i^2 \right\} \quad (5.5.9)$$

A partir de donde se puede obtener la función de densidad conjunta de $y_{h+1}, y_{h+2}, \dots, y_T$

$$f(y_{h+1}, y_{h+2}, \dots, y_T) = f(a_{h+1}, a_{h+2}, \dots, a_T) \prod_{i=h+1}^T \frac{\partial y_i}{\partial a_i} = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{T-h} \exp \left\{ -\frac{1}{2\sigma_a^2} \sum_{i=h+1}^T (y_i - c_1 y_{i-1} - \dots - c_{a+p} y_{i-a-p} - d_0 x_{i-h} + \dots + d_{m+p} x_{i-m-p-h} + e_{i-1} + \dots + e_{a+q} a_{i-a-q})^2 \right\}$$

aplicando logaritmos se tiene:

$$\ln f(y_{h+1}, y_{h+2}, \dots, y_T) = -\frac{1}{2} (T-h) \ln 2\pi - \frac{1}{2} (T-h) \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_{i=h+1}^T S(\delta, w, \phi, \theta)$$

de donde se puede prescindir del primer término del lado derecho de la igualdad, por ser constante, y obtener la función de verosimilitud condicional

$$L(\delta, w, \phi, \theta, \sigma_a^2 / b, x_h, y_h, a_0) = -\frac{T-h}{2} \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_{i=h+1}^T S(\delta, w, \phi, \theta)$$

lo que reduce la estimación de parámetros a la minimización de

$$S(\delta, w, \phi, \theta) = \sum_{i=h+1}^T a_i^2 \quad (5.5.9)$$

La minimización de (5.5.9) se realizará mediante la generalización de los procedimientos explicados anteriormente (sección 5.2 y sección 5.3). Los cuales se pueden resumir en los siguientes pasos:

Paso 1) Para la utilización del algoritmo de minimización no_lineal se requiere de ciertos valores iniciales. En el caso de la función de transferencia, los valores iniciales de los parámetros w_i , $i=0,1,\dots,m$ y δ_j , $j=1,2,\dots,a$ pueden ser obtenidos resolviendo el sistema de ecuaciones que los relaciona con los \hat{v}_k :

$$\hat{v}(B) = \frac{\hat{w}(B)}{\hat{\delta}(B)}$$

mientras que los estimadores \hat{v}_k se obtienen de:

$$\hat{v}_k = r_{\alpha\beta}(k) \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha}$$

en donde $r_{\alpha\beta}(k)$ representa el coeficiente de correlación entre α_t y β_t ; que son dos series obtenidas por el método de preblanqueado y que tienen la característica de estar relacionadas por la misma función de transferencia que las series x_t y y_t , con la ventaja de que en este caso la variable explicativa α_t es ruido blanco.

Paso 2) Los estimadores iniciales de los parámetros involucrados en el proceso de inercia, serán los valores estimados para la serie univariate y_t .

Paso 3) Calcular los residuales, considerando los valores iniciales de los parámetros obtenidos en el primer y segundo paso.

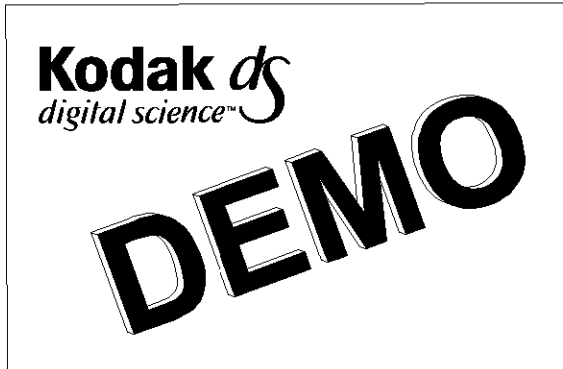
$$\hat{a}_t = \hat{y}_t - \hat{c}_1 \hat{y}_{t-1} - \hat{c}_2 \hat{y}_{t-2} - \dots - \hat{c}_{a+p} \hat{y}_{t-a-p} - \hat{d}_0 x_{t-h} + \hat{d}_1 x_{t-h-1} + \dots + \hat{d}_{m+p} x_{t-m-p-h} + \hat{e}_1 \hat{a}_{t-1} + \dots + \hat{e}_{a+q} \hat{a}_{t-a-q} \tag{5.5.10}$$

en donde $\hat{c}(B) = \hat{\delta}(B)\hat{\phi}(B)$, $\hat{d}(B) = \hat{\phi}(B)\hat{w}(B)$ y $\hat{e}(B) = \hat{\delta}(B)\hat{\theta}(B)$.

Paso 4) Aplicar el algoritmo de estimación no lineal explicado en la sección 5.2, para minimizar (5.5.10), hasta obtener convergencia.

6. Verificación

Esta etapa consiste en aplicar algunas pruebas para verificar que los supuestos en los que se ha basado la construcción del modelo se satisfacen. No sólo puede encerrarse esta etapa en el concepto de verificación sino también de corrección ya que si alguna de las pruebas indica que alguno de los supuestos se está violando, ella misma servirá para dar alguna idea de la correcta adecuación.



por lo cual, se aplicarán todas las pruebas para verificar si se está violando sólo alguno de los supuestos o simplemente de no estar violando ninguno.

Los modelos de series univariantes como el modelo de Box-Cox se dividirán en dos: a)

pruebas de verificación sobre los errores y b) pruebas de verificación sobre los parámetros.

6.1 Sobre los errores

Estas pruebas consisten en comprobar que los supuestos respecto a los errores; media cero, varianza constante, falta de correlación para cualquier retardo y distribución normal; se

satisfacen. Una de las formas más simples de hacerlo es a través del análisis de residuales, en donde, el residual es la medida que existe entre los valores observados y los valores estimados por el modelo, esto es:

$$\hat{a}_t = y_t - \hat{y}_t \tag{6.1.1}$$

Cuando el tamaño de la muestra es grande, se espera que los errores aleatorios y los residuales sean esencialmente iguales; y por esta razón, al analizar los residuales observados $\{\hat{a}_t\}$ se analiza básicamente lo que debería ser una realización del proceso de ruido blanco $\{a_t\}$.

Verificación de modelos para series temporales univariantes a través del análisis de residuales.

Se empezará por revisar si los residuales son mutuamente independientes. Para ello se calcula su función de autocorrelación mediante

$$r_k(\hat{a}) = \frac{\sum_1^{T-k} \left(\hat{a}_t - \bar{\hat{a}} \right) \left(\hat{a}_{t+k} - \bar{\hat{a}} \right)}{\sum_1^T \left(\hat{a}_t - \bar{\hat{a}} \right)^2} \tag{6.1.2}$$

en donde $\bar{\hat{a}} = \frac{1}{T} \sum_{t=1}^T \hat{a}_t$ es la media muestral de los T residuales.

Si los residuales están normalmente distribuidos con media cero, entonces se dirá que $r_k(\hat{a})$ es significativamente distinta de cero, con un 95% de confianza, si $r_k(\hat{a})$ es mayor

o igual a $2/T$. Esta prueba fue desarrollada por J. Durbin¹. No obstante esta prueba de significación estadística individual no es del todo válida para retrasos pequeños ($k < 3$). En los casos que el retraso es pequeño, $r_k(\hat{a})$ puede ser mucho menor que $2/T$, por lo cual si las autocorrelaciones se encuentran cercanas a estos límites, se puede considerar que se está violando este supuesto.

Otra prueba consiste en calcular el estadístico

$$Q = T(T+2) \sum_{k=1}^K (T-k)^{-1} r_k^2 \quad (6.1.3)$$

de Ljung y Box² para probar la hipótesis nula conjunta

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_K = 0$$

para los primeros K coeficientes (K debe de ser grande). Esta prueba de significación se conoce como prueba "portmanteau" y originalmente fue introducida por Box y Pierce, pero debido a que en ocasiones la aproximación lograda con su estadístico no era del todo apropiada su expresión fue modificada, por Ljung y Box, quedando el estadístico que se mostró con anterioridad, (6.1.3).

Si el modelo ajustado es apropiado el estadístico Q sigue aproximadamente una distribución asintótica χ^2 con $(T-m)$ grados de libertad, donde m es el número de parámetros estimados.

¹ J Durbin. *Testing for serial correlation in least-squares regression when some of regressors are lagged dependent variables* Econometrica, Vol 38, No. 3 (May, 1970).

² Ljung, G. Y Box, G.E.P. (1978). *On a measure of lack of fit in time series models*. Biometrika, 65, 297-303.

Por lo tanto se dirá que un modelo es inadecuado si

$$P(\chi^2(N - m) > Q) \text{ es pequeña, normalmente } 0.05.$$

En el caso de que las pruebas indiquen que los residuales no son mutuamente independientes, se recomienda graficar algunos valores de $r_k(\hat{a})$ y tratar de identificar

algún modelo *ARMA* para su representación mediante el uso de las herramientas $\left\{ r_k(\hat{a}) \right\}$ y $\left\{ \hat{\phi}_{kk}(\hat{a}) \right\}$.

Por ejemplo, supóngase que se ha estimado un modelo *ARMA*(p, q), para la serie univariante Z_t ,

$$\phi(B)Z_t = \theta(B)a_t \tag{6.1.4}$$

y la gráfica de autocorrelación entre residuales muestra que estos no son mutuamente independientes, por lo cual se sugiere ajustar un modelo *ARMA* para la serie $\{a_t\}$. A través de la *FAC* y de la *FACP* se llega a que la serie $\{a_t\}$ puede ser representada por el modelo:

$$\phi^*(B)a_t = \theta^*(B)b_t \tag{6.1.5}$$

en donde $\{b_t\}$ es un proceso de ruido blanco.

Por lo tanto, utilizando (6.1.5) en (6.1.4) se llega al siguiente modelo *ARMA*($p + p^*, q + q^*$)

$$\phi^*(B)\phi(B)Z_t = \theta^*(B)\theta(B)b_t \quad (6.1.6)$$

En donde los parámetros de $\phi(B)$ y de $\theta(B)$ serán obtenidos del modelo (6.1.4) y los parámetros de $\phi^*(B)$ y de $\theta^*(B)$ serán obtenidos de estimar el modelo (6.1.5).

Es así como se pueden ir corrigiendo las violaciones a los supuestos para quedarse con el modelo que mejor represente a la realidad.

El siguiente supuesto a verificar es la media cero de la serie $\{a_t\}$, para ello calcúlese la media de los residuales, como sigue

$$\bar{\hat{a}} = \frac{\sum_{t=1}^T \hat{a}_t}{T} \quad (6.1.7)$$

y su varianza mediante

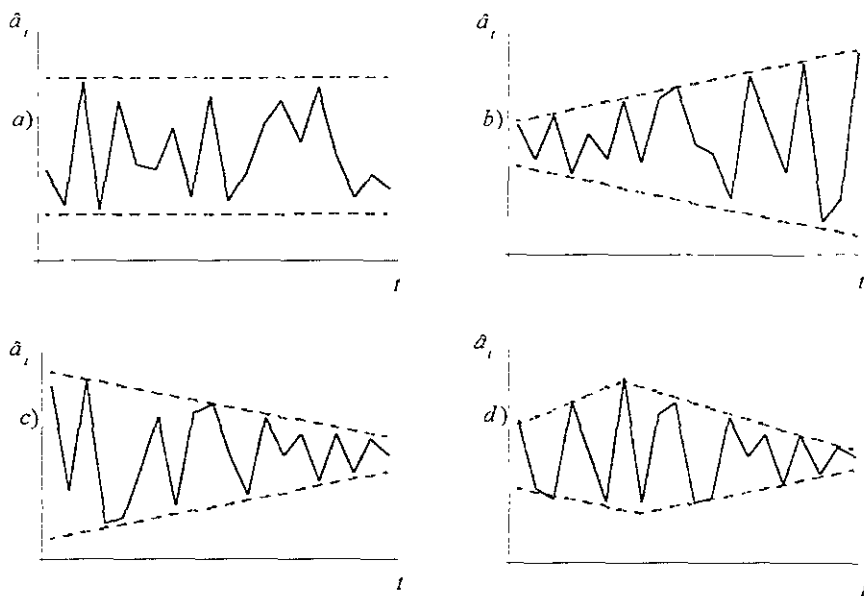
$$\hat{S}_a^2 = \frac{\sum \left(\hat{a}_t - \bar{\hat{a}} \right)^2}{T - q - p} \quad (6.1.8)$$

Debido a que se supone que \hat{a}_t sigue una distribución normal se espera que si $\bar{\hat{a}}$ es igual a cero, entonces:

$$\frac{\hat{a} - \bar{\hat{a}}}{\hat{S}_a \sqrt{T}} = \frac{\hat{a}}{\hat{S}_a \sqrt{T}} < 2$$

de donde se concluye, si la desigualdad anterior se cumple, no hay argumento para rechazar el supuesto media de $\{\hat{a}_t\}$ igual a cero de lo contrario, si la desigualdad es mayor o igual a dos se dirá que la media de los residuales es significativamente diferente de cero. Para que esta prueba sea válida es necesario que se haya concluido previamente que los residuales no están correlacionados.

El supuesto de varianza constante puede comprobarse a través de una gráfica de los residuales contra el tiempo y observar en ella si la varianza parece o no ser constante. En la *figura 6.1* se muestran los patrones que puede presentar la gráfica.



Algunos comportamientos típicos de la varianza de los residuales: a)varianza constante, b)varianza creciente, c)varianza decreciente y d)varianza no-constante.

Figura 6.1

En el caso de detectarse algún comportamiento no constante habrá que aplicar una transformación estabilizadora de varianza a la serie en cuestión.

Por último para revisar que \hat{a}_t sigue una distribución normal, se puede aplicar cualquier prueba de normalidad; histogramas, papel normal, empírica, etc.; que ayude a detectar si está o no violando este supuesto.

En el caso de que las violaciones sean muy notorias, puede aplicarse una transformación como la transformación potencia de Box-Cox que se vio en el primer capítulo.

Cabe mencionar que el supuesto de normalidad lo debe de cumplir la serie de errores aleatorios, $\{a_t\}$, por lo cual es de esperarse pequeñas violaciones al supuesto que no causaran problemas, ya que la serie de residuales $\{\hat{a}_t\}$ no tiene porque cumplir el supuesto exactamente, debido a que puede considerarse una sola realización del proceso $\{a_t\}$.

Verificación de los supuestos respecto a los errores para modelos de regresión dinámica a través del análisis de residuales

En el caso de los modelos de regresión dinámica se han considerado los supuestos utilizados para la construcción de cualquier modelo de regresión. Dichos supuestos son los siguientes:

- i) a_t es un proceso de ruido blanco,
- ii) la serie de errores aleatorios a_t es independiente de la serie explicativa x_t ,

Para comprobar que a_t es un proceso de ruido blanco se pueden aplicar las mismas pruebas de verificación que se explicaron para modelos de series de tiempo univariantes.

Ahora, para determinar si el término de error es independiente de la serie explicativa, se puede calcular su función de autocorrelación y compararse con su error estándar o bien aplicarse la prueba “portmanteau” entre la serie x_t y los residuales \hat{a}_t . Sin embargo, nótese

que si \hat{a}_t es independiente de x_t , también lo será de la serie preblanqueada α_t , pero debido a que la función de autocorrelación entre \hat{a}_t y x_t , puede no ser muy clara por la propia dependencia de x_t , es preferible trabajar con la serie preblanqueada α_t .

Para determinar si α_t es independiente de \hat{a}_t , calcúlese su función de autocorrelación $r_{\alpha\hat{a}}(k)$ y compárese con dos veces su error estándar, $\frac{2}{T}$, o bien aplíquese la prueba “portmanteau”

$$Q = T(T+2) \sum_{k=0}^K (T-k)^{-1} r_{\alpha\hat{a}}^2(k)$$

la cual puede ser aproximada por una distribución χ^2 con $(K+1) - M$ grados de libertad en donde M es el número de parámetros δ , y w , estimados en la función de transferencia

$$v(B) = \frac{w(B)}{\delta(B)}$$

En el caso de que alguna de las funciones de correlación cruzada, $r_{\alpha\hat{a}}(k)$ para ver que las series son independientes, de autocorrelación, $r_{\hat{a}}(k)$ para ver si es ruido blanco, o ambas no son nulas ¿en dónde está el error?, en la función de transferencia, en el modelo del proceso de inercia, o en ambos.

Para responder a esta pregunta, supóngase que el modelo verdadero es:

$$y_t = v(B)x_t + \psi(B)\alpha_t$$

y que se ha estimado la función de transferencia $v_0(B)$ dando como consecuencia un término de error b_t ,

$$y_t = v_0(B)x_t + \psi_0(B)b_t$$

entonces, se verifica que:

$$v(B)x_t + \psi(B)a_t = v_0(B)x_t + \psi_0(B)b_t$$

de donde

$$b_t = \frac{v(B) - v_0(B)}{\psi_0(B)} x_t + \frac{\psi(B)}{\psi_0(B)} a_t$$

a partir de aquí pueden surgir tres situaciones:

Caso 1) La función de transferencia es inadecuada es decir, $v(B)$ no es muy semejante a $v_0(B)$. En este caso, independientemente de que $\psi(B)$ sea muy similar a $\psi_0(B)$ o no, $r_b(k) \neq 0$ y $r_{ab} \neq 0$. Ya que si b_t esta correlacionada con x_t para algún retardo, también lo estará con ella misma. Por lo cual es muy importante primero establecer una adecuada función de transferencia antes de un proceso de inercia.

Caso 2) Si $v_0(B)$ es igual o muy aproximado a $v(B)$ pero $\psi(B)$ es diferente de $\psi_0(B)$ entonces el proceso de inercia es equivocado, ya que habrá autocorrelación entre los residuales, $r_b(k) \neq 0$, aunque no existirá correlación cruzada alguna entre \hat{b}_t y x_t , es decir $r_{ab} = 0$.

Caso 3) Ambos, la función de transferencia y el proceso de inercia, son correctos entonces $r_{ab} = 0$ y $r_b = 0$. El cual sería el caso más adecuado, que se pudiera encontrar.

6.1.1 Criterios de selección AIC y SBC

Un criterio de selección de modelo fue introducido por Akaike en 1973, consiste en aplicar la siguiente estadística basada en la información disponible y el número de parámetros estimados en el modelo:

$$AIC(M) = -2 \ln(\text{función de máxima verosimilitud}) + 2M \quad (6.1.1.1)$$

en donde M representa el número de parámetros en el modelo y la función de máxima verosimilitud según (5.1.15) para un modelo $ARMA(p, q)$ esta dada por:

$$L(\phi, \theta_0, \theta, \sigma_a^2) = -\frac{(T-p)}{2} \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} S(\phi, \theta_0, \theta) \quad (6.1.1.2)$$

con:

$$S(\phi, \theta_0, \theta) = \sum_{p+1}^T \alpha_i^2$$

al maximizar 6.1.1.2 con respecto a ϕ , θ_0 , θ y σ_a^2 de acuerdo a la teoría de estimación estudiada en la sección 5.1 de este capítulo se obtiene el estimador:

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}, \hat{\theta}_0, \hat{\theta})}{T-p}$$

utilizándolo en (6.1.1.2) se tiene:

$$L(\hat{\phi}, \hat{\theta}_0, \hat{\theta}, \hat{\sigma}_a^2) = -\frac{(T-p)}{2} \ln 2\pi \hat{\sigma}_a^2 - \frac{(T-p) \hat{\sigma}_a^2}{2\hat{\sigma}_a^2} =$$

$$\begin{aligned}
 &= -\frac{(T-p)}{2} \left(\ln \hat{\sigma}_a + \ln 2\pi \right) - \frac{(T-p)}{2} = \\
 &= -\frac{(T-p)}{2} \ln \hat{\sigma}_a - \frac{(T-p)}{2} (1 + \ln 2\pi) \tag{6.1.1.3}
 \end{aligned}$$

se puede prescindir del segundo término de (6.1.1.3) ya que este es constante. Por lo tanto, la estadística $AIC(M)$ queda definida como:

$$AIC(M) = (T-p) \ln \hat{\sigma}_a + 2M \tag{6.1.1.4}$$

y el mejor modelo será aquel que minimice dicha estadística.

