2 cy

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES

"A C A T L A N"

"UNA APLICACION DEL METODO DE BOX - JENKINS A INGRESOS POR TURISMO EN MEXICO"

T E S I S

QUE PARA OBTENER EL TITULO DE:

LICENCIADO EN MATEMATICAS APLICADAS Y COMPUTACION

P R E S E N T A :

ALEJANDRA ALANIS URIBE

Director: ACT. Ma. del Carmen González Videgaray

MEXICO, D. F. 1989

I N D I C E

INTRODUCCION

La necesidad de tener mayor precisiòn en los nuevos pronòsticos de los fenòmenos (econòmicos, financieros, fìsicos,...) ha propiciado el desarrollo de metodologìas,algoritmos y proyectos de "software". El anàlisis de series de tiempo ha venido creciendo creciendo poco a poco con el fin de analizar datos, predecir valores futuros, apoyar a la toma de decisiones y optimizar pèrdidas y ganancias.

Esta tesis tiene por objetivo el anàlisis de los Ingresos por Turismo en Mèxico utilizando el mètodo de pronòstico de Box y Jenkins. El modelo resultante que representa el comportamiento de Ingresos por Turismo es un modelo multiplicativo que contiene parte autorregresiva y parte estacional de medias mòviles.

AR(1) x SMA(1)

Este anàlisis se realiza utilizando dos paquetes de predicciòn, FORECAST PLUS y STATGRAPHICS, y se comparan los resultados obtenidos de cada uno de ellos presentando ventajas y desventajas.

En el capítulo 1 se presentan las bases teóricas de series de tiempo univariadas. Modelos Estacionarios, modelos estacionales, funciones de autocorrelación y autocorrelación parcial (teóricas y muestrales) hasta llegar a la parte correspondiente a pronósticos.

En el capítulo 2 se dan a conocer todo lo referente a los Ingresos por Turismo en México en los últimos 10 años,antecedentes, encuestas y estadísticas para conocer a fondo el fenómeno en estudio. Los datos que aparecen en este capítulo fueron proporcionados por la biblioteca y la hemeroteca del Banco de México.

En el capítulo 3 se explicarán por separado cada uno de los paquetes con el fin de dar a conocer como funcionan antes de realizar la aplicación y la comparacion de éstos.

En el capítulo 4 se realiza el análisis de predicción utilizando la teoría presentada en los capítulos anteriores así como la comparación entre los dos paquetes para ver cual es el más eficiente.

CAPITULO 1

CONCEPTOS BASICOS DE SERIES DE TIEMPO

La importancia de tener un buen modelo para pronóstico ha venido creciendo poco a poco ya que uno de los objetivos de analizar datos es predecir los valores futuros de las variables en las que se está interesado. Las situaciones en las que son requeridos los pronósticos varían ampliamente, por lo que se han desarrollado varias técnicas las cuales se dividen en dos categorias: cuantitativas y cualitativas. Las técnicas cuantitativas pueden aplicarse cuando se cumplen 3 condiciones:

1.- Se tiene información sobre el pasado.
2.- Esta información puede cuantificarse en forma de datos (numéricos).
3.- Puede asumirse que el patrón del pasado continuará en el futuro.

Estas técnicas cuantitativas varían considerablemente, ya que se fueron desarrollando para diferentes propósitos, cada una tiene sus propiedades, su cuidado y sus costos, que se tienen que considerar cuando se elige una técnica específica. Los procedimientos cuantitativos se dividen en 2 tipos:

1. Métodos Intuitivos: utilizan una extrapolación horizontal, estacional o de tendencia y se basan en la experiencia empírica. Estos métodos son fáciles de usar pero no son exactos, lo cual ha hecho que su uso haya venido disminuyendo. Sin embargo algunos negocios siguen utilizando los métodos intuitivos ya sea ó por que no conocen otros métodos ó porque prefieren usar aproximaciones subjetivas para pronosticar.

2. Métodos Formales: Estos métodos cada vez son más usados gracias a la disminución del uso de los métodos intuitivos y a su facilidad de uso en las computadoras. Pueden también incluir extrapolación, pero ésto se da en un camino estándar, usando una aproximación sistemática que pretende minimizar los errores de predicción. Existen varios métodos formales que son poco costosos y fáciles de usar y que además se pueden aplicar de manera

mecánica. Estos métodos se utilizan cuando es necesario pronosticar sobre un gran número de productos y cuando los errores de pronóstico de un producto no sean demasiado costosos.

Las personas que no están familiarizadas con métodos de predicción cuantitativos, frecuentemente piensan que el pasado no puede describir de manera exacta el futuro ya que todo esta cambiando constantemente. Después de cierta familiaridad con los datos, se ve con más claridad que nada queda igual, que la historia se repite sólo en cierto sentido.La aplicación del mejor método puede identificar las relaciones entre el factor a ser pronósticado y el tiempo mismo(o algunos otros factores) haciendo así que el pronóstico sea lo más exacto posible.

Para clasificar los métodos cuantitativos de pronóstico se considera fundamentalmente el modelo del que se trata en, vez del nivel de teoría estadística en la que se apoya cada método. Los modelos para pronóstico se clasifican en dos

a.- Series de Tiempo: La predicción del futuro se basa en valores pasados de la variable y/ó errores pasados. El objetivo es descubrir un patrón dentro de la serie de datos históricos y extrapolar este patrón al futuro.

b.- Modelos de Regresión (modelos causales): Asumen que el factor a ser pronosticado exhibe una relación causa-efecto mediante una ó más variables independientes. El propósito es descubrir la relación y usarla para pronosticar valores futuros de la variable dependiente.

Ambos tipos de modelos tienen sus ventajas, los modelos de series de tiempo son de mayor facilidad para pronosticar mientras que,los modelos causales pueden usarse con mayor éxito para la toma de decisiones. Todo esto se resume de manera esquemática en las figuras 1.1 y 1.2.

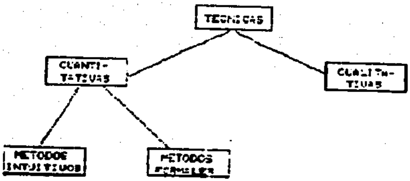


figura 1.1

figura 1.2

SERIES DE TIEMPO

DEFINICION: Una serie de tiempo es un conjunto de observaciones generadas en forma secuencial a travès del tiempo. La t-èsima observaciòn se denota por:

$$Y_t \quad , \quad y_t = Y_t - \hat{\mu}$$

donde $\hat{\mu}$ es el estimador de la media (generalmente se denota por Yt)

Las series de tiempo se clasifican en

1. **UNIVARIADA:** Descripciòn de una sola variable.
2. **MULTIVARIADA:** Descripciòn de màs de una variable.

En una serie de tiempo se pueden distinguir cuatro componentes ò tipos de patròn:

a.- TENDENCIA

Se define como el movimiento de arriba hacia abajo (o viceversa) de la serie de tiempo que se aproxima a una función. Por ejemplo, si tenemos un comportamiento que va desde los negativos a los positivos formando una recta (como lo muestra la figura 1.3), la serie tiene una tendencia lineal.

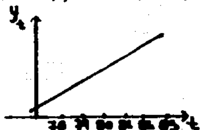


figura 1.3 tendencia lineal creciente

La tendencia se ve reflejada en el comportamiento de las funciones de autocorrelación y autocorrelación muestral, ésto se ve con más énfasis posteriormente.

b.- VARIACION ESTACIONAL

Cuando la serie se vê influenciada por factores estacionales (de estaciones), se dice que la serie tiene un comportamiento estacional ó que es una serie estacional. Usualmente se refiere a patrones anuales, ó cuatrimestrales, Un ejemplo grâfico lo muestra la figura 1.4

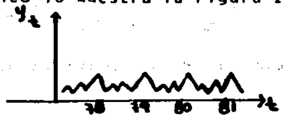


figura 1.4 variación estacional

c.- CICLO

Movimiento recurrente que se repite en intervalos mayores de un año. Este patrôn existe cuando los datos son influenciados por largas fluctuaciones que pueden estar asociadas, por ejemplo, con ciclos de negocios de alguna empresa. La diferencia entre estacional y cíclico es que la primera es de una duración constante y recurre en una base periódica regular, mientras que la segunda varía en duración y magnitud, obsérvese la figura 1.5.