Posteriormente en 1978, Swchwarz desarrollo un criterio también para seleccionar el modelo óptimo, éste consistente en elegir aquel que minimice la siguiente estadística:

$$SBC(M) = n \ln \hat{\sigma}_a + M \ln(T-p)$$

en donde $\hat{\sigma}_a$ es el estimador máximo verosímil de σ_a^2 y M el numero de parámetros estimados del modelo.

Tanto el criterio AIC y SBC puede ser generalizado para modelos *ARIMA* y funciones de transferencia de modelos de regresión dinámica.

6.2 Sobre los parámetros.

Además de los supuestos que se hicieron sobre los errores existen otros relacionados con los parámetros; se pueden mencionar los siguientes:

- i) *Que el modelo cumpla con el principio de parsimonia.* Esto se refiere a que el número de parámetros en el modelo no se pueda reducir. Para verificarlo hay que crear intervalos de confianza del siguiente estilo

$$\left(\hat{\theta} - 2 \sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + 2 \sqrt{\text{Var}(\hat{\theta})} \right)$$

para cada uno de los parámetros y verificar si el cero está o no dentro del intervalo. En el caso de que el cero esté considerado dentro del intervalo, indica que el parámetro puede ser omitido en el modelo. En algunos casos es evidente que el parámetro a pesar de que el cero se encuentra dentro del intervalo debe de permanecer para la buena interpretación de los datos, en este caso el parámetro deberá ser considerado en el modelo.

- ii) *Que el modelo sea admisible.* Para que el modelo cumpla con este supuesto se tiene que verificar que las raíces del polinomio de rezago $\phi_p(B)$ se encuentre fuera del círculo unitario para que la serie sea estacionaria así como verificar que las raíces de $\theta_q(B)=0$ también se encuentren fuera del círculo unitario para que el modelo sea invertible. En el caso de los modelos de regresión dinámicos también se requiere que las raíces del polinomio $\delta_v(B)$ en módulo sean menores o igual a uno para que el modelo sea estable.

- iii) *Que no haya multicolinealidad entre los parámetros.* Esto se refiere a que entre los parámetros no deben de estar altamente correlacionados, de manera que un cambio en alguno de ellos pueda compensarse mediante un cambio en alguno otro. Para ello se sugiere calcular correlaciones entre parejas de parámetros estimados.

Una manera de corregir este problema es mediante la eliminación de alguno de los parámetros; aunque si la inclusión de los parámetros –altamente correlacionados- es necesaria para una buena representación de los datos, ninguno de ellos debe ser eliminado.

7. Predicción

La última etapa sugerida por Box-Jenkins es la de utilización del modelo para los fines que fue construido; debido a que uno de estos fines, quizá el más importante, es el de la predicción de valores futuros de la variable de interés a partir del modelo construido se explicará aquí como encontrar el mejor valor óptimo de predicción .

El criterio que se usará para determinar el valor óptimo de y_{t+k} será el del error cuadrático medio mínimo, éste consiste en calcular el pronóstico \hat{y}_{t+k} , que satisfaga la siguiente condición

$$E\left[y_{t+k} - \hat{y}_{t+k}\right]^2 = \min_{\hat{y}_{t+k}} E\left[y_{t+k} - \hat{y}_{t+k}\right]^2 \quad (7.1.1)$$

Cuando no se conoce nada acerca de los valores previos de la serie, el pronóstico óptimo de y_{t+k} es su media $E(y_{t+k})$, ya que la constante que minimiza $E[y_{t+k} - c]^2$ es la esperanza de la variable.

Demostración:

$$\begin{aligned} E[y_{t+k} - c]^2 &= E[y_{t+k} - E(y_{t+k}) + E(y_{t+k}) - c]^2 = \\ &= E[y_{t+k} - E(y_{t+k})]^2 + 2[E(y_{t+k}) - c] E[y_{t+k} - E(y_{t+k})] + E[E(y_{t+k}) - c]^2 = \\ &= \text{Var}[y_{t+k}] + 2[E(y_{t+k}) - c] [E(y_{t+k}) - E(y_{t+k})] + E[E(y_{t+k}) - c]^2 = \\ &= \text{Var}[y_{t+k}] + E[E(y_{t+k}) - c]^2 \end{aligned}$$

en donde se puede observar que esta última expresión se minimiza cuando $c = E(y_{t+k})$.

Sin embargo, si se cuenta con T observaciones previas, entonces el predictor óptimo de y_{T+k} será la esperanza de y_{T+k} condicionada a los T valores observados, $y_T = (y_1, y_{T-1}, y_{T-2}, \dots)$, es decir

$$\hat{y}_{T+k} = E[y_{T+k} / y_T] \tag{7.1.2}$$

La demostración es análoga que la demostración para la función no condicionada. Además, cabe hacer notar que (7.1.2) no sólo minimiza el error cuadrático medio con respecto a la distribución condicionada, $y_{T+k} / y_T, y_{T-1}, y_{T-2}, \dots$, sino además también es el óptimo con respecto a la distribución conjunta $y_{T+k}, y_1, y_{T-1}, y_{T-2}, \dots$.

La demostración, está basada en el siguiente teorema:

La esperanza de una función $g(x, y)$ respecto a la distribución conjunta de dos variables puede obtenerse calculando la esperanza respecto a la distribución condicionada y / x , seguida de la esperanza respecto a la distribución de x

$$E_{xy}[g(x, y)] = E_x[E_{y/x}(g(x, y))] \tag{7.1.3}$$

Demostración:

$$\begin{aligned} E_{xy}[g(x, y)] &= \iint g(x, y) f(x, y) dx dy = \iint g(x, y) f(y/x) f(x) dx dy = \\ &= \iint [g(x, y) f(y/x)] f(x) dx = \int E_{y/x}(g(x, y)) f(x) dx = \\ &= E_x[E_{y/x}(g(x, y))] \end{aligned}$$

Ahora, para demostrar que $\hat{y}_{T+k} = E[y_{T+k} / y_T, y_{T-1}, y_{T-2}, \dots]$ minimiza el error cuadrático medio respecto a la distribución conjunta de (y_{T+k}, y_T) considérese otro predictor cualesquiera \tilde{y}_{T+k} . Entonces, el error cuadrático de ese nuevo predictor medio \bar{y}_{T+k} está dado por:

$$E[y_{t+k} - \bar{y}_{t+k}]^2 \tag{7.1.4}$$

el cual puede reescribirse como:

$$\begin{aligned} E[y_{t+k} - \bar{y}_{t+k}]^2 &= E\left[y_{t+k} - \hat{y}_{t+k} + \hat{y}_{t+k} - \bar{y}_{t+k}\right]^2 = \\ &= E\left[y_{t+k} - \hat{y}_{t+k}\right]^2 + E\left[\hat{y}_{t+k} - \bar{y}_{t+k}\right]^2 + 2E\left[\left(\hat{y}_{t+k} - \bar{y}_{t+k}\right)\left(y_{t+k} - \hat{y}_{t+k}\right)\right] \end{aligned} \tag{7.1.5}$$

para demostrar que el doble producto, $E\left[\left(\hat{y}_{t+k} - \bar{y}_{t+k}\right)\left(y_{t+k} - \hat{y}_{t+k}\right)\right]$ en (7.1.5) es igual a cero, se utilizará el teorema que se demostró anteriormente.

Debido a que las esperanzas en (7.1.5) se toman respecto a la distribución conjunta (y_t, y_{t+k}) entonces

$$\begin{aligned} E_{y_t, y_{t+k}}\left[\left(\hat{y}_{t+k} - \bar{y}_{t+k}\right)\left(y_{t+k} - \hat{y}_{t+k}\right)\right] &= E_{y_t}\left\{E_{y_{t+k}|y_t}\left[\left(\hat{y}_{t+k} - \bar{y}_{t+k}\right)\left(y_{t+k} - \hat{y}_{t+k}\right)\right]\right\} = \\ &= E_{y_t}\left\{\left(\hat{y}_{t+k} - \bar{y}_{t+k}\right)\left[E_{y_{t+k}|y_t}\left(y_{t+k} - \hat{y}_{t+k}\right)\right]\right\} = E\left\{\left(\hat{y}_{t+k} - \bar{y}_{t+k}\right)\left[E_{y_{t+k}|y_t}\left(y_{t+k}\right) - E_{y_{t+k}|y_t}\left(\hat{y}_{t+k}\right)\right]\right\} \\ &= E_{y_t}\left[\left(\hat{y}_{t+k} - \bar{y}_{t+k}\right)\left(\hat{y}_{t+k} - \hat{y}_{t+k}\right)\right] = 0 \end{aligned}$$

de esta manera, (7.1.5) se reduce a:

$$E[y_{t+k} - \bar{y}_{t+k}]^2 = E\left[y_{t+k} - \hat{y}_{t+k}\right]^2 + E\left[\hat{y}_{t+k} - \bar{y}_{t+k}\right]^2 \tag{7.1.6}$$

en donde se puede observar que $E[y_{T+k} - \hat{y}_{T+k}]^2$ se hará mínimo si $\hat{y} = \hat{y}_{T+k}$. Por lo tanto

$$\hat{y}_{T+k} = E[y_{T+k} / \underline{y}_T]$$

es el predictor óptimo de y_{T+k} a partir de la función de distribución conjunta $y_{T+k}, y_T, y_{T-1}, \dots$.

De manera teórica se ha demostrado que \hat{y}_{T+k} minimiza el error cuadrático medio, sin embargo se desconoce la distribución de y_{T+k} y por lo tanto \hat{y}_{T+k} también es desconocido.

De acuerdo con lo anterior se tendrá que buscar alguna manera equivalente de obtener \hat{y}_{T+k} a partir de los valores conocidos, $y_T, y_{T-1}, y_{T-2} \dots$ y de los parámetros que se estimaron en el modelo ajustado.

Se considerará primero la predicción de valores futuros a partir de un modelo $ARMA(p, q)$ posteriormente se explicará la predicción para series de tiempo no necesariamente estacionarias y cuyo modelo ajustado es un modelo $ARIMA(p, d, q)$, finalmente se hará la generalización para modelos de regresión dinámica.

7.1 Predicción a partir de un modelo $ARMA(p, q)$.

Si se considera a la serie estacionaria y_t , para la cual fue ajustado el siguiente modelo $ARMA(p, q)$

$$\phi_p(B)y_t = \theta_0 + \theta_q(B)\alpha_t \tag{7.1.7}$$

en donde por simplicidad, pero sin pérdida de generalidad, se puede suponer que $\theta_0 = 0$. Entonces (7.1.7), por ser un modelo estacionario, puede ser reescrito mediante una representación de media móvil

$$\begin{aligned} y_t &= \psi(B)a_t \\ &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \end{aligned} \tag{7.1.8}$$

en donde

$$\psi(B) = \frac{\theta_q(B)}{\phi_p(B)} = \sum_{j=0}^{\infty} \psi_j B^j \tag{7.1.9}$$

con $\psi_0 = 1$.

Si a partir de t , se desea pronosticar a y_{t+k} , un pronóstico cualquiera puede obtenerse como una combinación lineal de los valores de la serie $\{y_t\}$ y de los parámetros estimados del modelo elegido, éste pronóstico se denotará como $\tilde{y}_t(k)$, en donde t denota el origen y k el número de pronósticos adelantados a partir del origen.

Supóngase que en el tiempo $t=T$ se cuenta con las observaciones $y_T, y_{T-1}, y_{T-2}, \dots$. Un pronóstico que se obtenga como combinación lineal de $y_T, y_{T-1}, y_{T-2}, \dots$ se puede obtener por consiguiente de la combinación lineal de $a_T, a_{T-1}, a_{T-2}, \dots$, de acuerdo a (7.1.8) como sigue:

$$\tilde{y}_t(k) = \psi_k^* a_T + \psi_{k+1}^* a_{T-1} + \psi_{k+2}^* a_{T-2} + \dots \tag{7.1.10}$$

en donde ahora para obtener el pronóstico óptimo, denotado como $\hat{y}_T(k)$, se tienen que determinar los valores $\psi_k^*, \psi_{k+1}^*, \psi_{k+2}^*, \dots$ de manera que hagan mínimo $E\left(y_{T+k} - \hat{y}_T(k)\right)^2$

Para este fin, desarróllese el error cuadrático medio de la siguiente manera:

$$\begin{aligned} E\left(y_{T+k} - \hat{y}_T(k)\right)^2 &= \\ &= E\left(y_T^2(k)\right) - 2E\left(y_{T+k} \hat{y}_T(k)\right) + E\left(\hat{y}_T(k)^2\right) \\ &= E\left(\sum_{i=0}^{\infty} \psi_i a_{T+k-i} \cdot \sum_{j=0}^{\infty} \psi_j a_{T+k-j}\right) - 2E\left(\sum_{i=0}^{\infty} \psi_i a_{T+k-i} \cdot \sum_{j=0}^{\infty} \psi_{k+j}^* a_{T-j}\right) + E\left(\sum_{i=0}^{\infty} \psi_{k+i}^* a_{T-i} \cdot \sum_{j=0}^{\infty} \psi_{k+j}^* a_{T-j}\right) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E\left(\psi_i a_{T+k-i} \cdot \psi_j a_{T+k-j}\right) - 2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E\left(\psi_i a_{T+k-i} \cdot \psi_{k+j}^* a_{T-j}\right) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E\left(\psi_{k+i}^* a_{T-i} \cdot \psi_{k+j}^* a_{T-j}\right) \end{aligned}$$

debido a que $E(a_{t+i} \cdot a_{t+j}) = 0$ para $i \neq j$, se tiene que

$$\begin{aligned} E\left(y_{T+k} - \hat{y}_T(k)\right)^2 &= \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2 - 2\sigma_a^2 \sum_{i=k}^{\infty} \psi_i \psi_i^* + \sigma_a^2 \sum_{i=0}^{\infty} \psi_{k+i}^{*2} = \\ &= \sigma_a^2 \sum_{i=0}^{k-1} \psi_i^2 + \sigma_a^2 \left[\sum_{i=k}^{\infty} \psi_i^2 - 2 \sum_{i=k}^{\infty} \psi_i \psi_i^* + \sum_{i=k}^{\infty} \psi_i^{*2} \right] = \\ &= \sigma_a^2 \sum_{i=0}^{k-1} \psi_i^2 + \sigma_a^2 \sum_{i=k}^{\infty} (\psi_i - \psi_i^*)^2 \end{aligned} \tag{7.1.11}$$

Se puede observar que el mínimo de la expresión (7.1.11) se obtiene haciendo $\psi_i = \psi_i^*$, de manera que, el error cuadrático medio estará dado por:

$$E\left[y_{T+k} - \hat{y}_T(k)\right]^2 = \sigma_a^2 \sum_{i=0}^{k-1} \psi_i^2$$

de acuerdo con (7.1.10)

$$\begin{aligned} \hat{y}_T(k) &= \psi_k a_T + \psi_{k+1} a_{T-1} + \psi_{k+2} a_{T-2} + \dots = \\ &= \sum_{j=k}^{\infty} \psi_j a_{T+k-j} \end{aligned} \tag{7.1.12}$$

Por otro lado, de acuerdo a la definición de esperanza condicional, se tiene que

$$E(a_{T-j} / y_T, y_{T-1}, y_{T-2}, \dots) = \begin{cases} a_{T-j} & \text{si } j \geq 0 \\ 0 & \text{si } j < 0 \end{cases}$$

de donde se sigue que la esperanza de y_{T+k} dada toda la información hasta el instante T es igual a:

$$\begin{aligned} E(y_{T+k} / y_T) &= E\left(\sum_{j=0}^{\infty} \psi_j a_{T+k-j} / y_T\right) = \\ &= E\left(\sum_{j=0}^{k-1} \psi_j a_{T+k-j} / y_T\right) + E\left(\sum_{j=k}^{\infty} \psi_j a_{T+k-j} / y_T\right) = \\ &= \sum_{j=k}^{\infty} \psi_j a_{T+k-j} \end{aligned} \tag{7.1.13}$$

Y por lo tanto, de comparar (7.1.12) con (7.1.13) se puede concluir que

$$y_T(k) = E(y_{T+k} / y_T, y_{T-1}, \dots) \tag{7.1.14}$$

El error del pronóstico (7.1.14) estará dado por

$$\begin{aligned}
 e_T(k) &= y_{T+k} - \hat{y}_T(k) \\
 &= \sum_{j=0}^{\infty} \psi_j a_{T+k-j} - \sum_{j=k}^{\infty} \psi_j a_{T+k-j} = \sum_{j=0}^{k-1} \psi_j a_{T+k-j}
 \end{aligned}
 \tag{7.1.15}$$

de donde se puede observar que

$$E(e_T(k) / \underline{y}_T) = 0 \quad \text{y} \quad \text{Var}(e_T(k) / \underline{y}_T) = \sum_{j=0}^{k-1} \psi_j^2 \sigma_a^2
 \tag{7.1.16}$$

es decir que los pronósticos $\hat{y}_T(k)$ son insesgados además de que se cumple la siguiente desigualdad:

$$\text{Var}(e_T(k) / \underline{y}_T) - \text{Var}(e_T(k-1) / \underline{y}_T) = \sum_{j=0}^{k-1} \psi_j^2 \sigma_a^2 - \sum_{j=0}^{k-2} \psi_j^2 \sigma_a^2 = \psi_{k-1}^2 \sigma_a^2 \geq 0 \quad \text{para } k \geq 1$$

es decir, que entre más alejados estén los pronósticos del origen T , menor será su precisión debido a que la varianza es más grande.

Con $k=1$ en (7.1.15) se obtiene

$$e_T(1) = y_{T+1} - \hat{y}_T(1)$$

$$= \sum_{j=0}^{\infty} \psi_j a_{t+1-j} - \sum_{j=1}^{\infty} \psi_j a_{t+1-j} = a_{t+1} \quad (7.1.17)$$

lo cual significa que los errores un período hacia delante no están correlacionados. A partir de $k \geq 2$ los errores de pronóstico se encuentran correlacionados, ya que como indica la expresión (7.1.15) la serie $\{e_t(k)\}$ sigue un proceso $MA(k-1)$.

El pronóstico $\hat{y}_t(k)$ en (7.1.12) está dado en términos de la serie $\{a_t\}$, para tenerlo en términos de los valores observados previos, y_t , considérese conocido el modelo $\phi(B)y_t = \theta(B)a_t$, es decir una vez que el modelo ha pasado con éxito las etapas anteriores de Box-Jenkins, — identificación, estimación y verificación—, entonces:

$$\begin{aligned} \hat{y}_t(k) &= E(y_{t+k} / \underline{y}_t) = \\ &= E[\phi_1 y_{t+k-1} + \dots + \phi_p y_{t+k-p} + a_{t+k} - \theta_1 a_{t+k-1} - \dots - \theta_q a_{t+k-q} / \underline{y}_t] = \\ &= \phi_1 E[y_{t+k-1} / \underline{y}_t] + \dots + \phi_p E[y_{t+k-p} / \underline{y}_t] + E[a_{t+k} / \underline{y}_t] - \\ &\quad - \theta_1 E[a_{t+k-1} / \underline{y}_t] - \dots - \theta_q E[a_{t+k-q} / \underline{y}_t] \quad (7.1.18) \end{aligned}$$

donde

$$E[y_{t+k-j} / \underline{y}_t] = \begin{cases} y_{t+k-j} & \text{si } j \geq k \\ \hat{y}_t(k-j) & \text{si } j < k \end{cases} \quad (7.1.19)$$

$$E[a_{t+k-j} / \underline{y}_t] = \begin{cases} a_{t+k-j} & \text{si } j \geq k \\ 0 & \text{si } j < k \end{cases} \quad (7.1.20)$$

Por (7.1.16) y (7.1.19) se tiene que

$$E[a_{t+k-1} / y_T] = \begin{cases} y_{t+k-1} - \hat{y}_{t+k-1} & \text{si } j \geq k \\ 0 & \text{si } j < k \end{cases} \quad (7.1.21)$$

En donde ahora se tiene una expresión de los errores aleatorios no observables, en términos de los valores observados de la serie $\{y_t\}$ y de sus pronósticos.

7.2 Pronóstico con modelos ARIMA

Supóngase, al igual que en la sección anterior, que se cuenta con T observaciones pero que en este caso y_t sigue un modelo ARIMA(p, d, q):

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_h y_{t-h} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (7.2.1)$$

el operador $\varphi_h(B) = \phi_p \nabla^d$ con $h = p + d$, y $d > 0$.