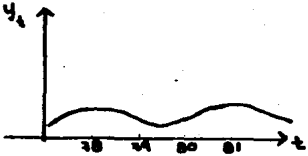


figura 1.5 Ciclo

d.- <u>VARIACION ALEATORIA</u>

Esta variación se debe a factores externos ó aleatórios. Este patrón se presenta cuando los valores de los datos fluctûan alrededor de una media constante . Si se eliminan los tres componentes anteriores resulta una variación aleatória. También se dice que es un patrón horizontal el que presenta la serie:

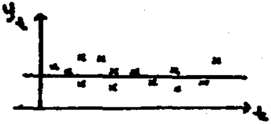


figura 1.6 variación aleatoria

Se puede presentar el caso de tener una serie de datos que incluye combinaciones de los cuatro patrones anteriores, que es lo que generalmente sucede, es decir

$$Y_t = \text{patrón} + \text{error aleatorio}$$

patrón puede ser estacionalidad, ciclo,etc

Cuando se tiene una serie que es completamente aletoria se define como RUIDO BLANCO,y se representa como:

$$Y_t = e_t$$

e_t = error aleatorio ($e_t \sim N(0,\sigma^2)$)

El objetivo del análisis de series de tiempo es

1. Describir el fenómeno en estudio.
2. Explicar el comportamiento del fenómeno.
3. Pronosticar datos futuros.
4. Controlar toma de decisiones.

ESTACIONARIDAD

Se dice que una serie es ESTACIONARIA cuando NO sigue una tendencia, NO contiene ciclos, NI es estacional, unicamente contiene fluctuaciones aleatorias. Si se tiene una serie que cumple con lo anterior, la media y la varianza de la serie son CONSTANTES. Gráficamente se observa que si la serie es estacionaria los datos deberán comportarse horizontalmente, a lo largo del eje de las x's como lo muestra la figura 1.7

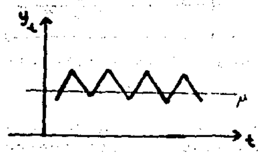


figura 1.7 serie estacionaria

Se dice que una serie Y es estacionaria cuando cumple con

i. **Media Constante**

$E(Y_t) = Y_t$

ii. **Varianza Constante** :

$$E(Y_t - E(Y_t))^2 = E[Y_{t+h} - E(Y_{t+h})]^2 = \sigma_Y^2$$

$$\hat{\sigma}_Y^2 = s^2 = \frac{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}{n}$$

iii. La Autocorrelación sólo depende del intervalo k. La autocorrelación en el intervalo k se define como

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\{var(Y_t)\, var(Y_{t+k})\}^{1/2}} \quad , \quad -1 \le \rho_k \le 1$$

$$\text{Si: } cov(Y_t, Y_{t+k}) = \gamma_0 \quad y \quad var(Y_t) = var(Y_{t+k}) = \gamma_1 \;, \Rightarrow \rho_k = \frac{\gamma_0}{\gamma}$$

El caso del modelo de RUIDO BLANCO

$$Y_t = e_t$$

Donde

e_t =error aleatorio

$e \sim N(0, \sigma^2)$

$E(e_t\, e_s)=0$ para toda $s \neq t$ (los errores son independientes)

es estacionario ya que cumple que la media y la varianza son constantes para toda t y las autocorrelaciones sólo dependen del intervalo k.

Cabe hacer notar que la estacionaridad(1) garantiza que no existen cambios significativos en la estructura del modelo, de lo contario sería difícil (ó imposible) predecir valores futuros; por lo tanto si la serie original NO ES estacionaria habrà que hacerla estacionaria por medio de transformaciones y/ó diferencias.

1.1 EL METODO DE BOX-JENKINS

Ya se mencionaron los elementos de una serie de tiempo, ahora corresponde tratar el método de pronóstico para series de tiempo, Box y Jenkins pero antes es necesario tomar en cuenta los siguientes conceptos:

- Un modelo puramente Autorregresivo indica que el valor actual depende de valores pasados.
- Un modelo de Medias Móviles indica que los promedios

(1)Es importante no confundir **ESTACIONARIDAD** con **ESTACIONALIDAD**.

de los errores pasados son usados para calcular nuevos valores.

- Un modelo ARMA es un modelo Auterregresivo y de Medias Mòviles.
- Un modelo ARIMA es un modelo Autorregresivo y de Medias Mòviles con transformaciones y/ò diferencias.

Los modelos Autorregresivos (Autoregresive (AR)) fueron primeramente introducidos por Yule (1926) y poco despuès generalizados por Walker(1931), mientras que los modelos de Medias Mòviles (Moving Average (MA)) primero fueron usados por Stultzky (1937). Fuè Wold (1938) el que provee fundamentos teòricos del proceso combinado ARMA. Los modelos ARMA se han desarrollado en dos direcciones, una es la identificaciòn eficiente y procedimientos de estimaciòn (para AR, MA y modelos mixtos) y la otra es la extensiòn de los resultados para incluir modelos estacionales.

Los modelos Autorregresivos y de Medias Mòviles (ARMA) fueron estudiados de manera extensiva por George Box y Gwilym Jenkins (1970) y sus nombres han sido usados como sinònimos de el proceso general ARMA aplicado al anàlisis de series de tiempo, predicciòn y control. Box y Jenkins juntaron, de una manera màs comprensiva, la informaciòn relevante requerida para estudiar y usar modelos ARMA de series de tiempo univariadas.

El mètodo de Box-Jenkins, para modelos de series de tiempo,es un mètodo para encontrar un modelo ARIMA, para un conjunto de datos, que represente adecuadamente el comportamiento de èstos.

La idea que maneja Box-Jenkins es el tomar un modelo e ir pasandolo por el filtro autorregresivo, por el de medias mòviles, hasta resultar el modelo de ruido blanco.

Esquemàticamente:

figura 1.8

Este método es un poco más complicado que algunos otros métodos de pronóstico, sin embargo es uno de los más utilizados. ya que ha resultado uno de los más eficientes. Actualmente ya está incluido en muchos paquetes estadísticos, lo cual hace que se conozca y se aplique a diferentes tipos de problemas. Este método comprende 3 etapas:

1.-IDENTIFICACION

El objetivo de esta etapa es seleccionar un modelo que represente el comportamiento de los datos. En esta etapa es en donde se identifica si existe tendencia, ciclos y/ó estacionalidad; y se obtiene una serie estacionaria en caso de ser necesario.

Se puede identificar un modelo examinando las funciones de autocorrelación (ACF) y autocorrelación parcial muestrales (PACF) (ver sec. 1.2.3), sin embargo este procedimiento requiere del juicio de cada persona que no siempre da el mejor resultado de identificación, ya que

1. Las funciones de autocorrelación y autocorrelacion parcial pueden no indicar claramente un modelo específico.

2. Pueden indicar más de un modelo.

Por ésto, los modelos que sugieren las ACF y las PACF son llamados modelos "tentativos", los cuales son realmente el resultado de esta etapa; ésto se debe a que es difícil identificar un modelo definitivo cuando se trata de datos reales. Una sugerencia para llevar cierto orden dentro de esta etapa es seguir los siguientes tres puntos:

a.- OBTENER UNA SERIE ESTACIONARIA:

Si la serie de tiempo original no es estacionaria resultará una autocorrelación degenerada que será un obstáculo para la identificación. Para lograr la estacionaridad de la serie habrá que realizar EL NUMERO APROPIADO de diferencias (no sobrediferenciar porque se complica el modelo) y/ó tomar alguna transformación.

b.- EXAMINAR LAS AUTOCORRELACIONES Y AUTOCORRELACIONES

PARCIALES (de preferencia graficamente):

Una forma pràctica de identificar el modelo es observar la gràfica de las autocorrelaciones y autocorrelaciones parciales ya que para cada tipo de modelo tienen un comportamiento especial. Por ejemplo, si los coeficientes de la fn. de autocorrelaciòn disminuyen hasta cero de forma exponencial (i.e. rapidamente a cero)podrìa tratarse de un modelo autorregresivo (AR). Si este mismo comportamiento se observa en la fn de autocorrelaciòn parcial podrìa tratarse de un modelo de medias mòviles (MA). Y si lo mismo se observa en las dos gràficas de las funciones se trata de un modelo autorregresivo y de medias mòviles (ARMA).(VER SECC. 1.2)

c.- EXAMINAR LAS CORRELACIONES DE LOS RESIDUALES:

El analizar las gràficas de las correlaciones de los residuales ayuda a detectar què es lo que falta por considerar en el modelo tentativo ò què es lo que sobra. Si el modelo ya es el adecuado las correlaciones se tendràn que comportar como ruido blanco, es decir, los coeficientes tendràn que tener valores aproximadamente igual a cero. El encontrar valores significativamente distintos de cero puede ayudar a determinar el orden de los modelos AR, MA ò ARMA ò a incluir tèrminos AR ò MA no considerados anteriormente.

2.-ESTIMACION

Una vez identificados los modelos "tentatïvos" se tendràn que estimar los paràmetros que minimicen la suma de cuadrados de los residuales (diferencia entre el valor observado y el valor estimado de la serie). Si el modelo es puramente autorregresivo los paràmetros pueden ser estimados como en el caso de regresiòn mùltiple, pero si el modelo incluye medias mòviles la minimizaciòn de los errores de la fn. de verosimilitud requiere mètodos de optimizaciòn no lineales.

3.-EXAMEN DEL MODELO

En esta ùltima etapa se verifica si el modelo es el adecuado. Para èsto se pueden hacer dos cosas: realizar el anàlisis de los residuales, ya que el comportamiento de èstos refleja si falta algo por considerar en el modelo y sobre especificar el modelo, i.e ya teniendo identificado y estimado el modelo ARMA(p,q), se puede estimar un modelo ARMA(p+1,q) y un ARMA(p,q+1) y observar si el paràmetro que se agregó es significativo ó no.

Es recomendable tener cuidado con no complicar el modelo, por lo que es de gran utilidad comenzar con el modelo lo más sencillo posible que pudiera reflejar el comportamiento en estudio e ir agregando parámetro por parámetro. La etapa de Examen del Modelo también se conoce como DIAGNOSTICO del modelo.

Para terminar, una vez que se pasó por estas tres etapas y se tiene ya el modelo se procede a hacer el pronóstico de nuevos valores.

Falta mencionar un concepto de gran importancia que siempre se deberá tener en cuenta para el análisis de series de tiempo que es el concepto de PARQUEDAD (parsimony) que nos dice que se utilice el menor número de parámetros posible en el ajuste del modelo para un conjunto de datos. Dicho de otra manera, al ajustar un modelo se deberá tratar que éste sea el más sencillo posible pero que a su vez represente el comportamiento del fenómeno.

Por último, para actualizar el modelo cuando hay datos nuevos, se pueden realizar dos cosas:

i) Incluir estos datos nuevos dentro de las observaciones que ya se tienen y seguir pronosticando con el modelo anteriormente ajustado, es decir suponer que estos datos nuevos no cambian el modelo ya encontrado.

ii) Incluir estos datos nuevos y ver si con ellos cambia el modelo, es decir realizar todo el procedimiento incluyendo las nuevas observaciones.

1.2 IDENTIFICACION DE MODELOS.

Para poder identificar el modelo adecuado que represente un comportamiento en particular es necesario conocer los tipos de modelos con los que se cuenta. Es por esto que en este tema se verân detenidamente todos ô casi todos los tipos más comunes y sus características, así como el comportamiento de sus funciones de autocorrelación y autocorrelación parcial. Para entender mejor la etapa de identificación se efectúa la explicación empezando con el modelo más sencillo hasta el más complejo.

1.2.1 MODELOS ESTACIONARIOS

A) MODELOS AUTORREGRESIVOS

El modelo AUTORREGRESIVO de orden p, AR(p) ô ARMA(0,1), se representa como

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \dots + \emptyset_p Y_{t-p}$$

$\emptyset_i$ es el parámetro autorregresivo asociado a Y_{t-i} .
$e_t \sim N(0, \sigma^2)$
$E(e_t, e_s) = 0 \qquad s \neq t$
$E(Y_{t-i}, e_t) = 0$

MODELO AUTORREGRESIVO DE PRIMER ORDEN

AR(1) se expresa como

$$Z_t = \emptyset Z_{t-1} + e_t$$

Se utiliza Z_t ya que pueden no ser los datos originales, es decir, se realizaron diferencias y/ô transformaciones para tener una serie estacionaria. Y se omite el término δ porque se supone Z_t desviaciones con respecto a la media.

En seguida se analizarán cada una de las condiciones de estacionaridad y así fijar, en caso de ser necesario, restricciones para los parámetros y así poder lograr la estacionaridad.

1. MEDIA CONSTANTE

$E(Z_t)=0$

$E(\phi_1 Z_{t-1}+e_t)=0$

en este caso no hay restricción para ϕ_1 ya que con cualquier valor que tome puede llegar a cumplirse esta igualdad.

2. VARIANZA CONSTANTE

$$Var(Z_t)= E(Z_t - E(Z_t))^2 = E(Z_t^2) - [E(Z_t)]^2$$

COMO $E(Z_t)=0$

$$Var(Z_t)= E(Z_t^2)$$
$$= E(\phi_1 Z_{t-1}+e_t)^2 = E(\phi_1^2 Z_{t-1}^2 + 2\phi_1 Z_{t-1}e_t + e_t^2)$$
$$= \phi_1^2 E(Z_{t-1}) + 2\phi_1 E(Z_{t-1}e_t) + E(e_t^2)$$

COMO El PROCESO ES ESTACIONARIO

$$E(Z_t^2) = E(Z_{t-1}^2) = var\, Z_t$$
$$Var(Z_t) = \phi_1^2 Var(Z_t) + \sigma^2$$

$$var(Z_t) = \frac{\sigma^2}{1-\phi_1^2}$$

En este caso sí existe restricción para ϕ_1, para que la varianza sea positiva.

$$\phi_1 < 1$$

entonces la condición de estacionaridad es

$$/\phi_1/ < 1 \quad ó \quad -1 < \phi_1 < 1$$

3. AUTOCORRELACIÓN

Sea

$$\gamma_0 = var(Z_t) = E(Z_t^2)$$
$$\gamma_1 = cov(Z_t\ Z_{t-1}) = E(Z_t\ Z_{t-1})$$
$$\gamma_2 = cov(Z_t\ Z_{t-2}) = E(Z_t\ Z_{t-2})$$

La covarianza dividida entre la varianza es igual al coeficiente de autocorrelación ,que es el que se quiere encontrar.

$$\frac{\gamma_1}{\gamma_0} = \frac{cov}{var} = \rho_1 \quad , \quad \frac{\gamma_2}{\gamma_0} = \rho_2 \quad , \quad \ldots$$

$$\gamma_1 = E(Z_t Z_{t-1}) = E[(\phi_1 Z_{t-1} + e_t) Z_{t-1}]$$

$$\gamma_1 = E(\phi_1 Z_{t-1}^2 + e_t Z_{t-1}$$

$$\gamma_1 = \phi_1 E(Z_{t-1}^2) + E(Z_{t-1} e_t)$$

$$\gamma_1 = \phi_1 E(Z_{t-1}^2)$$

$$\gamma_1 = \phi_1 \gamma_0$$

$$\frac{\gamma_1}{\gamma_0} = \underline{\underline{\phi_1 = \rho_1}}$$

$$\gamma_2 = E(Z_t Z_{t-2}) = E[(\phi_1 Z_{t-1} + e_t) Z_{t-2}]$$

$$\gamma_2 = E(\phi_1 Z_{t-1} Z_{t-2} + Z_{t-2} e_t)$$

$$\gamma_2 = \phi_1 E(Z_{t-1} Z_{t-2}) + E(Z_{t-2} e_t)$$

$$\gamma_2 = \phi_1 \gamma_1$$

$$\gamma_2 = \phi_1 (\phi_1 \gamma_0)$$

$$\gamma_2 = \phi_1^2 \gamma_0 \quad , \quad \frac{\gamma_2}{\gamma_0} = \underline{\underline{\phi_1^2 = \rho_2}}$$

En general, si el proceso es AR(1)

$$\rho_k = \begin{cases} 1 , & k = 0 \\ \emptyset_1^k , & k > 0 \end{cases}$$

como $/ \emptyset_1 / < 1$ la fn.de autocorrelación será decreciente en forma exponencial tendiendo a cero como se observa en la figura 1.9 del ejemplo siguiente.