La predicción óptima, de y_{t+k} será la esperanza condicionada de esta variable dada toda la información hasta el instante T , $y_T = (y_t, y_{t-1}, \dots)$. De esta manera

$$\begin{aligned} \hat{y}_t(k) &= E[y_{t+k} / y_T] = E[\varphi_1 y_{t+k-1} + \dots + \varphi_h y_{t+k-h} + a_{t+k} - \theta_1 a_{t+k-1} - \dots - \theta_q a_{t+k-q} / y_T] \\ &= \varphi_1 E[y_{t+k-1} / y_T] + \dots + \varphi_h E[y_{t+k-h} / y_T] + \\ &\quad + E[a_{t+k} / y_T] - \theta_1 E[a_{t+k-1} / y_T] - \dots - \theta_q E[a_{t+k-q} / y_T] \end{aligned} \quad (7.2.2)$$

en donde al igual que en la sección anterior se cumple lo siguiente

$$E[y_{t+k-j} / y_T] = \begin{cases} y_{t+k-j} & \text{si } j \geq k \\ \hat{y}_t(k-j) & \text{si } j < k \end{cases}$$

Además

$$(7.2.3)$$

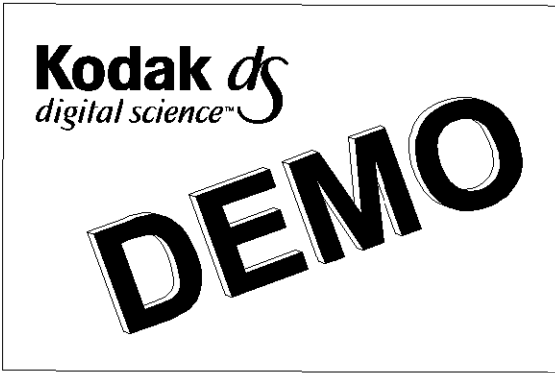
$$E[a_{t+k-j} / y_t] = \begin{cases} a_{t+k-j} & \text{si } j \geq k \\ 0 & \text{si } j < k \end{cases}$$

Tomando en cuenta (7.2.3) en (7.2.2) se puede observar que para $k > q$, $\hat{y}_t(k)$ dependerá sólo de la parte autorregresiva del modelo,

$$\hat{y}_t(k) = \varphi_1 E[y_{t+k-1} / y_t] + \dots + \varphi_h E[y_{t+k-h} / y_t]$$

Para $k > q$

las predicciones



$k > q$

en donde se
este último :

; y no sobre T , ya que

En esta sección se ha hecho la generalización de lo que se analizó en la sección anterior, sin embargo si a la serie y_t se le aplican las transformaciones necesarias para tener una serie estacionaria $\nabla^d T(y_t)$, la predicción de valores utilizando todo lo expuesto en la sección anterior puede ser factible.

7.3 Pronóstico utilizando modelos de regresión dinámica.

Sean x_t y y_t series estacionarias, relacionadas mediante el siguiente modelo de regresión dinámico:

$$(1 - B)^d y_t = \frac{w(B)}{\delta(B)} (1 - B)^d B^h x_t + \frac{\theta(B)}{\phi(B)} a_t \tag{7.3.1}$$

Para obtener el pronóstico de y_{t+k} a partir de (7.3.1) se siguen básicamente los pasos de la sección anterior.

Es así como el valor óptimo de y_{t+k} está dado por la esperanza de la variable, condicionada a la información que se tenga hasta el momento.

$$\hat{Y}_t(k) = E[Y_{t+k} / Y_t, \dots, Y_1; X_t, \dots, X_1] \quad (7.3.2)$$

Obsérvese que no sólo se requerirá ajustar el modelo de regresión dinámico para establecer la relación entre la serie de entrada y la serie respuesta, sino además es necesario construir un modelo univariante para la serie explicativa $\{X_t\}$ debido a que se necesita conocer su comportamiento hasta el instante de tiempo $t - b$, —puede construirse siguiendo los pasos que se han venido desarrollando para series de tiempo univariantes—. Supóngase que este modelo es:

$$\phi_x(B)x_t = \theta_x(B)\alpha_t \quad (7.3.3)$$

donde las raíces de $\phi_x(B) = 0$ se encuentran fuera del círculo unitario y α_t es un proceso de ruido blanco con media cero y varianza σ_α^2 .

Considerando (7.3.3) en (7.3.1) se tiene

$$y_{t+k} = \frac{w(B)\theta_x(B)}{\delta(B)\phi_x(B)} B^k \alpha_t + \frac{\theta(B)}{\phi(B)(1-B)^d} a_t \quad (7.3.4)$$

Sea

$$u(B) = \frac{w(B)\theta_x(B)}{\delta(B)\phi_x(B)} B^h = u_0 + u_1 B + u_2 B^2 + \dots \quad (7.3.5)$$

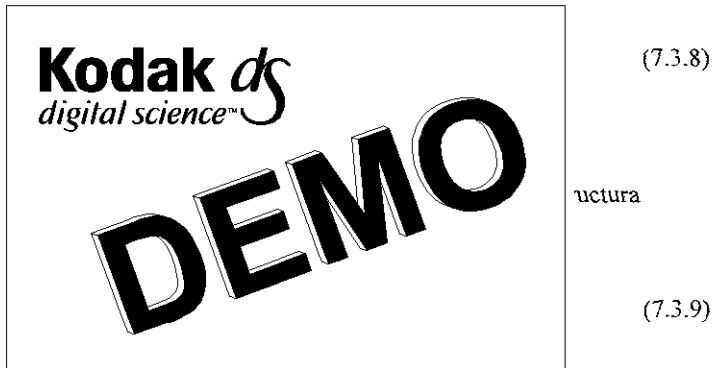
y

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \dots \quad (7.3.6)$$

de manera que Y_t puede ser expresado como:

$$\begin{aligned} y_t &= u(B)\alpha_t + \psi(B)\alpha_t \\ &= \sum_{j=0}^{\infty} u_j \alpha_{t-j} + \sum_{j=0}^{\infty} \psi_j \alpha_{t-j} \end{aligned} \quad (7.3.7)$$

y una vez teniendo la expresión (7.3.7); Y_{t+k} puede expresarse como:



Es así como, se esperz

uctura

(7.3.9)

en donde se deberán determinar los valores de u_{k+1}^* y de ψ_{k+1}^* , $j = 0, 1, \dots$, que hagan mínimo $E\left[y_{t+j} - \hat{y}_t(k)\right]^2$. Para ello obsérvese que

$$\begin{aligned}
 y_{t+k} - \hat{y}_t(k) &= \sum_{j=0}^{\infty} u_j \alpha_{t+k-j} + \sum_{j=0}^{\infty} \psi_j a_{t+k-j} - \sum_{j=0}^{\infty} u_{k+j}^* \alpha_{t-j} - \sum_{j=0}^{\infty} \psi_{k+j}^* a_{t-j} = \\
 &= \sum_{j=0}^{k-1} [u_j \alpha_{t+k-j} + \psi_j a_{t+k-j}] - \sum_{j=0}^{\infty} [u_{k+j}^* - u_{k+j}] \alpha_{t-j} - \sum_{j=0}^{\infty} [\psi_{k+j}^* - \psi_{k+j}] a_{t-j}
 \end{aligned}$$

ya que α_i y a_i son independientes con media cero y varianza σ_α^2 y σ_a^2 respectivamente, se sigue que

$$\begin{aligned}
 E \left[y_{t+k} - \hat{y}_t(k) \right]^2 &= \sum_{j=0}^{k-1} (\sigma_\alpha^2 u_j^2 + \sigma_a^2 \psi_j^2) + \sum_{j=0}^{\infty} \sigma_\alpha^2 (u_{k+j}^* - u_{k+j})^2 + \\
 &\quad + \sum_{j=0}^{\infty} \sigma_a^2 (\psi_{k+j}^* - \psi_{k+j})^2
 \end{aligned} \tag{7.3.10}$$

el cual se minimiza cuando $u_{k+j}^* = u_{k+j}$ y $\psi_{k+j}^* = \psi_{k+j}$, para $j = 0, 1, 2, \dots$

Por lo tanto, el predictor óptimo de y_{t+k} es igual a:

$$\hat{y}_t(k) = E(y_{t+k} / y_t, x_t) = \sum_{j=0}^{\infty} u_{k+j} \alpha_{t-j} + \sum_{j=0}^{\infty} \psi_{k+j} a_{t-j}$$

Para calcular el valor de predicción de Y_{T+k} a partir de T observaciones conocidas, se puede reescribir el modelo general (7.3.1) de la siguiente manera:

$$c(B)y_t = d(B)x_{t-h} + e(B)a_t \tag{7.3.11}$$

en donde

$$c(B) = \delta(B)\phi(B)(1-B)^d$$

$$= 1 - c_1 B - c_2 B^2 - \dots - c_{p+a+d} B^{p+a+d}$$

$$d(B) = w(B)\phi(B)(1-B)^d$$

$$= d_0 - d_1 B - d_2 B^2 - \dots - d_{p+m+d} B^{p+m+d}$$

y

$$e(B) = \delta(B)\theta(B)$$

$$= 1 - e_1 B - e_2 B^2 - \dots - e_{q+a} B^{q+a}$$

De manera que si se cuenta con T observaciones, el predictor óptimo de y_{t+k} se obtendrá de la siguiente manera:

$$\hat{y}_t(k) = E[y_{t+k} / y_t, x_t] =$$

$$c_1 \hat{y}_t(k-1) + \dots + c_{p+a+d} \hat{y}_t(k-p-a-d) +$$

$$+ d_0 \hat{x}_t(k-b) - \dots - d_{p+m+d} \hat{x}_t(k-b-p-m-d) +$$

$$+ E[a_t / y_t, x_t] - e_1 E[a_{t-1} / y_t, x_t] - \dots - e_{q+a} E[a_{t-q-a} / y_t, x_t].$$

en donde

$$\hat{y}_T(j) = \begin{cases} y_{t+j}, & j \leq 0 \\ \hat{y}_T(j), & j > 0 \end{cases}$$

$$\hat{x}_T(j) = \begin{cases} x_{t+j}, & j \leq 0 \\ \hat{x}_T(j), & j > 0 \end{cases}$$

$$E[a_{T+j} / y_T, x_T] = \begin{cases} y_{T+j} - \hat{y}_{T+j}, & j \leq 0 \\ 0, & j > 0 \end{cases}$$

El error del pronóstico esta dado por:

$$\begin{aligned} e_T(k) &= y_{T+k} - \hat{y}_T(k) = \\ &= \sum_{j=0}^{\infty} u_j \alpha_{T+j} + \sum_{j=0}^{\infty} \psi_j a_{T+k-j} - \sum_{j=0}^{\infty} u_{k+j} \alpha_{T-j} - \sum_{j=0}^{\infty} \psi_{k+j} a_{T+j} = \\ &= \sum_{j=0}^{k-1} u_j \alpha_{T+k+j} + \sum_{j=0}^{k-1} \psi_j a_{T+k-j} \end{aligned}$$

con:

$$E[e_T(k)] = 0 \quad \text{y} \quad \text{Var}[e_T(k)] = \sigma_u^2 \sum_{j=0}^{k-1} u_j^2 + \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2 \quad (7.3.12)$$

7.4 Intervalos de confianza

Es importante que al hacer el pronóstico se obtengan sus limites de confianza. El error del pronóstico de $\hat{y}_T(k)$ como se vio anteriormente esta dado por:

$$e_T(k) = y_{T+k} - \hat{y}_T(k) = \sum_{j=0}^{k-1} \psi_j a_{T+k-j} \quad (7.4.1)$$

para los casos de modelos de series univariantes, en el caso de predicción con modelos de regresión dinámica el error del pronóstico es:

$$e_i(k) = \sum_{j=0}^{k-1} u_j \alpha_{1+k-j} + \sum_{j=0}^{k-1} \psi_j \alpha_{j+k-j} \quad (7.4.2)$$

haciendo uso del supuesto^o de que $a_t \approx N[0, \sigma_a^2]$ y $\alpha_t \approx N[0, \sigma_\alpha^2]$ para toda t implica que

$$e_i(k) \approx N[0, \text{Var}(e_i(k))] \quad (7.4.2)$$

en donde de acuerdo a (7.1.16), para los modelos univariantes

$$\text{Var}(e_i(k)) = \sum_{j=0}^{k-1} \psi_j^2 \sigma_a^2 \quad (7.4.3)$$

y para los modelos de regresión dinámica:

$$\text{Var}[e_T(k)] = \sigma_a^2 \sum_{j=0}^{k-1} u_j^2 + \sigma_u^2 \sum_{j=0}^{k-1} \psi_j^2 \quad (7.4.4)$$

Esto permite obtener los límites de confianza cuando $\{\psi_1, \psi_2, \dots, \psi_{k-1}\}$ y $\{u_0, u_1, \dots, u_{k-1}\}$ son conocidos. En general estos parámetros son desconocidos y se estiman con las observaciones muestrales. Sin embargo a pesar de que (7.4.3) es sólo una aproximación se pueden construir intervalos del $100(1 - \alpha)\%$ de confianza, dados por:

$$\hat{y}_T(k) \pm z_{\alpha/2} \left(\sum_{j=0}^{k-1} \psi_j^2 \right)^{1/2} \sigma_a \quad (7.4.5)$$

$$\hat{y}_T(k) \pm z_{\alpha/2} \left(\sum_{j=0}^{k-1} u_j^2 \right)^{1/2} \sigma_a + \left(\sum_{j=0}^{k-1} \psi_j^2 \right)^{1/2} \sigma_u \quad (7.4.6)$$

¹ El supuesto de normalidad se tuvo que haber verificado en la etapa correspondiente.

donde $z_{\alpha/2}$ es el punto porcentual tal que $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$ si $Z \approx N(0,1)$. El intervalo de confianza (7.4.5) se utiliza en modelos de series de tiempo univariantes y (7.4.6) para pronósticos originados de modelos de regresión dinámica.

7.5 Adaptación de pronósticos

El pronóstico $\hat{y}_T(1)$ deja de tener importancia en cuanto se conoce el valor real y_{T+1} de la serie, al igual que los pronósticos subsecuentes $\hat{y}_T(2)$, $\hat{y}_T(3)$, $\hat{y}_T(4)$, ... , ya que ahora es deseable tener los pronósticos a partir de $T+1$ en lugar de T . Para evitar que los pronósticos se vuelvan obsoletos se pueden ir adaptando al modelo las nuevas observaciones conocidas, aprovechando de esta manera toda la información histórica con que se cuenta.

Para ello supóngase que se obtuvo el pronóstico $\hat{y}_T(k)$ a partir de T observaciones y del modelo $ARIMA(p, d, q)$. De acuerdo con (7.1.12) se tiene que el pronóstico escrito en su representación general es:

$$\hat{y}_T(k) = \sum_{j=k}^{\infty} \psi_j a_{T+k-j} \quad (7.5.1)$$

ahora supóngase que se tiene el pronóstico a partir de una nueva información, es decir a partir de $T+1$:

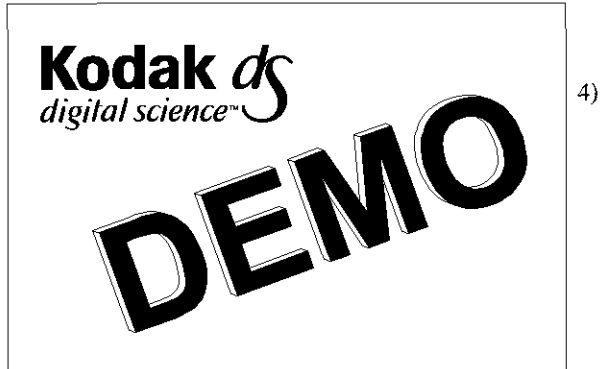
$$\hat{y}_{T+1}(k-1) = \sum_{i=k-1}^{\infty} \psi_i a_{T+k-i} \quad (7.5.2)$$

Al restar las dos expresiones anteriores, se tiene:

$$\hat{y}_{j+1}(k-1) - \hat{y}_j(k) = \psi_{k-1} \alpha_{j+1} \quad (7.5.3)$$

de esta manera, cuando se tenga el valor real de la serie y_{j+1} y se calcule el error que hubo entre lo que se pronosticó y su valor real se obtendrá:

mediante el cual se podrán adaptar:



En el caso de modelos de regresión dinámica el procedimiento es muy parecido exceptuando porque en este caso, como es de suponerse, se debe de actualizar los valores observados tanto para la serie independiente como para la serie dependiente.

Los pronósticos para los modelos de regresión dinámica de acuerdo a la expresión (7.3.9) pueden ser obtenidos de:

$$\hat{y}_j(k) = \sum_{j=0}^{\infty} u_{k+j} \alpha_{j-j} + \sum_{j=0}^{\infty} \psi_{k+j} \alpha_{j+1}$$

de donde al igual que (7.5.3) se obtiene

$$\hat{y}_{j+1}(k-1) - \hat{y}_j(k) = u_{k-1} \alpha_{j+1} + \psi_{k-1} \alpha_{j+1} \quad (7.5.5)$$

y una vez observados los valores de la serie x_{T+1} así como y_{T+1} se calcularán los errores entre el pronóstico y el valor real de la siguiente manera:

$$\alpha_{T+1} = x_{T+1} - \hat{x}_T(1) \quad \text{y} \quad a_{T+1} = y_{T+1} - \hat{y}_T(1) \quad (7.5.6)$$

que serán utilizados para la adaptación de los valores conocidos para el calculo del pronóstico futuros:

$$\hat{y}_{T+1}(k-1) = \hat{y}_T(k) + u_{k-1}\alpha_{T+1} + \psi_{k-1}a_{T+1} \quad (7.5.7)$$

Con esto concluye la parte teórica de construcción de modelos de regresión dinámica. Con todos estos elementos se espera que se pueda analizar, construir un modelo y pronosticar valores futuros para una serie, a través de un modelo univariante de series de tiempo o mediante un modelo de regresión dinámica.

Capítulo 3

Aplicación de modelos de regresión dinámica.

1. Introducción

En este capítulo se ejemplificará con un caso de la economía mexicana la teoría aportada en capítulos anteriores, la finalidad es que a través de él se puedan aclarar los conceptos, la utilidad y por supuesto el alcance de estos modelos.

Como se vio en el capítulo anterior los modelos de regresión dinámica tienen la capacidad de explicar una variable a través del comportamiento histórico de ella misma y de otra variable cuya influencia sobre la primera se considera relevante, esta cualidad permitirá, por ejemplo, exhibir el impacto nocivo que causa un aumento general de precios de bienes de consumo y de factores productivos, conocido como inflación, sobre el desarrollo de otras variables económicas importantes como es el caso de la tasa de interés.

La diferencia entre la tasa de rendimiento que se obtiene de una inversión y la tasa de inflación se denomina *tasa real* y puede ser positiva o negativa. Es evidente que para incentivar el ahorro, el rendimiento que ofrece un país debe de rebasar la tasa de inflación

por el periodo correspondiente. En México, la tasa líder es la tasa de rendimiento a 28 días de los CETES porque es el instrumento de deuda más líquido del mercado y tiene la garantía del gobierno federal, considerándose así “*la tasa carente de riesgo*” del mercado financiero mexicano. Esta tasa tiene una estrecha relación con la inflación ya que el gobierno puede influir en ella permitiendo así reaccionar ante los cambios inflacionarios. La tasa de interés que se utilizará para el análisis será la *tasa de interés real* obtenida mediante la diferencia de la tasa de inflación menos las tasa de CETES a 28 días. lo que también se conoce como tasa de interés deflactada al Índice Nacional de Precios al Consumidor y obtenida mediante la fórmula:

$$TR_t = \left(\frac{1 + TC_t}{1 + I_t} \right) - 1$$

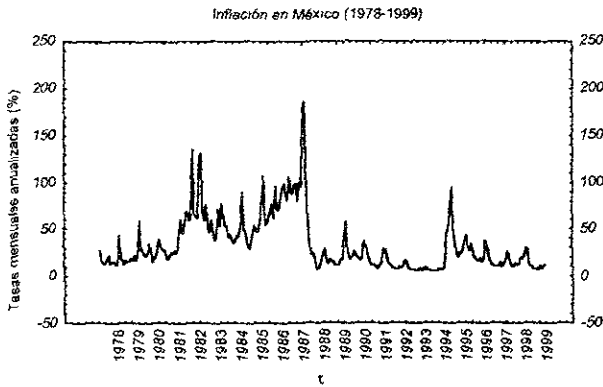
en donde TR es la tasa mensual de interés real, TC representa a la tasa de CETES a 28 días e I es la tasa de inflación mensual.