EJEMPLO:

AR(1) $\quad Zt = 0.8Z_{t-1} + e_t$

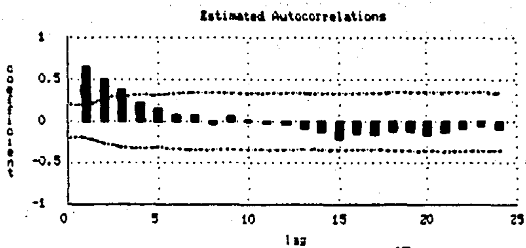


figura 1.9
ACF AR(1)

FUNCION DE MEMORIA

Ahora se expresará a $Z_t = \emptyset_1 Z_{t-1} + e_t$ en términos de los errores encontrando la forma GOLPE O CHOQUE ALEATORIO(2) que se obtiene mediante sustituciones sucesivas Z_t , Z_{t-1} ....

$$Z_{t-1} = \emptyset_1 Z_{t-2} + e_{t-1}$$

(2) La forma golpe aleatorio de AR(1) es un proceso de medias móviles de orden infinito

sustituyendo en Z_t

$$Z_t = \phi_1(\phi_1 Z_{t-2} + e_{t-1}) + e_t$$
$$= \phi_1^2 Z_{t-2} + \phi_1 e_{t-1} + e_t$$

$$Z_{t-2} = \phi_1 Z_{t-3} + e_{t-2}$$

ahora

$$Z_t = \phi_1^2(\phi_1 Z_{t-3} + e_{t-2}) + \phi_1 e_{t-1} + e_t$$
$$= \phi_1^3 Z_{t-3} + \phi_1^2 e_{t-2} + \phi_1 e_{t-1} + e_t$$
$$\vdots$$
$$Z_t^{(n)} = \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \cdots + e_t$$

Esta ecuación resultante es la FORMA GOLPE O CHOQUE ALEATORIO (RANDOM SHOCK, por su asociación con modelos físicos de choques de partículas), del modelo que es equivalente a AR(1).

Los ϕ_i's indican cual es la influencia de los errores en Z_t. Si el proceso es estacionario ($|\phi_1| < 1$) el efecto de los errores se disipa. A los coeficientes de los errores se les llama COEFICIENTES DE MEMORIA y a su gráfica FUNCION DE MEMORIA.

MODELO AUTORREGRESIVO DE ORDEN MAYOR

AR(p) se expresa como

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + e_t$$

Para fines prácticos se analizará un proceso AR(2) ya que para p>2 los cálculos se complican.

Sea un modelo AR(2)

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t$$

1. MEDIA CONSTANTE

$E(Z_t)=0$

$E(\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t)=0$

$E(Z_t) = \phi_1 E(Z_{t-1}) + \phi_2 E(Z_{t-2}) + e_t = 0$

Dado que se está hablando de un proceso estacionario hay que recordar que $E(Z_t)=E(Z_{t-1})=\ldots=E(Z_{t-p})=0$ por lo que claramente se observa que nada implica restricción para ϕ_1 y ϕ_2.

2. VARIANZA CONSTANTE

$var(Z_t) = \gamma_0 = E(Z_t^2)$

$$Var(Z_t) = E[(\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t) Z_t]$$
$$= E(\phi_1 Z_t Z_{t-1} + \phi_2 Z_t Z_{t-2} + e_t Z_t)$$
$$= \phi_1 E(Z_t Z_{t-1}) + \phi_2 E(Z_t Z_{t-2}) + E(e_t Z_t)$$
$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2 \quad \ldots \quad (1)$$

$$\gamma_1 = E(Z_t Z_{t-1})$$
$$= E[(\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t) Z_{t-1}]$$
$$= E(\phi_1 Z_{t-1}^2 + \phi_2 Z_{t-1} Z_{t-2} + e_t Z_{t-1})$$
$$= \phi_1 E(Z_{t-1}^2) + \phi_2 E(Z_{t-1} Z_{t-2}) + E(e_t Z_{t-1})$$
$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1 \quad \ldots \quad (2)$$

$$\gamma_2 = E(Z_t Z_{t-2})$$
$$= E[(\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t) Z_{t-2}]$$
$$= E(\phi_1 Z_{t-1} Z_{t-2} + \phi_2 Z_{t-2}^2 + Z_{t-2} e_t)$$
$$= \phi_1 E(Z_{t-1} Z_{t-2}) + \phi_2 E(Z_{t-2}^2) + E(Z_{t-2} e_t)$$
$$\gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0 \quad \ldots \quad (3)$$

DE (2)

$$\gamma_1 - \phi_2 \gamma_1 = \phi_1 \gamma_0$$
$$\gamma_1 (1 - \phi_2) = \phi_1 \gamma_0$$

$$\underline{\gamma_1 = \frac{\phi_1}{(1-\phi_2)} \gamma_0} \quad , \quad \underline{\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi_1}{(1-\phi_2)}}$$

SUSTITUYENDO γ_1 EN (3)

$$\gamma_2 = \phi_1 \left(\frac{\phi_1}{1-\phi_2} \gamma_0 \right) + \phi_2 \gamma_0$$

$$\underline{\gamma_2 = \gamma_0 \left(\frac{\phi_1^2}{1-\phi_2} + \phi_2 \right)} \quad , \quad \underline{\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1^2}{(1-\phi_2)} + \phi_2}$$

SUSTITUYENDO γ_1 y γ_2 EN (1)

$$\gamma_0 = \phi_1\left(\frac{\phi_1}{1-\phi_2}\gamma_0\right) + \phi_2\left(\frac{\phi_1^2}{1-\phi_2} + \phi_2\right)\gamma_0 + \sigma^2$$

$$= \frac{\phi_1^2}{1-\phi_2}\gamma_0 + \frac{\phi_1^2\phi_2}{1-\phi_2}\gamma_0 + \phi_2^2\gamma_0 + \sigma^2$$

$$\gamma_0 = \left(\frac{\phi_1^2(1+\phi_2)}{(1-\phi_2} + \phi_2^2\right)\gamma_0 + \sigma^2$$

$$\frac{\phi_1^2(1+\phi_2)}{1-\phi_2} \cdot \frac{1+\phi_2}{1+\phi_2} = \frac{\phi_1^2(1+\phi_2)^2}{(1-\phi_2^2)}$$

$$\gamma_0 = \left(\frac{\phi_1^2(1+\phi_2)^2}{(1-\phi_2^2)} + \phi_2^2\right)\gamma_0 + \sigma^2$$

$$\gamma_0\left(1 - \frac{\phi_1^2(1+\phi_2)^2}{(1-\phi_2^2)} - \phi_2^2\right) = \sigma^2$$

$$\underline{\gamma_0 = \frac{\sigma^2}{\left(1 - \frac{\phi_1^2(1+\phi_2)^2}{(1-\phi_2^2)} - \phi_2^2\right)}}$$

PARA QUE $\gamma'_0 > 0$, ES NECESARIO QUE

$$1 - \frac{\phi_1^2(1+\phi_2)^2}{(1-\phi_2^2)} - \phi_2^2 > 0$$

$$\frac{\phi_1^2(1+\phi_2)^2}{(1-\phi_2^2)} + \phi_2^2 < 1 \quad \ldots (4)$$

CONDICION 1.- $(1-\phi_2^2) > 0$, $|\phi_2| < 1$

$$\boxed{-1 < \phi_2 < 1}$$

MULTIPLICANDO (4) por $(1-\phi_2^2)$

$$\phi_1^2(1+\phi_2)^2 + \phi_2^2(1-\phi_2^2) < (1-\phi_2^2)$$
$$\phi_1^2(1+\phi_2)^2 + \phi_2^2(1-\phi_2^2) - (1-\phi_2^2) < 0$$
$$(1-\phi_2^2)(\phi_2^2-1) + \phi_1^2(1+\phi_2)^2 < 0$$
$$-(1-\phi_2^2)^2 + \phi_1^2(1+\phi_2)^2 < 0$$
$$\phi_1^2(1+\phi_2)^2 < (1-\phi_2^2)^2$$