2. Modelo univariante para la serie inflación en México

El análisis consistirá en crear un modelo univariante para la serie inflación en México, durante el periodo Enero de 1978 a Diciembre de 1999 con un total de 264 observaciones, que ayude a identificar la función de transferencia para la serie tasas de interés reales de acuerdo a los pasos de construcción de modelos de regresión dinámica explicados en el capítulo anterior.

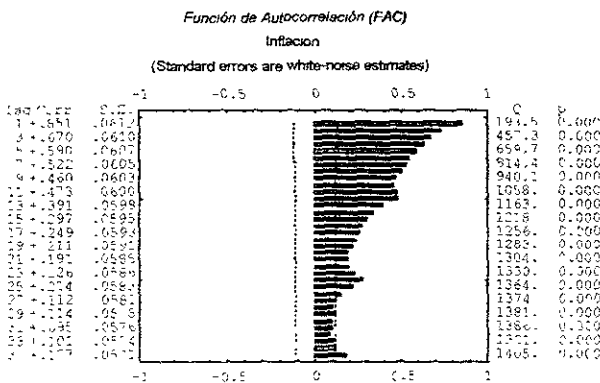
2.1 Identificación del modelo

Siguiendo los pasos de identificación univariante se presenta a continuación la gráfica de la serie original (*gráfica 3.1*) en donde se puede observar de primera instancia una clara inconsistencia en la varianza además de presentar ciclos estacionales de duración anual teniendo como consecuencia una media no constante.

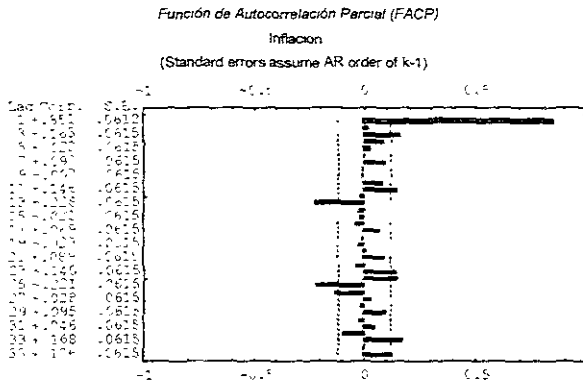


Gráfica 3.1. Tasas de inflación en México (1978-1999)

Las inconsistencias de estacionalidad se confirman al analizar la *función de autocorrelación* que presenta un coeficiente notablemente mayor en los rezagos múltiples de 12 y un lento decrecimiento en el resto de los *rezagos* (ver gráfica 3.2). Asimismo en la *función de autocorrelación parcial* se observa un coeficiente en el primer rezago muy grande para corresponder a un modelo estacionario (ver gráfica 3.3).



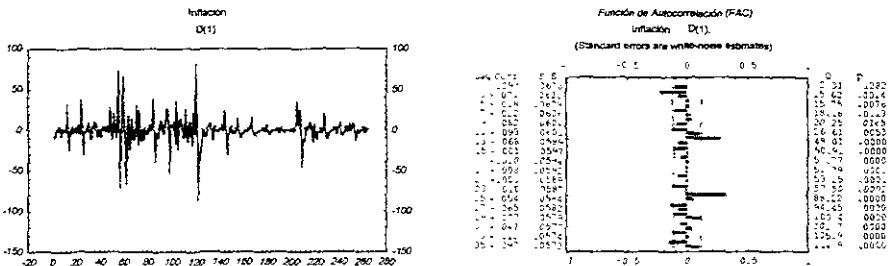
Gráfica 3.2. Función de autocorrelación simple (FAC)



Gráfica 3.3. Función de autocorrelación parcial (FACP)

Debido a las características que presenta la serie original se procederá a buscar la manera de estabilizarla a través de transformaciones, diferenciaciones y desestacionalidad.

Las diferenciaciones y/o la aplicación del operador estacional crea valores negativos por lo cual se recomienda primero aplicar la transformación estabilizadora de varianza. Sin embargo, habrá que comprobar primero que la transformación es requerida y que la inestabilidad de la varianza no se debe únicamente a la media no constante que presenta la serie. A continuación se aplicará una primera diferenciación a la serie original para ver si de esta manera se estabiliza la media y por ende la varianza (ver gráfica 3.4)



Gráfica 3.4 serie $\nabla X_t = (1 - B)X_t$

En la gráfica 3.4 se nota una estabilización en la media no siendo así en la varianza lo cual se confirma al observar la *FAC* ya que no presenta ningún patrón de decrecimiento y por el contrario en los últimos rezagos se nota un crecimiento de los coeficientes conforme aumenta el rezago. Dado que la inestabilidad de la varianza no sólo es estacional es inútil intentar diferenciar las observaciones anuales utilizando el operador ∇^{12} para tratar de estabilizar la serie.

Por lo tanto, como se había mencionado anteriormente es necesario la aplicación de una transformación potencia⁶ la cual consiste en elegir un valor para λ de tal manera que satisfaga la relación:

$$\sigma_t / \mu_t^{1-\lambda} = \text{constante} \quad \text{para } t = 1, 2, 3, \dots, T$$

en donde σ_t y μ_t representan la desviación estándar y la media de la variable X_t , respectivamente y T es el número de observaciones disponibles de la serie $\{X_t\}$.

Ya que en cada momento t se tiene solamente una observación de X_t , y por lo tanto no es posible obtener una estimación de σ_t y de μ_t , para cada t se sugiere dividir a las N observaciones de la serie en H grupos que contengan $R = (T - n)/H$ observaciones contiguas cada uno, dejando fuera de los cálculos a un total de n observaciones ($0 \leq n \leq R$) ya sean del principio o del final de la serie; lo importante es que cada grupo tenga el mismo número de observaciones para calcular estimaciones de la media y de la desviación estándar dentro de cada uno de ellos; de esta forma se tendrán H parejas de valores comparables con las cuales puede realizarse el siguiente cuadro comparativo:

⁶ Ver sección 5.2 del primer capítulo

Cálculos para seleccionar la potencia λ

Grupo	-1	-0.5	0	0.5	1
1	$S_1 X_1^2$	$S_1 X_1^{1.5}$	$S_1 X_1$	$S_1 X_1^{0.5}$	S_1
2	$S_1 X_2^2$	$S_1 X_2^{1.5}$	$S_1 X_2$	$S_1 X_2^{0.5}$	S_2
⋮					
h	$S_1 X_h^2$	$S_1 X_h^{1.5}$	$S_1 X_h$	$S_1 X_h^{0.5}$	S_h
⋮					
H	$S_1 X_H^2$	$S_1 X_H^{1.5}$	$S_1 X_H$	$S_1 X_H^{0.5}$	S_H

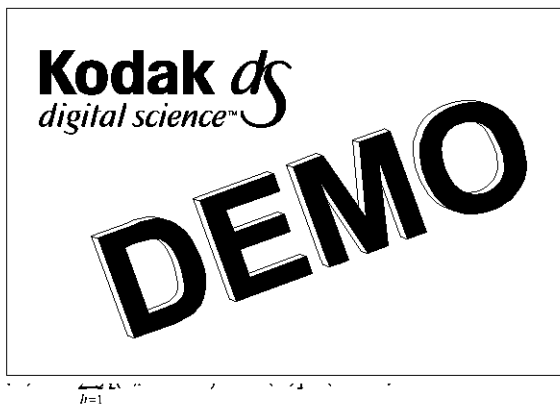
Coeficiente de Variación	$CV(-1)$	$CV(-0.5)$	$CV(0)$	$CV(0.5)$	$CV(1)$
--------------------------	----------	------------	---------	-----------	---------

en donde, si $X_{h,r}$ es la r -ésima observación del grupo H ,

$$X_h = \sum_{r=1}^R X_{h,r} \quad R, \quad S_h = \sum_{r=1}^R (X_{h,r} - X_h)^2 \quad (R-1) \tag{2.1.1}$$

además el coeficiente de media $M(\lambda)$, es decir

$$CV(\lambda) = de(\lambda)$$



sobre la

$$\tag{2.1.2}$$

$$\tag{2.1.3}$$

Ya que se desea que la relación entre la media y la varianza sea constante, se elegirá la potencia que proporcione el mínimo coeficiente de variación, debido a que ese valor satisfecerá de manera aproximada la relación:

(2.1.4)

$$\frac{S_h}{X_h^{1-\lambda}} = \text{constante}, \quad \text{para } h=1, \dots, H$$

y así la serie transformada resultante será

$$T(Z_t) = \begin{cases} Z_t^\lambda & \text{si } \lambda \neq 0 \\ \ln(Z_t) & \text{si } \lambda = 0 \end{cases}$$

Debe tenerse en cuenta que este método está basado en aproximaciones y por lo tanto conviene utilizarlo para sugerir una transformación potencia más no determinarla.

Ya que la serie inflación muestra un marcado comportamiento estacional, el número de observaciones contenida en cada grupo de estudio de la serie puede ser 12 debido a que las observaciones son mensuales, por lo tanto:

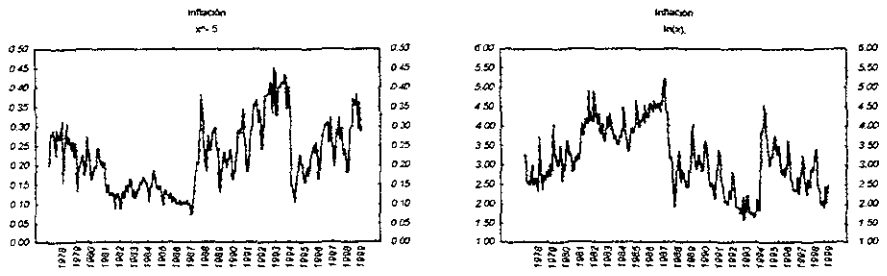
$$N = 264, n = 0, H = 12, R = 22$$

A continuación se muestran los coeficientes de variación obtenidos para la serie *inflación* para los diferentes valores del parámetro λ :

λ	$CV(\lambda)$
-1.5	0.93274
-1.0	0.66424
-0.5	0.47203
0.0	0.47701
0.5	0.65318
1.0	0.86481
1.5	1.06404

Estos resultados sugieren que $1 - \bar{X}_t$ es una buena transformación para la estabilización de la varianza ya que el coeficiente de variación se minimiza en $\lambda = -0.5$. O bien, también $\ln(X_t)$ puede ser utilizado ya que la diferencia entre los coeficientes de variación de estas dos transformaciones resulta insignificante en comparación con el resto.

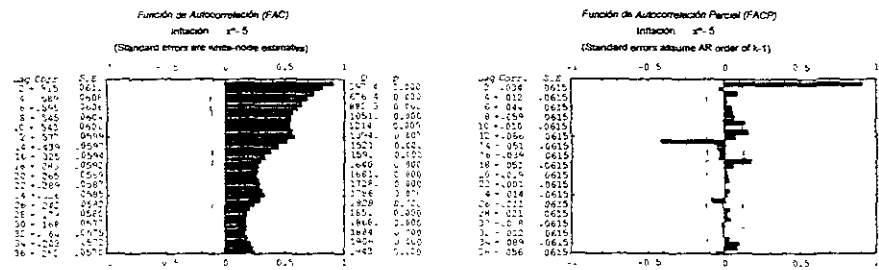
Por lo tanto, se mostrara a continuación la gráfica de ambas transformaciones, *gráfica 3.5*, y posteriormente en la *gráfica 3.6* se presentan las funciones de autocorrelación simple y parcial respectivas a cada una de ellas.



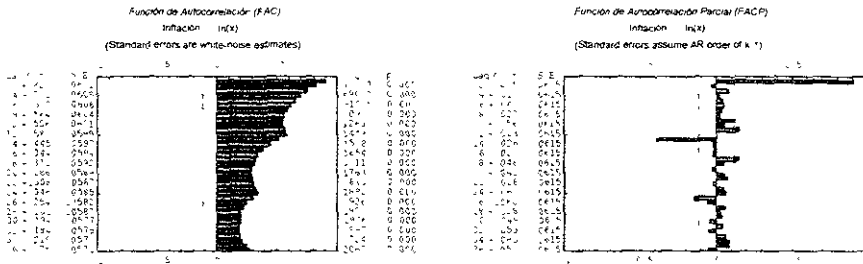
$$x_t = 1 - \bar{X}_t$$

$$x_t = \ln(X_t)$$

Gráfica 3.5 Transformaciones estabilizadoras de varianza



$$a) x_t = 1 - \bar{X}_t$$

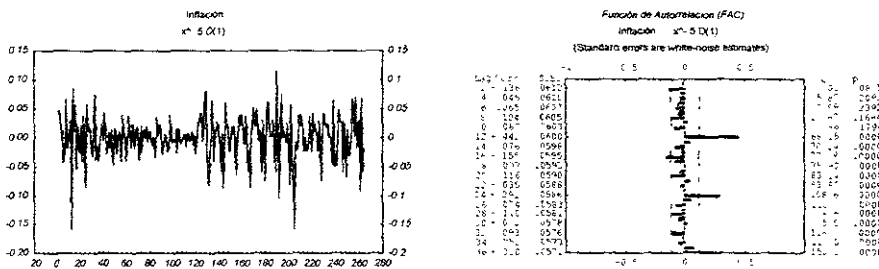


$$b) x_t = \ln(X_t)$$

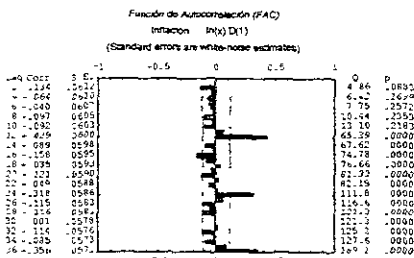
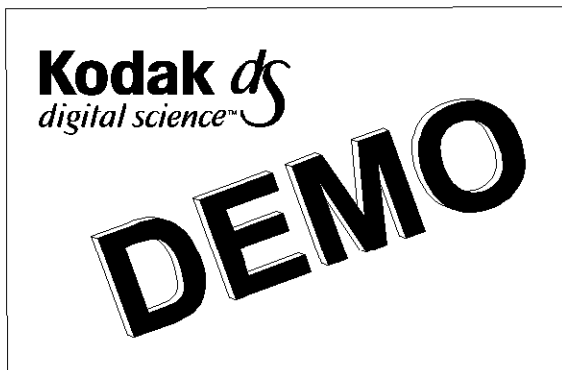
Gráfica 3.6 FAC y FACP de la serie transformada a) $x_t = 1 - X_t$ y b) $x_t = \ln(X_t)$

En la gráfica 3.5 se nota una estabilización de la varianza después de aplicada la transformación, sin embargo aún no se puede considerar una serie estacionaria ya que los correlogramas correspondientes a la FAC y a la FACP no corresponden a series de ese tipo, principalmente por la falta de patrones de decrecimiento y el coeficiente alto que muestra la FACP en el primer rezago indicando que se requiere de una estabilización.

En la siguiente gráfica se observa las series resultantes y su FAC respectiva para una primera diferenciación



$$a) \nabla X_t = (1 - B)1 - X_t$$

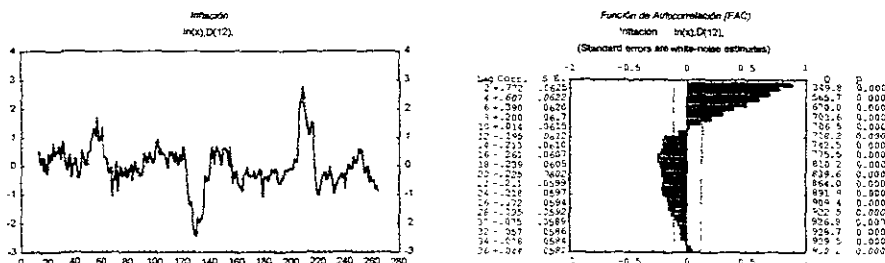


$$(1 - B)\ln(X_t)$$

Gráfica 3.7 Resultados de una primera diferenciación ∇x_t ,

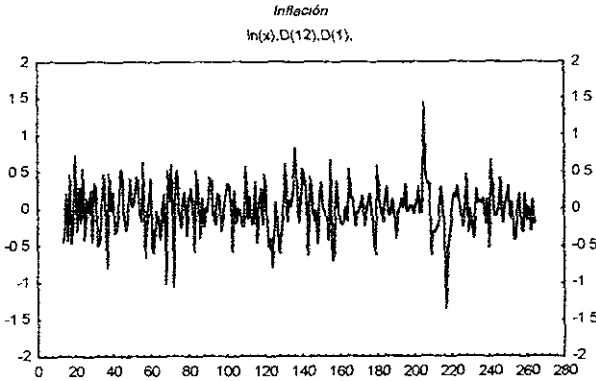
Una vez estabilizada la media, resulta más fácil diagnosticar el comportamiento de la varianza, obsérvese que a través de la segunda transformación se obtiene una serie más estable en la varianza a comparación de la primera, por lo que a partir de este momento se trabajará únicamente con $x_t = \ln(X_t)$.

Desde un principio obsérvese un pronunciado comportamiento estacional en la serie, además en la FAC se nota un decrecimiento lento de los coeficientes múltiples de doce, por lo que a continuación se aplicara el operador de diferencia estacional ∇_{12} a la serie transformada, ver gráfica 3.8.



Gráfica 3.8 Serie transformada $\nabla_{12} x_t = \nabla_{12} \ln(X_t)$

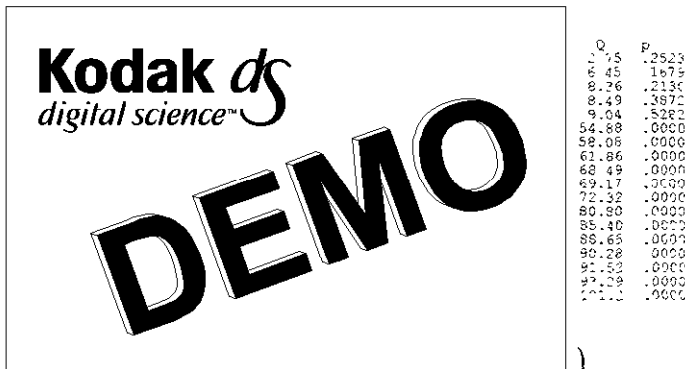
El aplicar sólo alguno de los dos operadores, ya sea el de diferenciación o el de desestacionalización por separado ha dado series no del todo estacionarias, por lo cual a continuación se muestra la combinación de estas dos últimas estabilizaciones.



Gráfica 3.9 Serie transformada $\nabla_{12} \nabla x_t = (1 - B^{12})(1 - B) \ln(X_t)$

Aparentemente se ha estabilizado la media y la varianza, sin embargo habrá que analizar la *FAC* y la *FACP* para ver que éstas se asemejen al comportamiento de series estacionarias. Para ello, observar las gráficas 3.10 y 3.11 en donde se presentan estas funciones.

Función de Autocorrelación (FAC)
Inflacion ln(x).D(1).D(12).



La gráfica muestra una *FAC* semejante a las gráficas teóricas de series estacionarias, por lo tanto se tratará de establecer el o los modelos tentativos que describan de la mejor manera a la serie y no violen los supuestos que han sido utilizados para su construcción.

Parte no estacional o regular del modelo

En la función de autocorrelación, si se ignoran los coeficientes altos debido a la estacionalidad que se presentan en el rezago 12 y sus múltiplos, se observa que la mayoría de los coeficientes se encuentran por debajo de las bandas de significación y el resto apenas si las toca por lo que se empezará a conjeturar algunos modelos basados en su comportamiento:

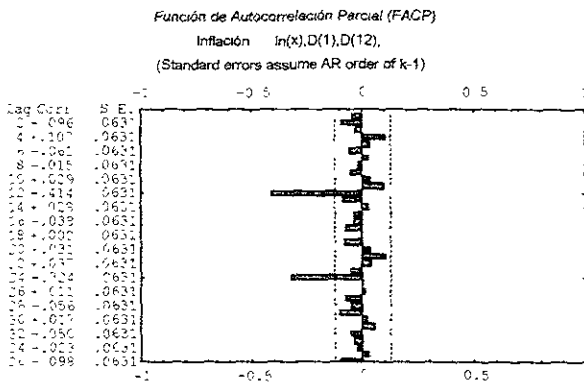
1. Si todos los coeficientes, de la parte regular, no son significativamente mayores a cero se puede suponer un modelo sin componentes autorregresivos y de media móvil representado por un $ARIMA(0,1,0)$.
2. Si se observan los coeficientes que están más cercanos a las bandas de significación se distingue un patrón de comportamiento alrededor de los rezagos estacionales. Es decir, los coeficientes 11, 13, 23, 25 y 35 se muestran más grandes a comparación del resto, lo que hace suponer un modelo con estructura de media móvil de orden uno. Cabe destacar que en los primeros rezagos la función no es la más adecuada para tratar de asociarla a un modelo de este estilo ya que el primer coeficiente se muestra disminuido en comparación al segundo y tomando en cuenta el comportamiento decreciente a cero mostrado alrededor de los coeficientes estacionales con cambio de signo hace suponer más bien un modelo autorregresivo de segundo orden.
3. Otro patrón sobresaliente entre los coeficientes es el que se refiere al cuarto rezago y sus múltiplos presentados alrededor de los coeficientes estacionales los cuales sugieren un modelo con estructura autorregresiva de orden cuatro por el decrecimiento mostrado entre los rezagos 4, 16 y 28.
4. La *FAC* de un modelo con estructura mixta se distingue por no presentar un patrón de comportamiento en los primeros q rezagos (donde q es el orden del polinomio de

media móvil) y posteriormente decrecer de acuerdo a la parte autorregresiva del modelo. En este caso, en los primeros tres coeficientes no se puede decir nada acerca del comportamiento de la función y posteriormente en los siguientes rezagos los coeficientes decrecen a cero. Por lo tanto, se puede suponer un modelo $ARIMA(2,1,3)$, tomando en cuenta las observaciones hechas en el segundo punto, o bien un modelo $ARIMA(4,1,3)$ considerando lo expuesto en el punto anterior.

Parte estacional del modelo

1. Hay un coeficiente significativamente distinto de cero que parece corresponder a un modelo estacional $ARIMA(0,1,1)_{12}$
2. También hay un marcado decrecimiento a cero que bien puede ser originado por un modelo autorregresivo de orden dos por el cambio de signo $ARIMA(2,1,0)_{12}$.

Ahora hay que analizar la función de autocorrelación para ratificar que los modelos supuestos cumplen también con los requerimientos de esta función.



Gráfica 3.11 FACP de $\nabla_{12}\nabla\ln(X_t)$

Al observar la gráfica de la FACP se puede decir que:

Parte no estacional o regular del modelo

1. Esta función es congruente con el modelo supuesto de estructuras autorregresiva y de media móvil nulas ya que aquí tampoco se observan coeficientes significativamente diferentes de cero conforme a las barras de significación.
2. En esta gráfica se suele observar en los primeros rezagos y hacia arriba de los coeficientes estacionales la función de autocorrelación parcial de la parte regular. Para este caso, hacia arriba del coeficiente 12 la función muestra decrecimiento a cero lo que podría indicar que efectivamente la estructura corresponde a un modelo de media móvil de primer orden, sin embargo alrededor del coeficiente 24 no ocurre lo mismo ya que el tercer rezago a partir de este rompe con el patrón de decrecimiento a cero indicando que quizá se debiese de considerar algún polinomio autorregresivo o de media móvil de orden tres.
3. El cuarto coeficiente de correlación parcial sobresale dentro de los primeros y después hay un decrecimiento confirmando un modelo autorregresivo de orden cuatro.
4. Como se menciona con anterioridad, el tercer coeficiente a partir del 24 rezago hacia arriba rompe con el decrecimiento esperado a cero lo que puede considerarse un indicador de un componente de media móvil de orden tres. Se puede pensar en un modelo autorregresivo en lugar de media móvil, no obstante en los primeros rezagos de la *FAC* no hay ningún coeficiente que confirme una estructura de este tipo.

Parte estacional del modelo

1. Ya que los coeficientes múltiplos de doce decrecen rápidamente a cero, el modelo $ARIMA(0,1,1)_{12}$ puede ser el adecuado.
2. Sin embargo, obsérvese que esta función presenta los dos primeros rezagos estacionales altamente significativos por lo que se puede pensar en un modelo con estructura autorregresiva de orden dos $ARIMA(2,1,0)_{12}$, o bien en una combinación entre ambos, de la forma $ARIMA(2,1,1)_{12}$.

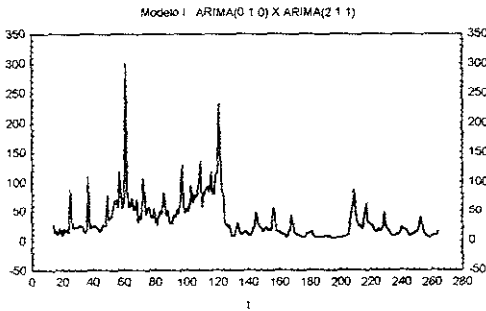
En resumen, se estimarán los siguientes modelos escogiendo el que resulte mejor de acuerdo a las pruebas de verificación mencionadas en el capítulo anterior:

- I $ARIMA(0,1,0) \times ARIMA(2,1,1)_{12}$: $\Phi_2(B^{12}) \nabla \nabla_{12} \ln(X_t) = \Theta_1(B^{12}) \alpha_t$
- II $ARIMA(0,1,1) \times ARIMA(2,1,1)_{12}$: $\Phi_2(B^{12}) \nabla \nabla_{12} \ln(X_t) = \theta_1(B) \Theta_1(B^{12}) \alpha_t$
- III $ARIMA(2,1,0) \times ARIMA(2,1,1)_{12}$: $\Phi_2(B^{12}) \phi_2(B) \nabla \nabla_{12} \ln(X_t) = \Theta_1(B^{12}) \alpha_t$
- IV $ARIMA(4,1,3) \times ARIMA(2,1,1)_{12}$: $\Phi_2(B^{12}) \phi_4(B) \nabla \nabla_{12} \ln(X_t) = \theta_1(B) \Theta_1(B^{12}) \alpha_t$

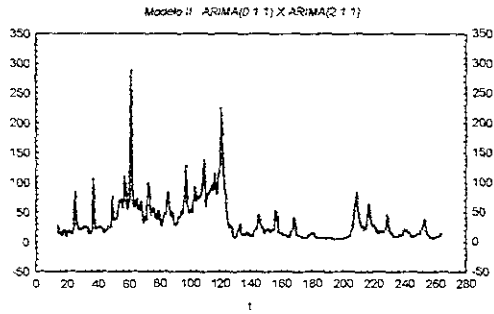
2.2 Estimación y verificación del modelo univariante para la serie inflación.

La estimación de parámetros se realizará por el método de máxima verosimilitud condicionado (explicado en el capítulo 2 sección 5). Observar los resultados que se muestran a continuación, a partir de la serie transformada $\nabla_{12} \nabla x_t = (1 - B^{12})(1 - B) \ln(X_t)$.

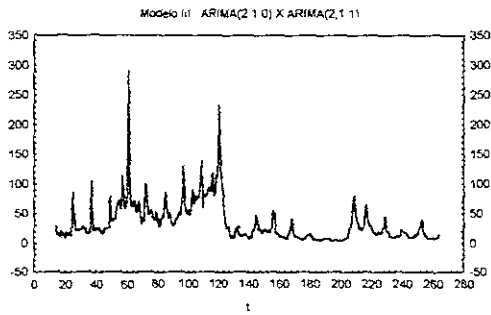
Ver gráficas 3.12.



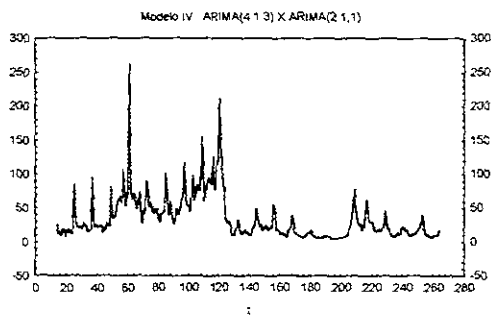
$$(1 + 06B^{12} + .13B^{24})x_t = (1 - .56B^{12})z_t$$



$$(1 + 0.5B^{12} + 1.3B^{24})x_t = (1 - 0.9B)(1 - 0.58B^{12})x_t$$



$$(1 - 0.3B^{12} + 1.1B^{24})(1 + 0.8B + 0.9B^2)x_t = (1 - 0.6B^{12})x_t$$



$$(1 + 0.1B^{12} + 0.8B^{24})(1 - 0.42B + 0.53B^2 - 0.82B^3 - 0.9B^4)x_t = (1 - 0.52B + 0.49B^2 - 0.92B^3)(1 - 0.6B^{12})x_t$$

Gráfica 3.12 Modelos estimados para la serie inflación

Los resultados se muestran muy parecidos, sin embargo habrá que verificar que se cumplen los supuestos en los que esta basada su construcción. A continuación, la siguiente tabla resume la media, la desviación estándar de los residuales así como las estadísticas *AIC* y *SBC* de los modelos estimados las cuales ayudarán a elegir el modelo más adecuado a través de su análisis:

Modelo	Residuales	\hat{a}_t	\hat{S}_a	AIC	SBC
ARIMA(0,1,0) × ARIMA(2,1,1) ₁₂	251	-0.00820	0.29887	97.103764	107.68012
ARIMA(0,1,1) × ARIMA(2,1,1) ₁₂	251	-0.00905	0.29784	97.789464	111.89128
ARIMA(2,1,0) × ARIMA(2,1,1) ₁₂	251	-0.00977	0.29673	98.138115	115.76538
ARIMA(4,1,3) × ARIMA(2,1,1) ₁₂	251	-0.02233	0.28831	94.04713	129.30166

Tabla 3.1 Resumen de residuales

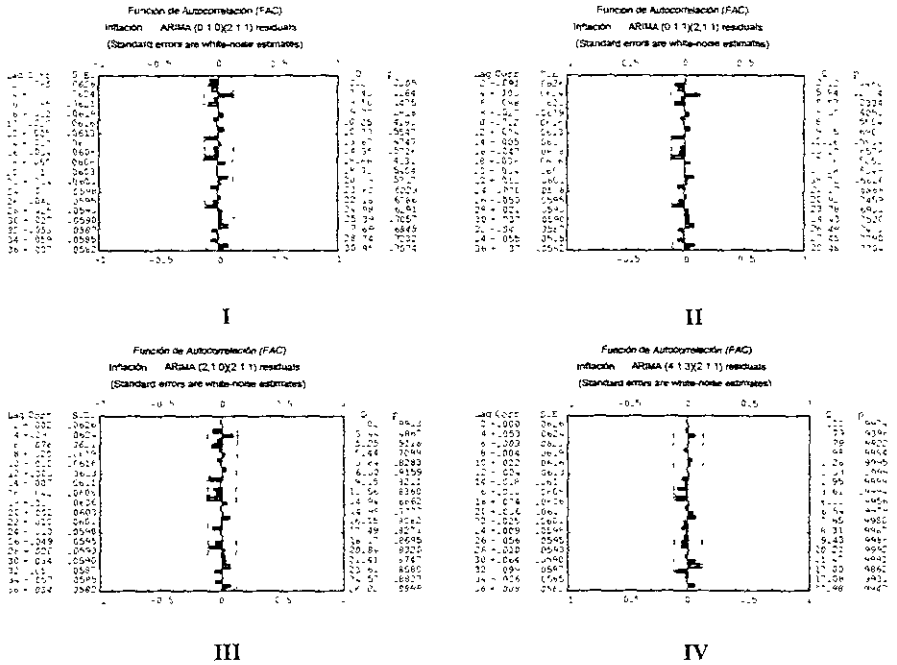
El primer supuesto a verificar es que los residuales tienen media cero. Para ello habrá que validar la siguiente estadística:

$$\frac{\hat{a}}{\hat{S}_a / \sqrt{251}} < 2$$

Modelo	Resultados de la estadística
I	0.434544
II	0.481394
III	0.521857
IV	1.226950

De acuerdo a los resultados ninguno de los cuatro modelos viola el supuesto de media igual a cero. Como ya se había mencionado con anterioridad la prueba es válida si los residuales

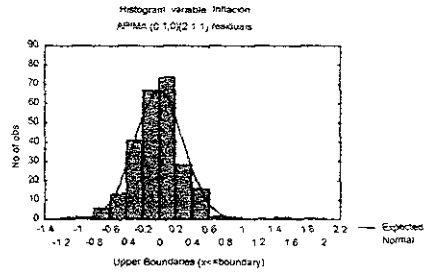
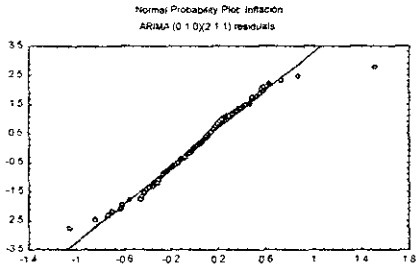
no son correlacionados para asegurar que \hat{S}_a es un buen estimador de la varianza. Para ello obsérvese las siguientes funciones de autocorrelación de los residuales de cada modelo.



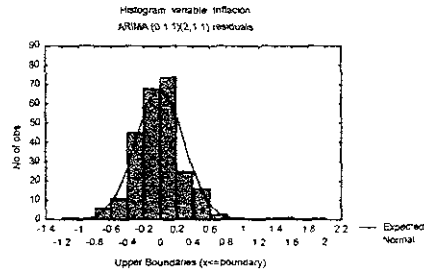
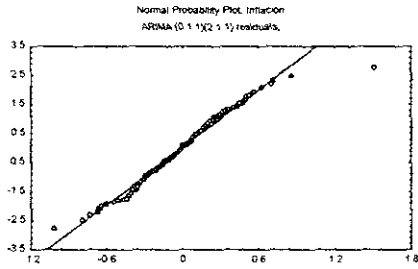
Gráfica 3.14 Funciones de autocorrelación de modelos estimados

Ninguna gráfica de autocorrelación está violando el supuesto de residuales no autocorrelacionados, sin embargo, nótese en la gráfica correspondiente al cuarto modelo las ventajas de utilizar un modelo con estructura mixta, los modelos con una solo tipo de polinomio ya sea autorregresivo o de media móvil no logran reducir los coeficientes de la manera que lo hace el cuarto modelo.

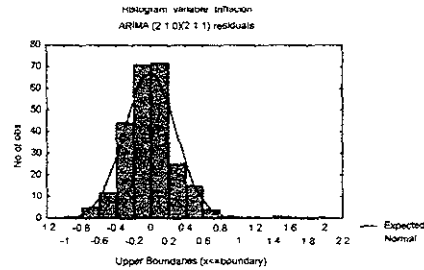
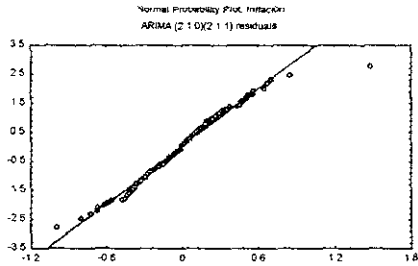
Se ha supuesto también que los residuales se distribuyen normalmente, lo cual puede comprobarse en las siguientes gráficas, en donde para estos modelos los residuales se encuentran muy cercanos a la recta normal.



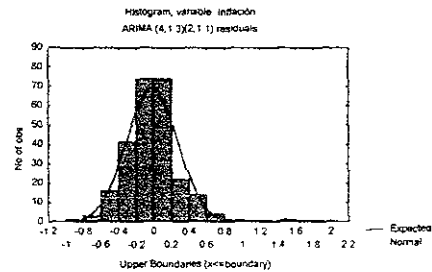
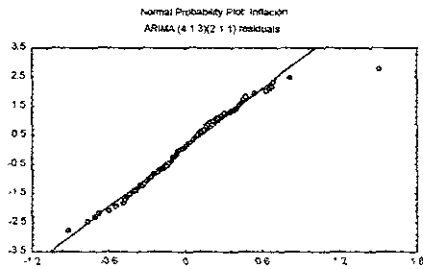
I



II



III



IV

Gráfica 3.15 Normalidad de residuales de modelos estimados

De igual manera que en las pruebas anteriores ningún modelo muestra grandes ventajas sobre los otros.

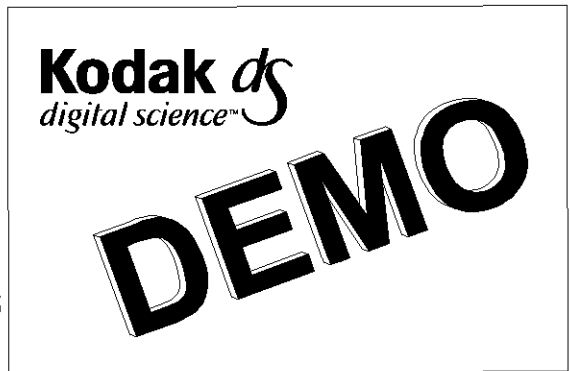
Basándose en las estadísticas *AIC* y *SBC* el mejor modelo es aquel sin componentes autorregresivos ni de media móvil $ARIMA(0,1,0) \times ARIMA(2,1,1)_2$ o bien el cuarto modelo $ARIMA(4,1,3) \times ARIMA(2,1,1)_1$. Habrá que realizar ahora las pruebas de verificación en cuanto a parámetros ya que pueden ser también una guía para la elección del modelo.

En la siguiente tabla se muestran los intervalos de confianza y los errores estándar para cada uno de los parámetros de los tres modelos estimados. cabe mencionar que ningún modelo presenta errores en cuanto a raíces mayores que uno para los polinomios autorregresivos y de media móvil. por lo cual todos ellos se pueden considerar estacionarios e invertibles.

Modelo	Parámetros	Estimador	Error estándar	Intervalos de confianza 95%	
				Inferior	Superior
I	Φ_1	-0.059576	0.1334144	-0.322346	0.203192
	Φ_2	-0.130833	0.0957043	-0.319330	0.057663
	Θ_1	0.563788	0.1251098	0.317375	0.810201
II	θ_1	0.092548	0.071816	-0.048901	0.233998
	Φ_1	-0.049506	0.129156	-0.303895	0.204881
	Φ_2	-0.126641	0.094327	-0.312431	0.059147
III	Θ_1	0.580366	0.120238	0.343542	0.817189
	ϕ_1	-0.082824	0.064331	-0.209534	0.043886
	ϕ_2	-0.094187	0.064675	-0.221575	0.033199
	Φ_1	-0.025115	0.122979	-0.267343	0.217111
	Φ_2	-0.113706	0.091834	-0.294589	0.067175
	Θ_1	0.604771	0.111697	0.384767	0.824776

Modelo	Parámetros	Estimador	Error estándar	Intervalos de confianza	
				95%	
				Inferior	Superior
IV	ϕ_1	0.424866	0.078315	0.270598	0.579135
	ϕ_2	-0.534352	0.069975	-0.672193	-0.396511
	ϕ_3	0.822815	0.069230	0.686442	0.959189
	ϕ_4	0.085671	0.070379	-0.052967	0.224308
	θ_1	0.515541	0.042002	0.430000	0.600170
	θ_2	-0.493603	0.042002	-0.570000	-0.410000
	θ_3	0.920519	0.042002	0.830000	0.990000
	Φ_1	-0.011706	0.042002	-0.080000	0.050000
	Φ_2	-0.082584	0.042002	-0.150000	-0.010000
	Θ_1	0.601306	0.042002	0.510000	0.690000

Tabla 3.2 Resultados obtenidos de la estimación



Se ha resaltado con negritas los parámetros cuyo intervalo de confianza no considera el cero, ya que en caso contrario se podrían ignorar esos parámetros en el modelo siempre y cuando no muestren ser indispensables en la interpretación de los datos.

Del análisis de residuales y de parámetros se concluye que cualquiera de los cuatro modelos puede ser elegido ya que ninguno viola algún supuesto además de que arrojan resultados muy parecidos entre sí, no obstante se debe de escoger el mejor conforme a las pruebas de verificación de modelos.