OBTENIENDO LA RAIZ CUADRADA

$$\phi_1(1+\phi_2) < (1-\phi_2^2)$$
$$\phi_1 < \frac{1-\phi_2^2}{(1+\phi_2)} \quad , \quad \phi_1 < \frac{(1-\phi_2)(1+\phi_2)}{(1+\phi_2)}$$

$$\phi_1 < (1-\phi_2)$$

$$\phi_1 + \phi_2 < 1$$

CONDICION 2.-

$$\boxed{\phi_1 + \phi_2 < 1}$$

Se obtuvieron 2 condiciones para los parâmetros del modelo AR(2):

condición 1: $/ \phi_2 / < 1$
condición 2: $\phi_1 + \phi_2 < 1$

3. AUTOCORRELACION

En el punto anterior ya se obtuvieron las covarianzas , . Ahora, para obtener los coeficientes de autocorrelación es necesario encontrar

$$\frac{\gamma_1}{\gamma_0} , \frac{\gamma_2}{\gamma_0} , \ldots$$

$$\frac{\gamma_1}{\gamma_0} = \rho_1 = \frac{\phi_1}{1-\phi_2} \qquad , \qquad \frac{\gamma_2}{\gamma_0} = \rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$$

En general

$$\rho_k = \begin{cases} \phi_1 / 1 - \phi_2 \ , & k=1 \\ \phi_1^2 / 1 - \phi_2 + \phi_2 \ , & k>1 \end{cases}$$

Como es obvio, si el calcular todo lo anterior para un AR(2) fuê laborioso, el aumentar el orden (p) del modelo estos câlculos serân cada vez mâs complicados.

Por último, el comportamiento de la función de autocorrelación de un modelo AR(p) decaerâ a cero en forma exponencial. Este comportamiento se puede observar las figuras 1.10 y 1.11 del siguiente ejemplo.

EJEMPLOS:

AR(2) $Zt = -0.6Z_{t-1} + 0.2Z_{t-2} + e_t$

ACF:

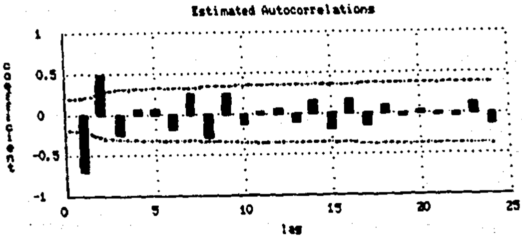


figura 1.10
ACF AR(2)

AR(2) $Zt = -0.8Z -0.6Z + e$

ACF:

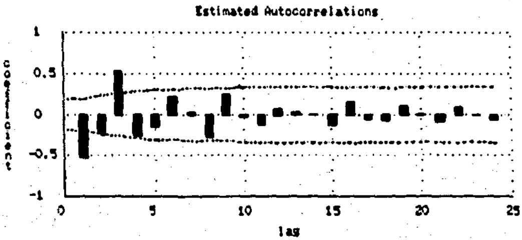


figura 1.11
ACF AR(2)

B) MODELOS DE MEDIAS MOVILES

El modelo de MEDIAS MOVILES de orden q, MA(q) ó ARMA(0,1), se representa como

$$Y_t = e_t - \Theta_1 e_{t-1} - \Theta_2 e_{t-2} - \dots - \Theta_q e_{t-q}$$

Θ_i es el parâmetro de medias môviles asociado a e_{t-i}, y el signo menos se introduce por convención.

Recuêrdese que
$e_t \sim N(0, \sigma)$
$E(e_t, e_s) = 0 \quad \forall \quad s \neq t$
$E(Y_{t-1}, e_t) = 0$

MODELO DE MEDIAS MOVILES DE PRIMER ORDEN

MA(1) se expresa como

$$Z_t = e_t - \Theta_1 e_{t-1}$$

Expresando MA(1) en têrminos de las observaciones y no de los errores es necesario encontrar la ecuaciôn en forma invertida:

$$e_t = Z_t + \Theta_1 e_{t-1}$$
$$e_{t-1} = Z_{t-1} + \Theta_1 e_{t-2}$$

entonces

$$Z_t = e_t - \Theta_1 (Z_{t-1} + \Theta_1 e_{t-2})$$
$$Z_t = e_t - \Theta_1 Z_{t-1} - \Theta_1^2 e_{t-2}$$
$$e_{t-2} = Z_{t-2} + \Theta_1 e_{t-3}$$
$$Z_t = e_t - \Theta_1 Z_{t-1} - \Theta_1^2 (Z_{t-2} + \Theta_1 e_{t-3})$$
$$Z_t = e_t - \Theta_1 Z_{t-1} - \Theta_1^2 Z_{t-2} - \Theta_1^3 e_{t-3}$$
$$\vdots$$
$$Z_t = e_t - \Theta_1 Z_{t-1} - \Theta_1^2 Z_{t-2} - \Theta_1^3 Z_{t-3} - \dots$$

Esta ecuación es la Ecuacion en Forma Invertida del MA(1), la cual corresponde a un AR(∞). Para que esta ecuación tenga algún sentido se tiene que cumplir que $|\theta_1| < 1$, que es la condición de invertibilidad, entonces los coeficientes que explican la influencia de las Z_t's en el valor actual serán cada vez más pequeños.

Analizando ahora para este modelo las condiciones de estacionaridad:

1. MEDIA CONSTANTE.

$$E(Z_t) = 0$$
$$E(e_t - \theta_1 e_{t-1}) = 0$$

como se observa no existe condición para θ_1.

2. VARIANZA CONSTANTE

$$\gamma_0 = E[(e_t - \theta_1 e_{t-1})^2] = E(e_t^2 - 2\theta_1 e_t e_{t-1} + \theta_1^2 e_{t-1}^2)$$
$$= E(e_t^2) - 2\theta_1 E(e_t e_{t-1}) + \theta_1^2 E(e_{t-1}^2)$$
$$\gamma_0 = \sigma^2 + \theta_1^2 \sigma^2 = \underline{\sigma^2(1+\theta_1^2)}$$

No existe restricción para θ_1 ya que para cualquier valor se cumple que $\gamma_0 > 0$ y será constante para toda Z_t.

3. AUTOCORRELACION

$$\gamma_1 = E(Z_t Z_{t-1}) = E[(e_t - \theta_1 e_{t-1}) Z_{t-1}]$$
$$= E(e_t Z_{t-1}) - \theta_1 E(Z_{t-1} e_{t-1})$$
$$E(Z_{t-1} e_{t-1}) = E(e_{t-1}^2 - \theta_1 e_{t-2} e_{t-1}) = \sigma^2$$

$$\underline{\gamma_1 = -\theta_1 \sigma^2} \quad , \quad \underline{\underline{\frac{\gamma_1}{\gamma_0} = \rho_1 = \frac{-\theta_1}{(1+\theta_1^2)}}}$$

$$\gamma_2 = E(Z_t Z_{t-2}) = E[(e_t - \theta_1 e_{t-1}) Z_{t-2}]$$
$$= E(e_t Z_{t-2} - \theta_1 Z_{t-2} e_{t-1}) = E(e_t Z_{t-2}) - \theta_1 E(Z_{t-2} e_{t-1})$$
$$\gamma_2 = 0$$
$$\vdots$$
$$\gamma_k = 0 , \quad k > 1$$

Generalizando, los coeficientes de autocorrelación para MA(1) quedan:

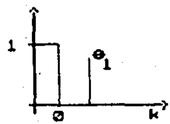
$$\rho_k = \begin{cases} \dfrac{-\theta_1}{(1+\theta_1^2)} & , \quad k=1 \\ 0 & , \quad k>1 \end{cases}$$

El comportamiento de la función de autocorrelación, será un valor significativo en k=1 y para k > 1, ρ_k tiene el valor de cero. Y será negativo ó positivo dependiendo del valor de θ_1. Un ejemplo de este comportamiento se observan en la figura 1.12 .

FUNCION DE MEMORIA.

En este caso el modelo ya está expresado en términos de los errores por lo que la función de memoria se obtiene directamente de la definición:

$$Z_t = a_t - \theta_1 a_{t-1}$$



funcion de memoria
MA(1)

EJEMPLO:

MA(1) $Zt = e_t + 0.8e_{t-1}$

ACF:

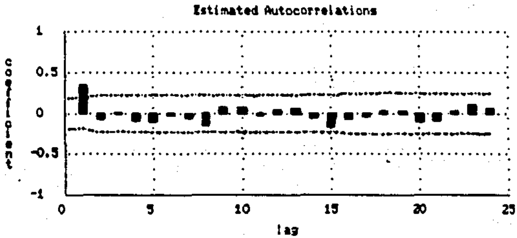


figura 1.12
ACF MA(1)

MODELOS DE MEDIAS MOVILES DE ORDEN MAYOR

De la misma forma que en autorregresivos se toma el modelo de segundo orden, q=2, para facilitar los càlculos.

El modelo MA(2) se expresa como

$$Z_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

Para encontrar la ecuaciòn en forma invertida:

$$e_t = Z_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$
$$e_{t-1} = Z_{t-1} + \theta_1 e_{t-2} + \theta_2 e_{t-3}$$

sustituyendo en Z_t.

$$Z_t = e_t - \theta_1 (Z_{t-1} + \theta_1 e_{t-2} + \theta_2 e_{t-3}) + \theta_2 e_{t-2}$$
$$Z_t = e_t - \theta_1 Z_{t-1} - (\theta_1^2 + \theta_2) e_{t-2} - \theta_1 \theta_2 e_{t-3}$$
$$\vdots$$
$$Z_t = e_t - \theta_1 Z_{t-1} - (\theta_1^2 + \theta_2) Z_{t-2} - \theta_1 (\theta_1^2 + 2\theta_2) e_{t-3} - \dots$$

Los coeficientes de los têrminos autorregresivos son

$$\left.\begin{array}{l} -\theta_1 \\ -(\theta_1^2+\theta_2) \\ -\theta_1(\theta_1^2+2\theta_2) \\ \vdots \end{array}\right\}$$ lo que interesa es que estos coeficientes sean cada vez menores.

Las condiciones de invertibilidad para MA(2) son

1. $\theta_2 + \theta_1 < 1$
2. $\theta_2 - \theta_1 < 1$
3. $/ \theta_2 / < 1$

Para el modelo general MA(q), q > 2, las condiciones de invertibilidad son complicadas y la estacionaridad està asegurada.

1. MEDIA CONSTANTE.

$E(Z_t)=0$

$E(a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2})=0$

No exite restricciôn alguna para θ_1 ni para θ_2.

2. VARIANZA CONSTANTE.

$$\gamma_0' = E(Z_t^2) = E\left[(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})\right]$$

$$= E(a_t^2 + \theta_1^2 a_{t-1}^2 + \theta_2^2 a_{t-2}^2)$$

$$= \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2$$

$$\underline{\gamma_0' = \sigma^2(1+\theta_1^2+\theta_2^2)}$$

No hay restricción para θi.

3. AUTOCORRELACION.

SACANDO LAS AUTOCOVARIANZAS.

$$\gamma_1 = E(Z_t Z_{t-1}) = E\left[(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})(e_{t-1} - \theta_1 e_{t-2} - \theta_2 e_{t-3})\right]$$

$$\gamma_1 = E(e_t e_{t-1} - \theta_1 e_t e_{t-2} - \theta_2 e_t e_{t-3} - \theta_1 e_{t-1}^2 + \theta_1^2 e_{t-1} e_{t-2} + \theta_1 \theta_2 e_{t-1} e_{t-3} - \theta_2 e_{t-2} e_{t-1} + \theta_1 \theta_2 e_{t-2}^2 + \theta_2^2 e_{t-2} e_{t-3})$$

$$\gamma_1 = -\theta_1 E(e_{t-1}^2) + \theta_1 \theta_2 E(e_{t-2}^2) = -\theta_1 \sigma^2 + \theta_1 \theta_2 \sigma^2$$

$$\underline{\underline{\gamma_1 = \theta_1(\theta_2 - 1)\sigma^2}}$$

$$\gamma_2 = E(Z_t Z_{t-2}) = E\left[(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})(e_{t-2} - \theta_1 e_{t-3} - \theta_2 e_{t-4})\right]$$

$$\gamma_2 = E(e_t e_{t-2} - \theta_1 e_t e_{t-3} - \theta_2 e_t e_{t-4} - \theta_1 e_{t-1} e_{t-2} + \theta_1^2 e_{t-1} e_{t-3} + \theta_1 \theta_2 e_{t-1} e_{t-4} - \theta_2 e_{t-2}^2 + \theta_1 \theta_2 e_{t-3} e_{t-2} + \theta_2^2 e_{t-2} e_{t-4})$$

$$\gamma_2 = E(-\theta_2 e_{t-2}^2) \qquad \underline{\underline{\gamma_2 = -\theta_2 \sigma^2}}$$

$$\gamma_3 = E(Z_t Z_{t-3}) = E\left[(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})(e_{t-3} - \theta_1 e_{t-4} - \theta_2 e_{t-5})\right]$$

$$\gamma_3 = E(e_t e_{t-3} - \theta_1 e_t e_{t-4} - \theta_2 e_t e_{t-5} - \theta_1 e_{t-1} e_{t-3} + \theta_1^2 e_{t-1} e_{t-4} + \theta_1 \theta_2 e_{t-1} e_{t-5} - \theta_2 e_{t-2} e_{t-3} + \theta_1 \theta_2 e_{t-2} e_{t-4} + \theta_2^2 e_{t-2} e_{t-5})$$

$$\underline{\gamma_3 = 0}$$

$$\frac{\gamma_1}{\gamma_0} = \frac{\theta_1(\theta_2 - 1)}{(1+\theta_1^2+\theta_2^2)}$$

$$\underline{\rho_1 = \frac{\theta_1(\theta_2 - 1)}{(1+\theta_1^2+\theta_2^2)}}$$

$$\frac{\gamma_2}{\gamma_0} = \frac{-\theta_2}{(1+\theta_1^2+\theta_2^2)}$$

$$\underline{\rho_2 = \frac{-\theta_2}{(1+\theta_1^2+\theta_2^2)}}$$

Claramente se ve que $\gamma_3=0$ ya que al efectuar el producto todos los $E(e_t e_s)=0$ ya que $s \neq t$. Entonces

$$\rho_3 = \gamma_3 / \gamma_0 = 0$$

en general para k queda:

$$\rho_k = \begin{cases} \theta_1(\theta_2 - 1)/1+\theta_1^2+\theta_2^2 & , k=1 \\ -\theta_2 / 1 + \theta_1^2 + \theta_2^2 & , k=2 \\ 0 & , k>=3 \end{cases}$$

De aquí se generaliza que para un MA(q), la función de autocorrelación se trunca después de q, dicho de otra forma, para k > q, $\rho_k=0$. Obsèrvese la figura 1.13 del siguiente ejemplo.

EJEMPLO:

$MA(2)\ Zt = e_t - 1.4\, e_{t-1} + 0.6\, e_{t-2}$

ACF:

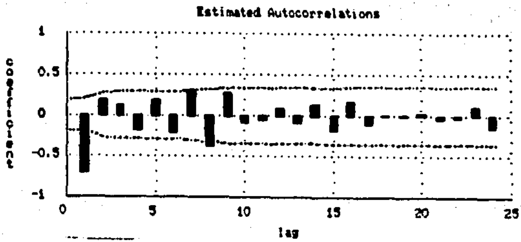


figura 1.13
ACF MA(2)

FUNCION DE MEMORIA

Es claro entonces que la función de memoria estará explicada sólo hasta los q términos.