El cuarto modelo es el que reduce la estadística *AIC* y el primer modelo el que reduce a *SBC*, sin embargo este último no puede ser considerado el mejor ya que presenta un coeficiente de correlación pronunciado en el cuarto rezago de la *FAC* de sus residuales. El segundo modelo, además de presentar estadísticas pequeñas obsérvese que los residuales de autocorrelación entre residuales en general se ven reducidos, de la misma manera que para

el tercer modelo. no obstante las estadísticas *AIC* y *SBC* de este ultimo se muestran una desventaja en comparación del primer y segundo modelo. Como puede observarse el modelo de media móvil reduce las estadísticas y presenta al mismo tiempo una gráfica de los residuos considerablemente más pequeños. En cuanto a las gráficas de normalidad, como ya se había mencionado las cuatro se muestran muy parecidas entre sí, sin embargo en los histogramas se observan algunas ventajas de acuerdo al tipo de modelo, como es el caso del modelo que tiene una estructura mixta que muestra mayor simetría conforme a la curva de normalidad.

En la tabla 3.2, la mayoría de los parámetros estimados del segundo y tercer modelo elegidos consideran al cero dentro de su intervalo de confianza por lo que se puede pensar en eliminarlos, reduciéndose ambos a la estructura del primer modelo. Por lo que los modelos a elegir solo están entre los modelos uno y cuatro. De estos dos modelos aún los parámetros autorregresivos estacionarios pueden eliminarse, sin embargo antes de estimar otros modelos con un número menor de parámetros hay que observar la matriz de correlación para ver si hay parámetros altamente correlacionados.

	Φ_1	Φ_2	Θ_1
Φ_1	1 000000	0 564316	0 878816
Φ_2	0 664316	1 000000	0 738464
Θ_1	0 878816	0 738464	1 000000

Modelo I

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	θ_1	θ_2	θ_3	Φ_1	Φ_2	Θ_1
ϕ_1	1 000000	-0 527769	0 664638	-0 679400	0 561967	-0 157896	0 015274	-0 013046	-0 010121	-0 087080
ϕ_2	-0 527769	1 000000	-0 510499	0 769929	-0 267394	0 689468	-0 206322	-0 063722	0 006334	0 014023
ϕ_3	0 664638	-0 510499	1 000000	-0 569569	0 331170	-0 533883	0 656833	-0 001652	-0 020117	-0 080627
ϕ_4	-0 679400	0 769929	-0 569569	1 000000	-0 037474	0 339976	-0 155306	0 000531	0 062698	0 084649
θ_1	0 561967	-0 267394	0 331170	-0 037474	1 000000	-0 280351	-0 019598	0 006524	-0 032429	-0 082253
θ_2	-0 157896	0 689468	-0 533883	0 339976	-0 280351	1 000000	-0 617174	-0 038395	-0 023738	-0 031745
θ_3	0 015274	-0 206322	0 656833	-0 155306	-0 019598	-0 617174	1 000000	-0 046099	0 047600	-0 040659
Φ_1	-0 013046	-0 063722	-0 001652	0 000531	0 006524	-0 038395	-0 046099	1 000000	0 577305	0 839307
Φ_2	-0 010121	0 006334	-0 020117	0 062698	-0 032429	-0 023738	0 047600	0 577305	1 000000	0 697433
Θ_1	-0 087080	0 014023	-0 080627	0 084649	-0 082253	-0 031745	-0 040659	0 839307	0 697433	1 000000

Modelo IV

Gráfica 3.16 Matriz de correlación de modelos estimados

Hay una marcada correlación entre los parámetros estacionales autorregresivos Φ_1 , Φ_2 y el parámetro estacional de media móvil Θ_1 de modo que lo más conveniente es que se reduzca el orden de alguno de los dos polinomios estacionales. En los dos modelos los parámetros Φ_1 y Φ_2 además de estar correlacionados consideran al cero dentro de su intervalo de confianza, por lo que la parte estacional puede ser descrita por un modelo únicamente de media móvil de orden uno.

De lo observado se pueden probar aún otros modelos con las modificaciones sugeridas por las pruebas de verificación hasta ahora realizadas. La primera modificación a realizar consistirá en eliminar el polinomio autorregresivo estacional, lo cual sólo se realizará en el primer y cuarto modelo ya que son los que mostraron un comportamiento más cercano a la serie real. Obsérvese además, que el parámetro ϕ_4 del polinomio autorregresivo del cuarto modelo considera al cero dentro de su intervalo de confianza y con una correlación algo significativa, mostrada en la matriz de correlación entre parámetros, por lo que también puede ser eliminado. De acuerdo a lo anterior expuesto, los modelos a probar son los siguientes:

Modelo	Residual	\hat{a}_t	$\hat{\sigma}_a$	AIC	SBC
V $ARIMA(0,1,0) \times ARIMA(0,1,1)_{12}$	251	-0.00817	0.300154	95.853588	99.379041
VI $ARIMA(3,1,3) \times ARIMA(0,1,1)_{12}$	251	-0.00979	0.291823	99.077396	123.75557
VII $ARIMA(4,1,3) \times ARIMA(0,1,1)_{12}$	251	-0.02255	0.288874	95.90252	124.10614

Tabla 3.3 Resumen de residuales

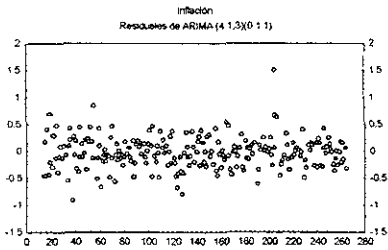
No se muestra el resto de las estadísticas y gráficas de estos tres últimos modelos debido a que la diferencia entre estas y los modelos anteriores (I y IV) resultan imperceptibles. En lo que respecta a las estadísticas *AIC* y *SBC* estas se ven favorecidas en el modelo $ARIMA(0,1,0) \times ARIMA(0,1,1)_{12}$ con respecto al $ARIMA(0,1,0) \times ARIMA(2,1,1)_{12}$. Nótese que

el sexto modelo no presenta ninguna ventaja exceptuando por la disminución de parámetros. Por último el séptimo modelo presenta unas estadísticas aceptables pero en el caso de la estadística AIC esta aumenta con respecto al modelo original $ARIMA(4,1,3) \times ARIMA(2,1,1)_{12}$ no ocurriendo lo mismo con la estadística SBC la cual se ve favorecida con la reducción de parámetros estacionales. Conjuntando toda esta información se concluye que el modelo más adecuado de acuerdo a las estadísticas AIC y SBC puede ser el modelo I, el modelo IV o el modelo VII, conforme al principio de parsimonia los modelos factibles se ven reducidos al I y al modelo VII, por último las funciones de autocorrelación simple y parcial sugieren al modelo VII como el más adecuado ya que como se puede observar en la gráfica 3.14 la inclusión de componentes autorregresivos y de media móvil disminuyen los coeficientes de estas funciones. Por lo tanto, se decide que el mejor modelo es el $ARIMA(4,1,3) \times ARIMA(0,1,1)_{12}$ dado que contiene componentes autorregresivos y de media móvil que cumplen con las pruebas de verificación de parámetros y ayudan a mejorar las pruebas de residuales.

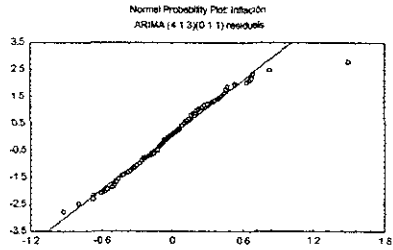
Tomando en cuenta que el modelo será utilizado para describir a la serie inflación considérese que el valor de ésta serie en un determinado mes depende en gran medida del comportamiento de los meses inmediato anteriores del mismo año así como la de años precedentes, si se utiliza el modelo $ARIMA(0,1,0) \times ARIMA(0,1,1)_{12}$ se estaría basando la descripción de la serie únicamente en la información histórica del mismo mes, en cambio al utilizar el modelo $ARIMA(4,1,3) \times ARIMA(0,1,1)_{12}$ además de estar tomando la información anual a través del componente estacional de media móvil incluye además el de los meses contiguos aprovechando de esta manera el comportamiento más cercano de la inflación la que de alguna manera puede reflejar el escenario que en ese momento atraviesa el país. Las gráficas y tablas para verificar que este modelo cumple con las demás pruebas de verificación se presentan a continuación:

Modelo	Parámetros	Estimador	Error estándar	Intervalos de confianza 95%	
				Inferior	Superior
VII	ϕ_1	0.424206	0.076568	0.273385	0.575027
	ϕ_2	-0.531233	0.066018	-0.661273	-0.401193
	ϕ_3	0.821192	0.065452	0.692266	0.950118
	ϕ_4	0.090798	0.069135	-0.045382	0.226979
	θ_1	0.513305	0.041375	0.431805	0.594805
	θ_2	-0.494282	0.048400	-0.589620	-0.398944
	θ_3	0.925402	0.045363	0.836046	1.014757
	Θ_1	0.640853	0.051539	0.539332	0.742374

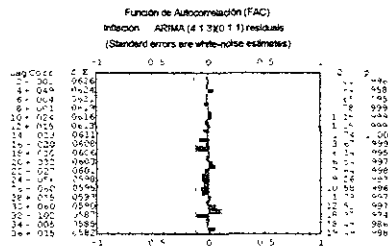
Tabla 3.4 Resultados obtenidos de la estimación para el análisis de parámetros.



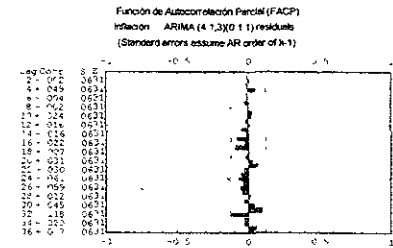
a)



b)



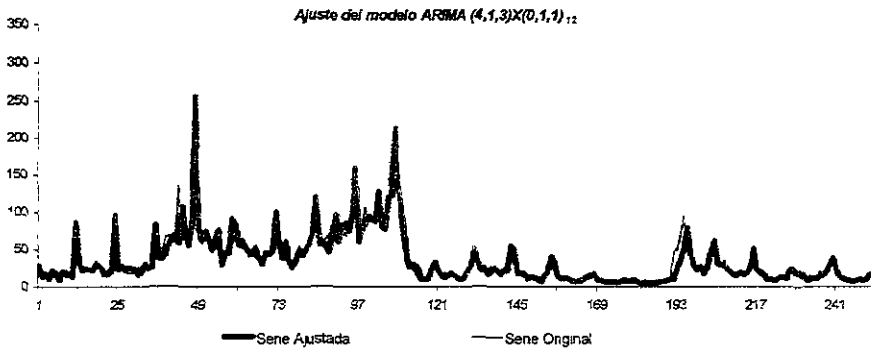
c)



d)

Gráfica 3.18 a) Residuales del modelo b) Normalidad de residuales c) FAC y d) FACP

A continuación se muestra la gráfica del modelo ajustado sobre la serie de inflación real, ver *gráfica 3.19*, obsérvese que a partir de 1987 el comportamiento es más estable por lo que la serie generada a partir del modelo tiene un mayor ajuste a la serie original.





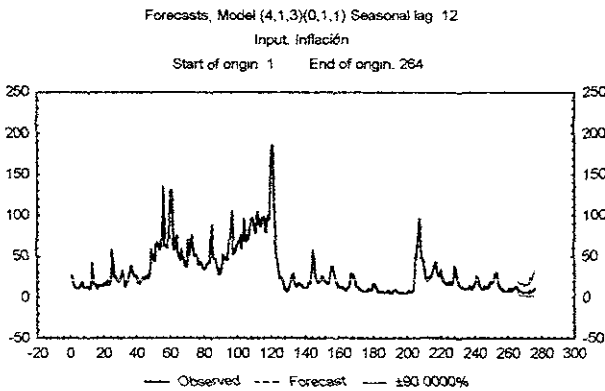
Gráfica 3.19 Serie original de inflación y modelo ajustado

2.3 Uso del modelo inflación

El objetivo de la creación de un modelo univariante para la inflación es estudiar los efectos que su comportamiento provoca sobre de otras variables económicas en las que se tenga interés.

Los valores futuros de la serie obtenidos a través del modelo ajustado se utilizarán posteriormente en el pronóstico de una segunda variable. En la *tabla 3.5* se muestran las tasas de inflación pronosticadas para el año 2000.

Fecha		 		
Enero	- 2000			
Febrero	- 2000			
Marzo	- 2000			
Abril	- 2000			
Mayo	- 2000			
Junio	- 2000	6.425223	2 364393	17.460498
Julio	- 2000	6 159509	2 156814	17.590555
Agosto	- 2000	5.558332	1.862652	16.586601
Septiembre	- 2000	8.488901	2 715901	26.533157
Octubre	- 2000	6.766834	2.084841	21.963331
Noviembre	- 2000	8.945703	2.677676	29.886215
Diciembre	- 2000	11.496124	3.337491	39 598865

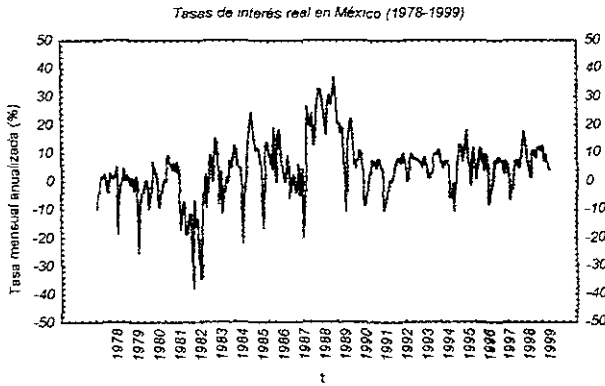


Gráfica 3.20 Predicción de la serie inflación en México

3. Modelos de regresión dinámica

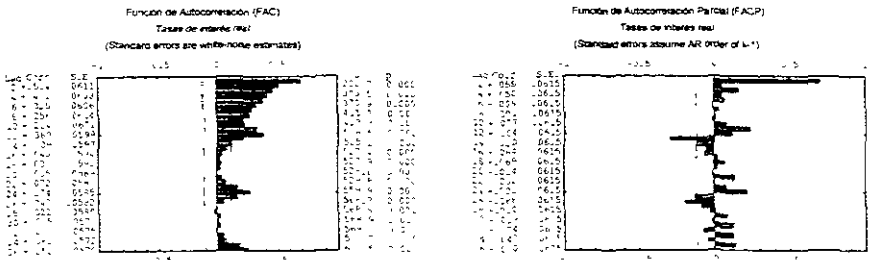
3.1 Análisis previo de la serie tasas de interés real.

Antes de iniciar el preblanqueado de series para intentar ajustar un modelo que refleje el impacto de la inflación sobre la serie tasas de interés real habrá que ver si esta es estacionaria o se necesitará aplicar algunas transformaciones para conseguirlo. *Gráfica 3.21.*



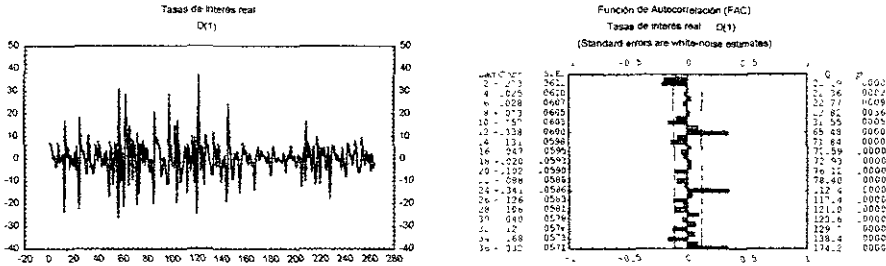
Gráfica 3.21 Tasas de interés real en México (1978-1999)

La gráfica de tasas de interés muestra claros indicios de no estacionariedad lo que se comprueba en la *FAC* y la *FACP* de la serie mostradas a continuación:



Gráfica 3.22 *FAC* y *FACP* de la serie tasas de interés real

En la *gráfica 3.22* además de confirmar la no estacionariedad de la serie se observa un marcado comportamiento estacional a 12 meses. En seguida, se presenta la primera diferenciación de la serie así como su *FAC* con el propósito de estabilizar media y de ser posible varianza.



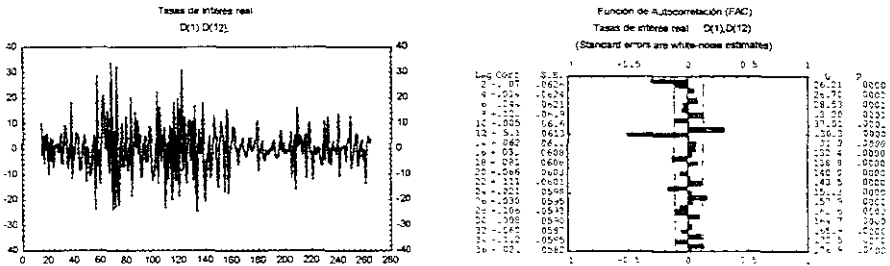
Gráfica 3.22 $\nabla Y_t = Y_t - Y_{t-1}$ y su *FAC*

La serie diferenciada presenta características de estacionariedad en la parte regular, sin embargo se intentará aplicar una función estabilizadora para confirmar que la serie no requiere de una transformación potencia.

De igual manera que se hizo para la serie inflación se escogerá aquella transformación que minimice el coeficiente de variación. A continuación se presentan los valores obtenidos para diferentes valores de λ :

λ	$CV(\lambda)$
-1.5	1.185917
-1.0	0.903017
-0.5	0.681084
0.0	0.526277
0.5	0.442267
1.0	0.424586
1.5	0.456613

Los coeficientes de variación vienen a confirmar que la serie no requiere una transformación para estabilizar la varianza. Con la primera diferenciación de la serie se logra la estabilización de la media, sin embargo si se fija la atención en los coeficientes múltiplos de doce, correspondientes a la parte estacional, se nota un lento decrecimiento a cero en la *FAC* de la *gráfica 3.22*. Por lo tanto, se sugiere aplicar el operador ∇_{12} a la serie ∇Y_t , obsérvese los resultados de la *gráfica 3.23*.



Gráfica 3.23 Serie $\nabla_{12}\nabla Y_t$ y su *FAC*

Con la aplicación del operador ∇_{12} se logró un prominente coeficiente en el primer rezago estacional y posteriormente un rápido decrecimiento a cero, en lo que respecta a la parte regular del proceso no presenta problemas de no estacionariedad. Es así como, se considerará a la serie $\nabla_{12}\nabla Y_t$ como una serie estacionaria y por lo tanto se logrará establecer una relación entre la serie de inflación y la de tasas de interés ya que ambas deben de ser estacionarias con el mismo número de diferenciaciones tanto estacionales como no estacionales.

3.2 Preblanqueado de la serie dependiente tasas de interés real

Una vez que se ha homologado el número de diferenciaciones que se aplicará a las series, se supondrá que la serie dependiente $y_t = \nabla_{12}\nabla Y_t$ sigue el mismo modelo que la serie independiente tasas de inflación:

$$\phi_4(B)x_t = \theta_3(B)\Theta_1(B^{12})\alpha_t$$

$$\phi_4(B)x_t = \theta_3(B)\Theta_1(B^{12})\alpha_t \tag{3.2.1}$$

es decir:

$$(1 - .42B + .53B^2 - .82B^3 - .09B^4)x_t = (1 - .51B + .49B^2 - .93B^3)(1 - .64B^{12})\alpha_t$$

en donde, α_t representa a la serie de residuales obtenida de ajustar el modelo univariante en la sección anterior para la serie inflación y x_t representa a la serie estacionaria $\nabla_{12}\nabla\ln(X_t)$ con desviación estándar:

$$\hat{\sigma}_\alpha = 0.28887$$

Al aplicar el mismo modelo univariante a la serie estacionaria $y_t = \nabla\nabla_{12}Y_t$, se obtiene una segunda serie de residuales, conocida como serie filtrada y denotada por β_t . El modelo univariante es el siguiente:

$$\phi_4(B)\nabla\nabla_{12}Y_t = \theta_3(B)\Theta_1(B^{12})\beta_t \tag{3.2.2}$$

respetando los valores obtenidos para los parámetros de los coeficientes, de la siguiente manera:

$$(1 - .42B + .53B^2 - .82B^3 - .09B^4)y_t = (1 - .51B + .49B^2 - .93B^3)(1 - .64B^{12})\beta_t$$

y la desviación estándar para la serie β_t es:

$$\hat{\sigma}_\beta = 7.85823$$

en la *tabla 3.5* se muestran los resultados de la función de correlación cruzada entre las series α_t y β_t , resultado del preblanqueado de series y la función de respuesta a impulsos.