C) MODELOS MEZCALDOS AUTORREGRESIVOS Y DE MEDIAS MOVILES ARMA(p,q)

Una extensión del modelo AR y MA, es una clase de modelo que contiene términos autorregresivos y términos de medias móviles. Tales modelos se conocen como MODELOS AUTORREGRESIVOS Y DE MEDIAS MOVILES de orden (p,q) (p por los términos AR y q por los términos MA), ARMA(p,q) y se expresan como

MODELOS MEZCLADOS DE PRIMER ORDEN ARMA(1,1)

El modelo mixto màs sencillo ARMA(1,1) se expresa como

$$Z_t = \phi_1 Z_{t-1} + e_t - \theta_1 e_{t-1}$$

analizando las condiciones de estacionaridad se encuentra que ahora existen restricciones para los paràmetros tanto de estacionaridad como de invertibilidad.

1. MEDIA CONSTANTE.

$E(Z_t)=0$

$E(\phi_1 Z_{t-1} + e_t - \theta_1 e_{t-1})=0$

en este caso no existe restricciòn para los paràmetros.

2. VARIANZA CONSTANTE.

$$\gamma_0 = E(Z_t^2) = E[(\phi_1 Z_{t-1} + e_t - \theta_1 e_{t-1})(\phi_1 Z_{t-1} + e_t - \theta_1 e_{t-1})]$$

$$= E(\phi_1^2 Z_{t-1}^2 + \phi_1 Z_{t-1} e_t - \phi_1 \theta_1 Z_{t-1} e_{t-1} + \phi_1 e_t Z_{t-1} + e_t^2 - \theta_1 e_t e_{t-1} - \theta_1^2 Z_{t-1} e_{t-1} - \theta_1 e_t e_{t-1} + \theta_1^2 e_{t-1}^2)$$

$$\gamma_0 = \phi_1^2 \gamma_0 - 2\phi_1 \theta_1 \sigma^2 + \sigma^2 + \theta_1^2 \sigma^2$$

$$\gamma_0 - \phi_1^2 \gamma_0 = \sigma^2 (1 + \theta_1^2 - 2\phi_1 \theta_1)$$

$$\gamma_0 = \frac{\sigma^2 (1 + \theta_1^2 - 2\phi_1 \theta_1)}{(1 - \phi_1^2)}$$

para que var(Z_t) > 0 se requiere que

$$(1 + \theta_1^2 - 2\phi_1 \theta_1) > 0 \qquad y \qquad (1 - \phi_1^2) > 0$$

$$-2\phi_1 \theta_1 + \theta_1^2 > -1 \qquad \qquad \phi_1^2 < 1$$

$$\theta_1 (2\phi_1 + \theta_1) < 1 \qquad \qquad |\phi_1| < 1$$

entonces la condiciòn de estacionaridad resultante es

CONDICION 1: $/ \phi_1 / < 1$

3. AUTOCORRELACION

$$\gamma_1 = E(Z_t Z_{t-1}) = E[(\phi_1 Z_{t-1} + e_t - \theta_1 e_{t-1}) Z_{t-1}]$$
$$= E(\phi_1 Z_{t-1}^2 + Z_{t-1} e_t - \theta_1 Z_{t-1} e_{t-1})$$

$$\gamma_1 = \phi_1 \gamma_0 - \theta_1 \sigma^2$$

El coeficiente de autocorrelación será

$$\frac{\gamma_1}{\gamma_0} = \frac{\phi_1 \gamma_0}{\gamma_0} - \frac{\theta_1 \sigma^2}{\gamma_0} = \phi_1 - \frac{\theta_1 \sigma^2}{(\theta_1^2 + 1 - 2\phi_1 \theta_1)}$$

$$\rho_1 = \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{(1 + \theta_1^2 - 2\phi_1 \theta_1)}$$

En general

$$\gamma_k = \phi_1 \gamma_{k-1} = \phi_1^k \gamma_1$$

$$\underline{\rho_k = \phi_1 \gamma_{k-1} = \phi_1^k \gamma_1}$$

Para encontrar la condición de invertibilidad

$$Z_t = \phi_1 Z_{t-1} + e_t - \theta_1 e_{t-1}$$
$$e_t = Z_t - \phi_1 Z_{t-1} + \theta_1 e_{t-1}$$
$$e_{t-1} = Z_{t-1} - \phi_1 Z_{t-2} + \theta_1 e_{t-2}$$

SUSTITUYENDO EN Z_t

$$Z_t = \phi_1 Z_{t-1} + e_t - \theta_1 (Z_{t-1} - \phi_1 Z_{t-2} + \theta_1 e_{t-2})$$
$$= \phi_1 Z_{t-1} + e_t - \theta_1 Z_{t-1} + \phi_1 \theta_1 Z_{t-2} - \theta_1^2 e_{t-2}$$
$$Z_t = (\phi_1 - \theta_1) Z_{t-1} + \phi_1 \theta_1 Z_{t-2} + e_t - \theta_1^2 e_{t-2}$$
$$\vdots$$
$$Z_t = (\phi_1 - \theta_1) Z_{t-1} + \theta_1 (\phi_1 - \theta_1) Z_{t-2} + \theta_1^2 (\phi_1 - \theta_1) Z_{t-3} + \dots + e_t$$

se llegó a un modelo AR(∞), y la condición de invertibilidad es que

CONDICION 2: $|\theta_1| < 1$

Ahora la función de autocorrelación queda

$$\rho_k = \begin{cases} (\phi_1 - \theta_1)(1-\phi_1\theta_1)/(1-2\phi_1\theta_1 + \theta_1^2), & k=1 \\ \phi_1^{k-1}\rho_1 & , k>1 \end{cases}$$

la gráfica es decreciente en forma exponencial, obsérvese la figura 1.14 del siguiente ejemplo.

EJEMPLO:

ARMA(1,1) $Z_t = 0.7Z_{t-1} + e_t + 0.3\, e_{t-1}$

ACF:

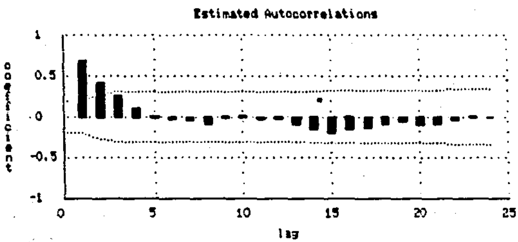


figura 1.14
ACF ARMA(1,1)

MODELOS MEZCLADO DE ORDEN MAYOR ARMA(p,q)

Los resultados obtenidos en ARMA(1,1) se generalizan para el modelo mixto de orden arbitrario. Cualquier modelo ARMA puede expresarse como un modelo MA de orden infinito. Similarmente puede expresarse como un AR(∞) siempre y cuando se cumplan las condiciones de invertibilidad. El modelo AR y el MA vienen a ser un caso particular de estos modelos mezclados, por ejemplo un AR(3) corresponde a un ARMA(3,0).

El modelo ARMA más general se expresa como

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

La función de autocorrelación de este modelo, ARMA(p,q) se comporta como la de un modelo AR(p) a partir del intervalo q+1, antes de ésto, refleja la parte de medias móviles

En general, la gráfica de un modelo ARMA decrece en forma exponencial (Recuérdese que los valores de k pueden también ser negativos). Observe este comportamiento en la figura 1.15, que es la gráfica de la función de autocorrelación de un modelo ARMA.

EJEMPLO:

ARMA(1,1) $Z_t = -0.5Z_{t-1} + e - 0.5e_{t-1}$

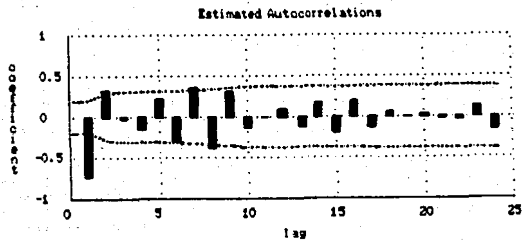


figura 1.15
ACF ARMA(1,1)

AUTOCORRELACION PARCIAL.

Las autocorrelaciones parciales se usan para medir el grado de asociación entre Y_t y Y_{t-k} cuando los efectos de lo demás retardos (Y_{t-1},.. $Y_{(t-1)}$) se eliminan.

El coeficiente de autocorrelación parcial para un intervalo k se denota como ρkk. El principal propósito de la función de autocorrelación parcial, en el análisis de series de tiempo es ayudar a identificar un modelo ARMA para predecir.