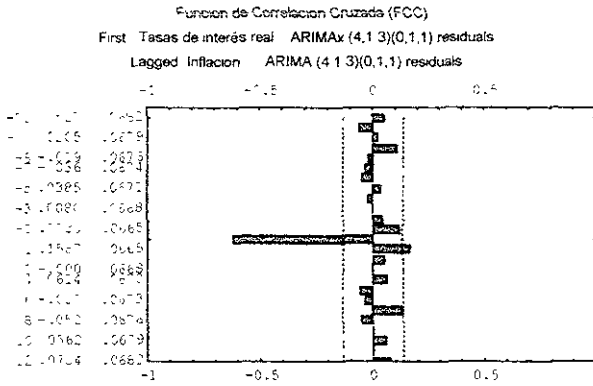
\hat{v}_k obtenida de la relación:

$$\hat{v}_k = r_{\alpha\beta} \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha}$$

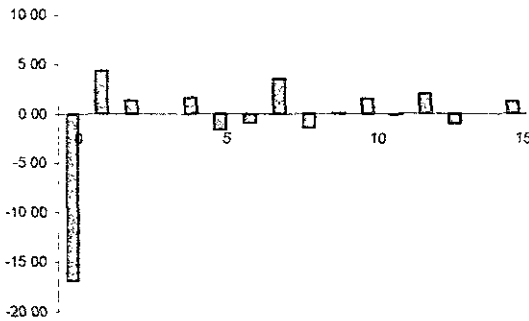
Rezago(k)	$r_{\alpha\beta}$	\hat{v}_k
0	-0.620356	-16.875549
1	0.158738	4.318148
2	0.051212	1.393125
3	-0.000114	-0.003093
4	0.061353	1.668993
5	-0.059432	-1.616716
6	-0.032843	-0.893437
7	0.131507	3.577383
8	-0.052029	-1.415349
9	0.004368	0.118819
10	0.056168	1.527938
11	-0.002931	-0.079744
12	0.075439	2.052165

Tabla 3.5 Resultados de la función de respuesta a impulsos

se ha señalado con negritas los coeficientes significativamente distintos de cero, los cuales se presentan en el rezago 1, 2 y posiblemente en el rezago 7 y 12. Gráficamente es más fácil la apreciación de la función de correlación cruzada, ver *gráfica 3.24* y *3.25*.



Gráfica 3.24 Función de correlación cruzada para las series preblanqueadas



Gráfica 3.25 Función de respuesta a impulsos (f_{ri})

Era de esperarse que la función de respuesta a impulsos se presentará de esta manera ya que el comportamiento de una tasa de interés depende en gran medida de la inflación prevista para el mes.

Debido a que la función de respuesta a impulsos de una función de transferencia es la análoga a una función de autocorrelación de un modelo *ARMA*, se puede decir que un polinomio de media móvil de primer orden puede utilizarse para la interpretación de los datos ya que el primer coeficiente destaca considerablemente del resto. Como ayuda para

determinar el orden de dichos polinomios se utilizarán posteriormente las estadísticas *AIC* y *SBC*, pero por el momento se empezará por estimar un modelo con función de transferencia:

$$v(B) = w_1(B) \tag{3.2.3}$$

A partir de las relaciones que se obtuvieron entre parámetros en el capítulo anterior se pueden obtener valores iniciales para los parámetros de la función de transferencia, igualando los coeficientes para obtener los valores de v_j :

$$\left(\hat{v}_0 + \hat{v}_1 B + \hat{v}_2 B^2 + \dots \right) = \hat{w}_0 + \hat{w}_1 B \tag{3.2.4}$$

debido a que no hay componentes equivalentes a media móvil en la función de transferencia los valores de v_j se obtiene directamente de la siguiente manera:

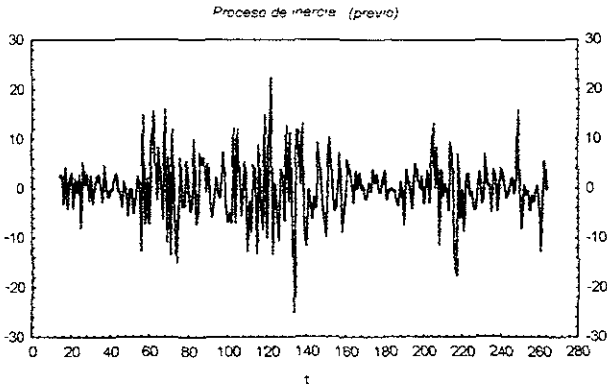
$$\hat{w}_0 = \hat{v}_0 = -16.875549$$

$$\hat{w}_1 = \hat{v}_1 = 4.318148$$

De estos estimadores preliminares se obtendrá el siguiente proceso de inercia preliminar:

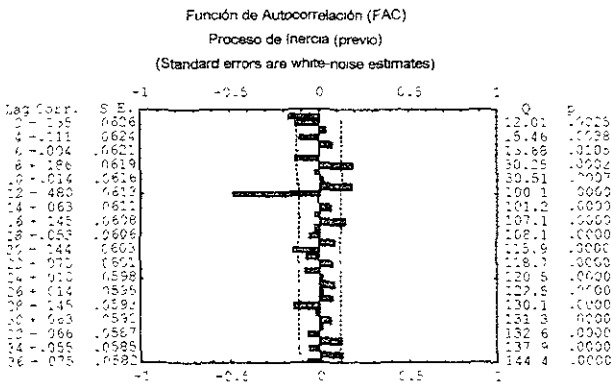
$$\hat{n}_t = y_t - \hat{w}_0 x_t - \hat{w}_1 x_{t-1} \tag{3.2.5}$$

obteniéndose la gráfica:

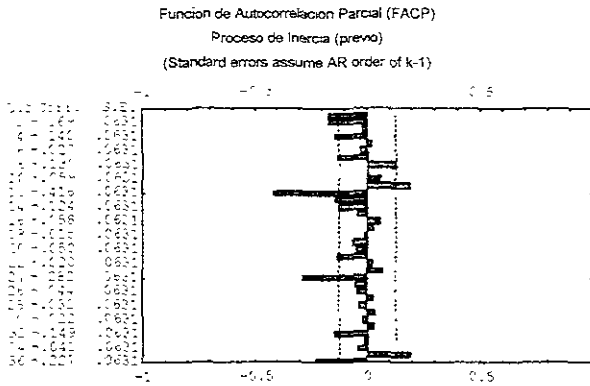


Gráfica 3.26 Proceso de inercia previo n_t

Para conjeturar un modelo univariante para el proceso de inercia hay que analizar las funciones de autocorrelación y de autocorrelación parcial, ver gráfica 3.28 y 3.29 respectivamente.



Gráfica 3.27 Función de autocorrelación de n_t



Gráfica 3.28 Función de autocorrelación parcial de \hat{n}_t

Como se puede observar aún quedan coeficientes significativos tanto de la parte no estacional como estacional. Un modelo autorregresivo de orden dos ó cuatro puede ser el indicado para interpretar el comportamiento del proceso inercia ya que en la *FACP* los dos primeros coeficientes son significativamente diferentes de cero y alrededor de los coeficientes estacionales destaca significativamente el cuarto rezago. Quizá un modelo más adecuado pudiese ser uno con estructura mixta, por lo que de acuerdo a las gráficas se sugiere aunar un polinomio de media móvil de primer orden.

En cuanto a la parte estacional, el comportamiento de las funciones indican que se trata de un modelo de media móvil, ya que el primer coeficiente es altamente significativo y luego en la *FACP* se observa solo decrecimiento. Por lo tanto, los modelos tentativos para el proceso de inercia son:

$$\phi_2(B)\hat{n}_t = \theta_1(B)\Theta_1(B^{12}) \quad (3.2.6)$$

y

$$\phi_4(B)\hat{n}_t = \theta_1(B)\Theta_1(B^{12}) \quad (3.2.7)$$

$$\phi_1(B)y_t = \theta_1(B)\Theta_1(B^{12}) \quad (3.2.7)$$

que al combinarlo con la función de transferencia se tiene el modelo completo de regresión dinámica:

$$y_t = w_1(B)x_t + \frac{\theta_1(B)\Theta_1(B^{12})}{\phi_2(B)} a_t \quad (3.2.8)$$

ó

$$y_t = w_1(B)x_t + \frac{\theta_1(B)\Theta_1(B^{12})}{\phi_1(B)} a_t \quad (3.2.9)$$

en donde x_t y y_t son las series transformadas estacionarias.

3.3 Estimación del modelo y verificación de los supuestos.

Una vez que se ha elegido el modelo tentativo, lo que sigue es la estimación de los parámetros utilizando el método de máxima verosimilitud condicionado. Como se vio en el capítulo 2 sección 5.5 el modelo (3.2.8) puede ser reescrito de la siguiente manera para poder ser estimado:

$$\phi_2(B)y_t = \phi_2(B)w_1(B)x_t + \theta_1(B)\Theta_1(B^{12})a_t \quad (3.2.10)$$

dando como resultado un modelo del siguiente estilo:

$$c_2(B)y_t = d_3(B)x_t + e_1(B)\Theta_1(B^{12})a_t \quad (3.2.11)$$

en donde $c_2(B) = \phi_2(B)$, $d_3(B) = \phi_2(B)w_1(B)$ y $e_1(B) = \theta_1(B)$ son los polinomios que componen el modelo adecuado para utilizar el método de máxima verosimilitud

condicionado explicado en el capítulo anterior. De igual manera se construye un modelo para la segunda opción de modelo de regresión dinámica (3.2.9).

En la *tabla 3.6* se muestra el resumen de las estadísticas de los residuales obtenidos para los dos modelos opcionales:

<i>Modelo</i>	<i>Residuales</i>	\hat{a}_1	$\hat{\sigma}_a$	<i>AIC</i>	<i>SBC</i>
I	248	-0.136594	4.756985	1498.9588	1527.0662
II	246	-0.178406	4.7398873	1493.1484	1535.2124

Tabla 3.6 Resumen de residuales modelos de regresión dinámica

Ninguno de los dos modelos muestra ventaja sobre el otro, sin embargo en cuanto a ahorro de parámetros es mejor el primer modelo, por lo que por el momento se considerara al primer modelo como mejor candidato que el segundo. No se olvide que en la *tabla 3.5* hubo un coeficiente significativo de la función de correlación cruzada en el doceavo rezago que podría estar indicando la posible presencia de un componente autorregresivo estacional en la función de transferencia al igual que también destaco el coeficiente en el séptimo rezago. A continuación se estimará el siguiente modelo:

$$c_2(B)y_t = d_0x_t + d_1x_{t-1} + d_2x_{t-2} + d_3x_{t-3} + d_{12}x_{t-12} + e_t(B)\Theta_1(B^{12})a_t \quad (3.2.12)$$

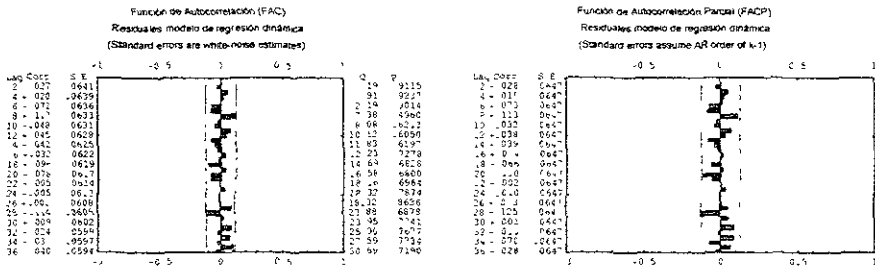
Obteniéndose los siguientes resultados en los residuales:

<i>Modelo</i>	<i>Residuales</i>	\hat{a}_1	$\hat{\sigma}_a$	<i>AIC</i>	<i>SBC</i>
III	239	-0.055518	4.832978	1453.5190	1484.8072

En este último modelo las estadísticas *AIC* y *SBC* muestran una disminución con respecto a los dos primeros modelos, dado que el primer modelo esta contenido en cuanto numero de

de que haya una sobreparametrización se estará regresando al primer modelo. El modelo incluyendo un séptimo componente representativo no mostró mayores ventajas que el modelo (3.2.12) por lo cual no se incluyen los resultados

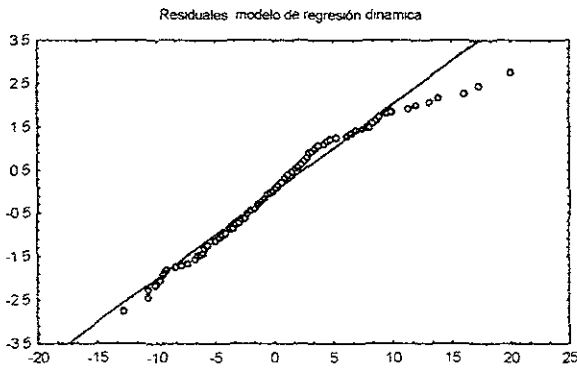
A continuación obsérvese las gráficas *FAC* y *FACP* de los residuales obtenidos del tercer modelo estimado. *Gráfica 3.30*.



Gráfica 3.30 Función de autocorrelación y de autocorrelación parcial de \hat{a}_t

En la función de autocorrelación no se observan coeficientes significativamente diferentes de cero, por lo que se puede concluir que el modelo de inercia parece ser el adecuado.

Para terminar con las pruebas de los residuales a continuación obsérvese la gráfica de normalidad para los residuales.



Gráfica 3.32 Gráfica de normalidad de residuales para el modelo III

La gráfica de normalidad muestra ciertas desviaciones sin embargo habrá que terminar de verificar todos los supuestos. quizá más adelante exista alguna otra violación que indique con más claridad el proceso a seguir para corregir el modelo. Con la gráfica de normalidad concluyen las pruebas para los residuales, a continuación se analizan los supuestos hechos a los parámetros. Obsérvese en la siguiente tabla los resultados obtenidos para los parámetros.

Modelo	Parámetros	Estimador	Error estándar	Intervalos de confianza 95%	
				Inferior	Superior
III	d_0	-17.977827	1.1196803	-19.097507	-16.858147
	d_1	4.585772	1.0879457	3.497826	5.673718
	d_2	2.264570	1.0834159	1.181154	3.347986
	d_3	0.131072	1.0821750	-0.951103	1.213247
	d_{12}	1.225209	1.0706475	0.154562	2.295857
	c_1	0.609749	0.1330601	0.476689	0.742809
	c_2	-0.048214	0.0785023	-0.126716	0.030288
	e_1	0.788469	0.1178772	0.670592	0.906346
	Θ_1	0.886269	0.0572653	0.829004	0.943534

Tabla 3.7 Resumen de parámetros obtenidos a través del modelo III

El modelo cumple con las características de estacionariedad para el proceso de inercia con parámetros en valor absoluto menor a la unidad.

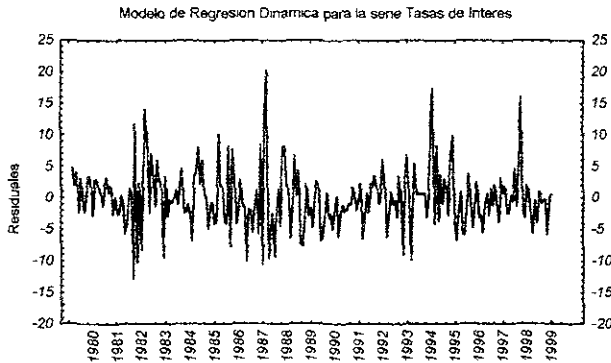
Por otro lado, hay parámetros que consideran al cero dentro de su intervalo de confianza por lo que podrían ser eliminados con el fin de tener un modelo admisible. Es así como se tiene el modelo:

$$c_1(B)y_t = d_0x_t + d_1x_{t-1} + d_2x_{t-2} + d_{12}x_{t-12} + e_1(B)\Theta_1(B^{12})a_t \quad (3.2.13)$$

Este modelo además de disminuir el número de parámetros en el modelo reduce las estadísticas *AIC* y *SBC* al igual que la varianza y la media. A continuación se muestra el resumen así como las gráficas de residuales, la estimación de los parámetros y para ver si el modelo es parsimonioso obsérvese la matriz de correlación entre los parámetros:

Modelo	Residuales	\hat{a}_t	$\hat{\sigma}_u$	AIC	SBC
IV	239	-0.066961	4.833516	1449.9571	1474.2923

Tabla 3.8 Resumen de residuales para el modelo IV



Gráfica 3.33 Gráfica de residuales para el modelo IV

En la gráfica de residuales, obsérvese que hay un claro patrón de comportamiento sexenal, en los años en que ha habido elecciones presidenciales y cambio de poder existe una o más variables políticas que afectan el comportamiento de las tasas de interés que ocasionan niveles más altos de los esperados de acuerdo al comportamiento de la inflación.

Los CETES se iniciaron a 14% en enero de 1978, subieron a 62% en marzo de 1983, llegaron a 157% en enero de 1988, en febrero de 1994 bajaron a un 10%, y volvieron a subir a en abril de 1995 a un 75%, bajando a niveles del 20% en febrero de 1997. A grandes rasgos se nota el ciclo sexenal que afecta claramente el comportamiento de las tasas de interés.

1995 a un 75%, bajando a niveles del 20% en febrero de 1997. A grandes rasgos se nota el ciclo sexenal que afecta claramente el comportamiento de las tasas de interés.

Modelo	Parámetros	Estimador	Error estándar	Intervalos de confianza 95%	
				Inferior	Superior
IV	d_0	-17.869325	1.1124202	-18.981745	-16.756905
	d_1	4.468072	1.0842110	3.383861	5.55228
	d_2	2.247798	1.0752389	1.172559	3.323037
	d_{12}	1.211200	1.0674367	0.143763	2.278637
	c_1	0.635348	0.1190427	0.516305	0.754391
	e_1	0.833737	0.0864805	0.747257	0.920218
	Θ_1	0.886450	0.0569657	0.829484	0.943415

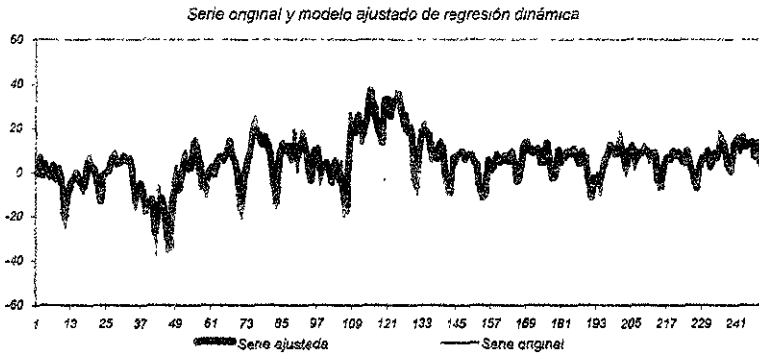
Tabla 3.8 Resumen de parámetros obtenidos a través del modelo IV

	w_0	w_1	w_2	w_{12}
w_0	1.000000	-0.122455	-0.077301	-0.065134
w_1	-0.122455	1.000000	-0.098684	-0.010227
w_2	-0.077301	-0.098684	1.000000	0.015758
w_{12}	-0.065134	-0.010227	0.015758	1.000000

Gráfica 3.33 Matriz de correlación entre los parámetros de la función de transferencia

La matriz no muestra ningún coeficiente altamente correlacionado, por lo que el modelo se puede considerar parsimonioso.

En la siguiente gráfica obsérvese a la serie original y a la serie generada a partir del modelo estimado:



Gráfica 3.35 Serie original y modelo estimado

En resumen, el modelo es aceptable conforme a las pruebas de verificación realizadas y conforme al ajuste mostrado a la serie real.

Ganancia

Como se vio anteriormente la ganancia se calcula como:

$$G = \sum_{i=0}^{\infty} v_i = -17.869325 + 4.468072 + 2.247798 + 1.211200 = -9.94225$$

lo cual quiere decir que un cambio unitario en el logaritmo de la inflación se ve reflejada totalmente transcurrido un año sobre las tasas de interés en -9.94225 .

Con esto se confirma el efecto negativo que tiene un aumento general en los precios sobre las tasas de interés reales.

Para este caso el retardo medio no puede ser interpretado ya que se requiere que todos los coeficientes de la serie $\{v_i\}$ sean todos positivos o todos negativos. Sin embargo, se puede decir que el efecto de la inflación sobre las tasas de interés es inmediato y posteriormente

empieza a nivelar durante el año principalmente en los dos meses inmediato posteriores al impacto.

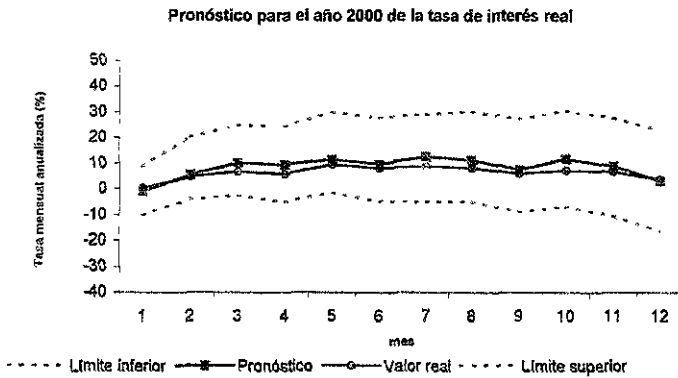
3.4 Pronóstico

La construcción de un modelo generalmente tiene un objetivo, y este es el de poder pronosticar el comportamiento de la serie en un futuro para la toma de decisiones, en la siguiente tabla se muestran los resultados del siguiente año:

<i>Fecha</i>	X_t	Y_t	<i>Intervalos de confianza Y_t</i>	
			<i>Límite inferior</i>	<i>Límite superior</i>
Enero - 2000	14.401694	-0.86086	-10.27454	8.55281
Febrero - 2000	9.564033	5.55284	-4.17988	19.97809
Marzo - 2000	7.966144	10.14332	-2.63659	24.71227
Abril - 2000	7.735887	9.39221	-5.28759	24.25108
Mayo - 2000	5.539638	11.38525	-1.50384	29.84245
Junio - 2000	6.425223	9.71288	-4.98622	27.79537
Julio - 2000	6.159509	12.80411	-5.10541	28.90944
Agosto - 2000	5.558332	11.29792	-5.13471	30.06708
Septiembre - 2000	8.488901	7.45112	-9.13848	27.05366
Octubre - 2000	6.766834	11.68796	-6.80267	30.48839
Noviembre - 2000	8.945703	8.87627	-10.56192	27.67295
Diciembre - 2000	11.496124	3.11698	-16.51933	22.7533

Tabla 3.9 Pronóstico de inflación y tasas de interés para el año 2000.

Comparando con los valores reales obtenidos se puede decir que el modelo es bueno ya que refleja claramente la relación dinámica a corto plazo y puede ser utilizado a reserva de las decisiones políticas que se ven reflejadas por el comportamiento sexenal.



Gráfica 3.36 Pronóstico comparado contra valores reales e intervalos de confianza

El pronóstico, aunque en la mayoría de los meses se muestra por encima de los valores conocidos de la serie tasas de interés, no se puede considerarse alejado e incluso la tendencia es muy semejante a la realidad, además obsérvese que los valores reales caen dentro de los intervalos de confianza. Cabe hacer notar que para el pronóstico de las tasas de interés para el año 2000 se requirió de los valores pronosticados para la inflación para el mismo año.

El pronóstico de las tasas de interés para el año 2001 se podría realizar a partir de la información obtenida hasta diciembre de 1999 sin embargo dado que actualmente se cuenta con la información real del año 2000 es importante y recomendable adaptarlas al modelo para la generación de nuevas predicciones.

Para hacer notorio la importancia de la adaptación de las predicciones a continuación se hará el ejercicio de cómo se pudieron ir adaptando las predicciones para el año 2000 con un mes de retraso, es decir una vez que se conoció el valor real de la inflación y de las tasas de interés para el mes de enero se hará el pronóstico modificado para febrero,

mes de retraso, es decir una vez que se conoció el valor real de la inflación y de las tasas de interés para el mes de enero se hará el pronóstico modificado para febrero, posteriormente una vez conocido el de febrero se obtendrá el de marzo, así sucesivamente. Estos resultados se presentan en la *tabla 3.10*.

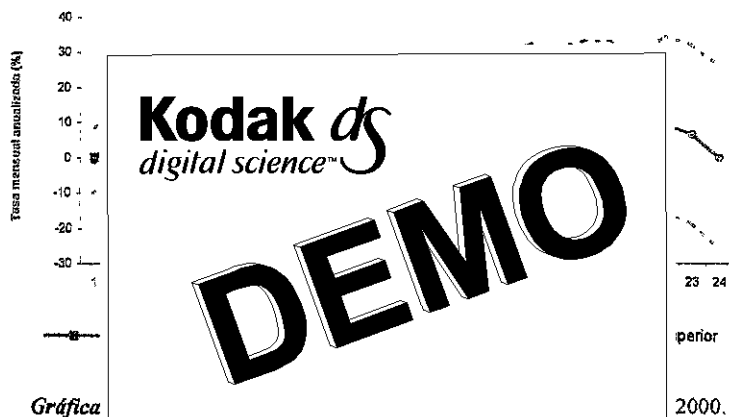
El modelo para la serie de inflación escrito en su representación general $MA(\infty)$ es la siguiente:

$$x_t = 1 + 0.91\alpha_{t-1} + 0.84\alpha_{t-2} + 0.75\alpha_{t-3} + 0.77\alpha_{t-4} + \dots \quad (3.3.1)$$

<i>Fecha</i>	X_t	Y_t	<i>Intervalos de confianza Y_t</i>	
			<i>Límite inferior</i>	<i>Límite superior</i>
Enero - 2000	14.401694	-0.8608600	-10.27453	8.55281
Febrero - 2000	10.574215	6.4192863	-5.659694	18.498266
Marzo - 2000	8.8148481	10.354345	-3.320085	24.028775
Abril - 2000	6.5078733	5.6813498	-9.08799	20.45069
Mayo - 2000	4.9895103	11.014331	-4.661309	26.689971
Junio - 2000	5.3117229	7.1692914	-9.221499	23.560081
Julio - 2000	6.5359057	9.0241769	-7.983253	26.031607
Agosto - 2000	4.3082855	9.1445799	-8.45631	26.74547
Septiembre - 2000	9.9735289	5.3693573	-12.72671	23.465427
Octubre - 2000	7.0774081	8.7103323	-9.935198	27.355862
Noviembre - 2000	10.311886	4.2588477	-14.85859	23.376288
Diciembre - 2000	13.355798	0.9482176	-18.6881	20.584538

Tabla 3.10 Pronóstico de inflación y tasas de interés para el año 2000 con información actualizada a un mes de retraso.

Pronósticos para el años 2000 y 2001



Gráfica

Fecha	X_t	Y_t	Intervalos de confianza Y_t	
			Límite inferior	Límite superior
Enero - 2000	17.446908	-4.107268	-24.48419	16.269652
Febrero - 2000	11.712732	4.0495337	-16.97637	25.075434
Marzo - 2000	7.8994238	8.417489	-13.20008	30.035059
Abril - 2000	7.8631318	6.9424389	-15.23285	29.117729
Mayo - 2000	5.823339	9.9273468	-12.8075	32.662197
Junio - 2000	7.1612177	8.128475	-15.09505	31.351995
Julio - 2000	5.6088518	9.7365812	-13.96698	33.440141
Agosto - 2000	6.2918308	9.3688682	-14.82292	33.560658
Septiembre - 2000	9.3320636	4.978753	-19.64757	29.605073
Octubre - 2000	7.2796149	9.377406	-15.71469	34.469506
Noviembre - 2000	9.3697023	6.5606732	-18.96838	32.089723
Diciembre - 2000	12.871867	-0.097398	-26.0867	25.891902

Tabla 3.10 Pronóstico de inflación y tasas de interés para el año 2001 con información actualizada a al año 2000.

Tabla 3.10 Pronóstico de inflación y tasas de interés para el año 2001 con información actualizada a al año 2000.

Al principio del año 2001 la tasa de interés se muestra negativa, sin embargo no hay que olvidar que este año se puede ver influenciado por una decisión sexenal que de acuerdo a lo observado en sexenios anteriores eleva la tasa de interés real.

La tasa de interés se ve modificada por decisiones políticas cada sexenio por la cada vez mayor competencia entre los partidos políticos, los movimientos de protesta organizados y no organizados, el papel del presidente, las políticas presupuestales, cambiarias, monetarias, comerciales y financieras, entre otras muchas. México ha alcanzado niveles de tasas de interés reales mayores que Estados Unidos de América, sin embargo el riesgo es mucho mayor y por lo cual las inversiones se deben de ver recompensadas con una tasa de interés mucho más atractiva que otros países.

Con esto se concluye el análisis de tasas de interés en México, cabe mencionar que el modelo aún puede ser mejorado si se incluyen una o más variables económicas y políticas que influyan en la serie, sobre todo porque en economía existen muchos factores externos e internos que influyen en su comportamiento, sin embargo este tema de regresión dinámica multivariado queda fuera del alcance de este trabajo.

Conclusiones

Los modelos de regresión dinámica al describir una variable a través de su propio comportamiento histórico y el de una segunda variable amplían el campo de datos que pueden ser analizados mediante la técnica de regresión, considerada una de las herramientas estadísticas más importante y utilizadas actualmente.

La característica más importante que deben de cumplir las series candidatas a formar parte de un modelo de regresión dinámica es ser *estacionarias*, lo cual significa a grandes rasgos que el comportamiento en cualquier periodo de tiempo debe de tener una media y varianza constantes así como autocorrelación igual entre una observación y otra separadas por un mismo periodo de tiempo denominado *rezago*, independiente del instante en que sean elegidas. Al principio se pudiese pensar que esta característica o condición reduce el campo de aplicación, sin embargo estos modelos pueden considerarse en ese aspecto flexibles ya que existen varias técnicas para hacer estacionarias a series que no lo son, aquí se explicaron las más comunes que son diferenciación o integración de series, transformaciones potencia para estabilización de varianza y desestacionalización.

En muchas series a analizar existen ciclos de comportamiento los cuales pueden clasificarse en ciclos cortos, medianos y largos. Los ciclos cortos pueden ser incluidos o considerados en los modelos de regresión dinámica a través de la desestacionalización de la variable, sin embargo los ciclos medianos y largos es difícil que se vean reflejados en el modelo debido a que se requerirían muchos datos históricos en donde se pueda tener una muestra considerable de estos comportamientos. Es por lo cual, la utilidad de estos modelos se ve reflejada principalmente en pronósticos a corto plazo. Es claro que, independientemente de los ciclos, entre más datos históricos se tengan para el análisis el resultado del modelo será mucho mejor, lo cual se verá reflejado en el poco error que muestren sus pronósticos con respecto a los valores reales.

Cuando la serie explicativa es conocida de antemano los modelos de regresión dinámica son de mayor utilidad, debido a que en caso contrario se debe de construir un modelo univariante para esta serie la cual esta expuesta a presentar desviaciones de la realidad.

Antes de construir un modelo se debe estar seguro de que la serie explicativa y la serie respuesta están altamente correlacionadas, en caso contrario es mejor la utilización de un modelo para series de tiempo univariantes. También, cabe hacer notar que este tipo de modelos están vinculados totalmente con el tiempo, por lo cual cuando una variable se quiere explicar como consecuencia del comportamiento de otra debe de tenerse en cuenta el factor del tiempo, de lo contrario, es mejor utilizar un modelo de regresión clásico especializado en reflejar situaciones estáticas, debido a que presentara las mismas ventajas.

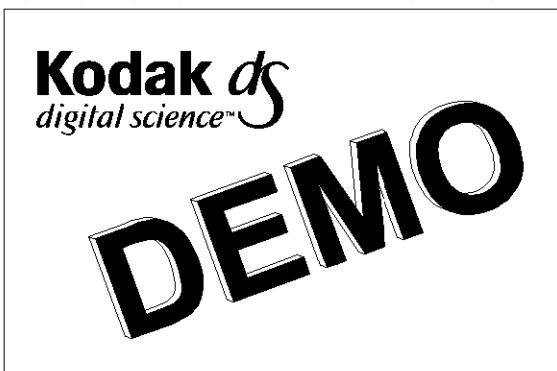
Por todo lo estudiado anteriormente es importante que cuando se construya un modelo de regresión dinámica se sigan las etapas de identificación, estimación, verificación y predicción en el orden que se sugiere, para poder asegurar que el modelo es el más adecuado.

En el estudio realizado en el presente trabajo se supuso que existe una alta correlación entre la serie inflación y la serie tasas de interés de tal manera que el comportamiento de la primera pudiera influir considerablemente en el comportamiento de la segunda. Como se pudo demostrar efectivamente la serie tasas de interés responde ante cualquier cambio en la inflación desde el primer instante y pierde fuerza transcurrido un año aproximadamente. Sin embargo, también se pudo demostrar que el comportamiento de las tasas de interés no solo depende de una sola variable, en este caso la inflación, sino que hay una o más variables que influyen en su comportamiento sobre todo a finales y principios del denominado ciclo sexenal.

Obsérvese además que los efectos de la inflación en general son malos para algunas variables económicas pero en este caso a través de este análisis se demostró que cualquier aumento en ella se ve reflejado en un decrecimiento de las tasas de interés reales.

Los CETES se iniciaron a 14% en enero de 1978, subieron a 62% en marzo de 1983, llegaron a 157% en enero de 1988, en febrero de 1994 bajaron a un 10%, y volvieron a subir a en abril de 1995 a un 75%, bajando a niveles del 20% en febrero de 1997. A grandes rasgos se nota el ciclo sexenal que afecta claramente el comportamiento de las tasas de interés.

Hay otros factores internos y externos que en los últimos años. Entre los factores internos económicos. En lo político, el ciclo se ve afectado por los partidos políticos, el papel cambiante de los grupos organizados y no organizados, y los trámites para tener un sistema político abierto y competitivo. En lo monetaria, cambiara, comercial y financiera. Los factores políticos y económicos pueden resumir en el efecto de la globalización.



Hay una gran interacción entre las economías, afectando o beneficiando, de alguna manera las situaciones por las que atraviesen otros países.

Cabe mencionar que durante los últimos veinte años, los rendimientos de las inversiones mexicanas han estado entre los más altos del mundo, incluso la tasa de rendimiento del CETE ha estado por arriba del Treasury bills (Tbills) de E.U. medidos en un mismo lapso. No obstante, estos altos rendimientos se han visto acompañados de un riesgo mucho mayor debido a la inestabilidad económica y política por la que a atravesado el país y las medidas que se han tenido que tomar para elevar las tasas de interés con el propósito de incrementar las inversiones. En resumen, el inversionista en México ha tenido la oportunidad de percibir rendimientos extraordinarios durante los últimos veinte años, pero con un alto nivel de riesgo.

Kodak *ds*
digital science™

DEMO

Bibliografía

Libros

- ☐ Anderson, T.W. (1971). *The Statical Analysis of Time Series*. Wiley & Sons, Jhon.
- ☐ Anderson, O.D. (1976). *Time Series Analysis and Forecasting. The Box-Jenkins aproach*. Butterworths. London and Boston.
- ☐ Bartlett, M.S. (1950). *Stochastic processes*. Cambridge University Press, Cambridge.
- ☐ Braun, Martín (1983). *Ecuaciones diferenciales y sus aplicaciones*. Grupo Editorial Iberoamérica.
- ☐ Bowerman, Bruce L.; O'Connell, Richard T. (1993). *Forecasting and Time Series. An applied approach*. Third Edition. Wadsworth Publishing Company. Belmont, California.
- ☐ Box, George E.P. and Jenkins, Gwilyn M. (1970). *Time Series: forecasting and control*. San Francisco Holden-Day.
- ☐ Burden, Richard L.; Douglas Faires J. (1985). *Análisis Numérico*. Grupo Editorial Iberoamérica.
- ☐ Chatfield, Christopher. (1979). *The analysis of time series: An Introduction*. Second Edition.
- ☐ Derrick, William R. (1986). *Ecuaciones diferenciales con aplicaciones*.
- ☐ Dhrymes, Phoebus (1968). *Time series, unit roots, and cointegration*. Academic Press.
- ☐ Diggle, Peter J. (1990). *Time Series. A Biostatistical Introduction*. Oxford Statistical Science Series.5.
- ☐ Granger, C.W.J. and Newbold, P. (1977). *Forecasting Economic Time Series*. Academic Press, New York.
- ☐ Grossman, Stanley I. (1981). *Elementary Differential Equations with Applications*. Second Edition.
- ☐ Guerrero, Victor M. (1991). *Análisis estadístico de series de tiempo económicas*. Universidad Autónoma Metropolitana, colección cbi.
- ☐ Harvey. (1981). *The econometric Analysis of Time Series*. Phillip Allan.

- ↳ Heyman Timothy. *Inversión en la globalización. Análisis y administración de las nuevas inversiones mexicanas* Editorial Milenio, S.A. de C.V.
- ↳ Judge et al. (1980). *The theory and practice of Econometrics*. Wiley.
- ↳ Judge, Hill, Griffiths, et al. (1980). Introduction to *The theory and practice of Econometrics*. Wiley.
- ↳ Kendall, Maurice and Keith, J. (1990). *Time Series*. Edward Arnold.
- ↳ Maddala, G.S. (1996). *Introducción a la Econometría*. Segunda Edición. Prentice-Hall Hispanoamericana, S.A.
- ↳ Mood, Alexander M. y Graybill, Franklin A. (1955). *Introducción a la Teoría de la Estadística*. McGraw – Hill Book Company
- ↳ Nemytskii and Stepanov. (1989). *Qualitative theory of Differential Equations*. General Publishing Company.
- ↳ Nerlove, Marc; Grather, David M.; Carvalho, José L. (1995). *Analysis of economic time series. A synthesis*.
- ↳ Novales, A. (1988). *Econometría*.
- ↳ Peña Sánchez de Rivera, Daniel. (1992). *Estadística Modelos y métodos*. Alianza Editorial.
- ↳ Peña Sánchez de Rivera, Daniel. (1992). *Estadística Modelos y métodos 2. Modelos lineales y series temporales*. Alianza Editorial.
- ↳ Wie, William W.S.; Addison Wesley. (1990). *Time Series Analysis Univariate and Multivariate Methods*.

Artículos

- Bartlett, M. S. (1946). "On the theoretical specification of sampling properties of autocorrelated time series". *Journal of Royal Statistical Society (Suppl.)* 7,211.
- Box, G.E.P. y Cox, D.R. "An analysis of transformations". *Journal of the Royal Statistical Society B-26*, 211.

- Guerrero, Víctor M. y Rojas, F. Javier. (1987). “Desestacionalización de Series de Tiempo Económicas: Una introducción a la economía”. Documento No. 54. Banco de México.
- Guerrero, Víctor M. y Rojas, F. Javier. (1987). “Desestacionalización de Series de Tiempo Económicas: Aplicación a los Indicadores de la actividad Industrial”. Documento No. 63. Banco de México.
- Hannan, E. J. (1963). “Regression for Time Series Analysis” Proceedings of a Symposium in Time Series Analysis. Wiley, New York.
- Hannan, E. J.; and Nicholls, D. F. (1972). “The estimation of mixed regression, autoregression, moving average, and distributed lag models”. *Econometrica* 40, 529-547.
- Haugh, Larry D. And Box, G. E. P. “Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Series”. *Journal of the American Statistical Association*.
- Nerlove, Marc (1972a). “Lags in economic behavior”. *Econometrica* 40, 221-251.
- Pérez López, Alejandro; Schwartz, Moisés J. (1999). “Inflación y Ciclos Económicos”. Documento de Investigación No. 9904. Dirección General de Investigación Económica. Banco de México.
- Quenouille, M.H. (1948). “Approximate test of correlation in time-series”. *Journal of the Royal Statistical Society B-11*, 68.
- Sims, Christopher A. (1974). “Seasonality in regression”. *Journal of the American Statistical Association* 69, 618-626.

Internet

- ☞ <http://inegi.gob.mx>
- ☞ <http://www.banxico.org.mx>
- ☞ <http://www.shcp.gob.mx>
- ☞ <http://www.mexico.businessline.com>
- ☞ <http://www.infolatna.com.mx>
- ☞ <http://www.cppure.com>