NOTA: La función de autocorrelación (acf) y la función de autocorrelación parcial (pacf), son sencillamente una gráfica de los valores correspondientes de k y kk para toda k, y su interpretación ayuda a visualizar cuando la serie de tiempo es estacionaria.

Las autocorrelaciones parciales se definen como el último término autorregresivo de un modelo AR(p); así son las p autocorrelaciones para cualquier proceso AR, ésto queda expresado en las siguientes ecuaciones

$$Z_t = \hat{\phi}_1 Z_{t-1} + e_t \quad \ldots \quad (a)$$

$$Z_t = \phi_1 Z_{t-1} + \hat{\phi}_2 Z_{t-2} + e_t \quad \ldots \quad (b)$$

$$\vdots$$

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \hat{\phi}_n Z_{t-n} \ldots (c)$$

Este sistema de ecuaciones nos dice que al construir un modelo autorregresivo puede checarse si la inclusión de una nueva Z_t en el modelo representará los datos en forma más adecuada. Supóngase que después de ajustar un modelo AR(k-1) se desea ver si lo adecuado es un AR(k); si el valor obtenido de /Øk/ es "grande", significa que debe incluirse Z_{t-k}. Este coeficiente mide "el exceso" de autocorrelación no tomada en cuenta en el modelo anterior. Dicho de otra forma, Øk mide el efecto "parcial" de Z_{t-k} para explicar el comportamiento de Z_t en un modelo que ya incluye $Z_{t-1}, \ldots, Z_{t-k+1}$. A este coeficiente autorregresivo se le conoce como COEFICIENTE DE AUTOCORRELACION PARCIAL (PACF) en el intervalo k, y se denota como ρkk (Økk).

PACF DE MODELOS AUTORREGRESIVOS

Si se trata de un modelo AR(1) solamente $\hat{\phi}_1$ será significativo mientras que $\hat{\phi}_2$, $\hat{\phi}_3$, ..., $\hat{\phi}_m$ no serán significativos (estadísticamente significativos).

Por lo tanto, en general, si se tienen p autocorrelaciones parciales significativas, se trata de un modelo AR(p). Observese en los ejemplos siguientes las figuras 1.16 y 1.17.

EJEMPLO:

AR(1) $Z_t = 0.8Z_{t-1} + e_t$

PACF:

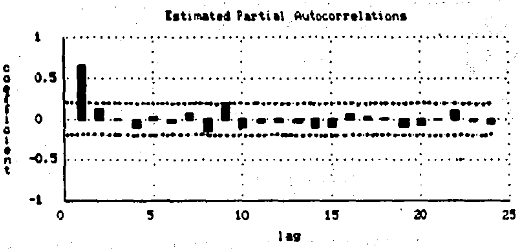


figura 1.16
PACF AR(1)

AR(2) $Z_t = -0.8Z_{t-1} - 0.6Z_{t-2} + e_t$

PACF:

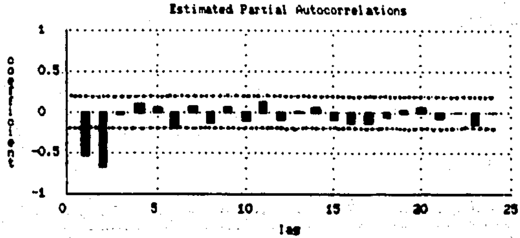


figura 1.17
PACF AR(2)

La función de autocorrelación parcial de los modelos autorregresivos es igual a la función de autocorrelación de los modelos de medias móviles.

PACF DE MODELOS DE MEDIAS MOVILES

Para este caso, habrà que reescribir el modelo en forma invertida, ya que por la definición de coeficiente de autocorrelación parcial es necesario tener el modelo en tèrminos de las observaciones pasadas.

Sea el modelo MA(1)

$$Z_t = e_t - \theta_1 e_{t-1}$$

reescribiendolo en forma invertida (ver modelos de medias moviles)

$$Z_t = -\theta_1 Z_{t-1} - \theta_1^2 Z_{t-2} - \theta_1^3 Z_{t-3} \ldots + e_t$$

Ahora Z_t es un modelo AR(∞) por lo que las autocorrelaciones parciales mostrarán un comportamiento similar a las autocorrelaciones de un modelo autorregresivo, es decir, decaen a cero en forma exponencial. Esto se generaliza para un modelo MA(q). En la figura 1.18 se puede observar el comportamiento de un modelo MA(1).

EJEMPLO:

MA(2) $Zt = e_t - 1.4e_{t-1} + 0.6e_{t-2}$

PACF:

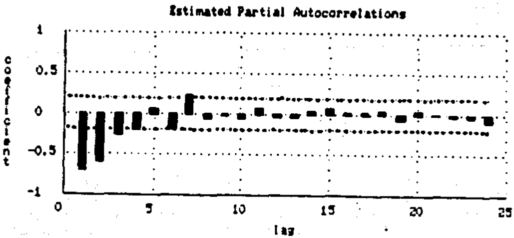


figura 1.18
PACF MA(2)

PACF DE MODELOS MEZCLADOS

Tómese un ARMA(1,1)

$$Z_t = \phi_1 Z_{t-1} + e_t - \theta_1 e_{t-1}$$

habrá que expresarlo en términos de las observaciones pasadas, entonces la forma invertida es:

$$Z_t = e_t + \underbrace{(\phi_1 - \theta_1)}_{\phi_1} Z_{t-1} + \underbrace{\theta_1(\phi_1 - \theta_1)}_{\phi_2} Z_{t-2} + \underbrace{\theta_1^2(\phi_1 - \theta_1)}_{\phi_3} Z_{t-3} + \ldots$$

se necesita que los coeficientes vayan disminuyendo por lo que la función de autocorrelación parcial decrece en forma exponencial. Esto se puede observar en el siguiente ejemplo, en la figura 1.19, que corresponde a un modelo ARMA(1,1).

EJEMPLO:

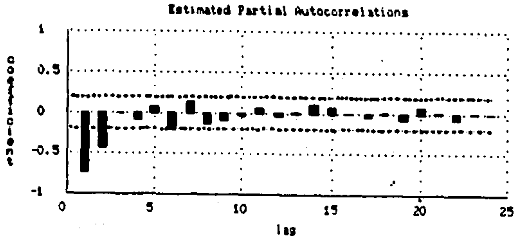


figura 1.19
PACF ARMA(1,1)

Por último, hay que recordar que la gráfica de las autocorrelaciones es completamente distinta a la gráfica de los datos. La gráfica de los datos ayuda a visualizar el patrón de los datos, mientras que la gráfica de las autocorrelaciones sumariza el patrón existente en los datos y revela ciertas características de ellos. En al siguiente tabla se muestra, a manera de resumen, como se comportan las acf's y pacf's de cada uno de los modelos.

MODELO	A C F	P A C F
AR(p)	decrece	se trunca despues de p
MA(q)	se trunca despues de q	decrece
ARMA(p,q)	decrece	decrece

tabla A

1.2.2 MODELOS NO ESTACIONARIOS ARIMA(p,d,q)

Cuando la función de autocorrelación y autocorrelación parcial decaen a cero muy lentamente se está frente a un proceso NO ESTACIONARIO y esta no estacionaridad también se observa en la gráfica de los datos originales ya que habrá una media y/o una varianza no constantes. Como un ejemplo sea el modelo AR(1):

$$Z_t = Z_{t-1} + e_t$$

$\phi_1 = 1$, por lo tanto es no estacionario (condición de estacionaridad, $|\phi_1| < 1$). Este modelo tiene una función de memoria constante. Para hacer que la serie sea estacionaria se pueden hacer diferencias ordinarias ($\nabla Z_t = Z_t - Z_{t-1}$), ésto se debe a que el modelo sigue un comportamiento de un polinomio de grado "n" el cual contiene un término constante al momento de restar se elimina éste término y se va estacionarizando la serie (realizar el número de diferencias necesario).

$$\nabla Z_t = Z_t - Z_{t-1}$$
$$Z_t - Z_{t-1} = W_t$$
$$W_t - W_{t-1} = V_t \text{ ⟶ SEGUNDAS DIFERENCIAS.}$$

obteniendo las primeras diferencias

$$Z_t - Z_{t-1} = Z_{t-1} + e_t - Z_{t-1}$$
$$Z_t - Z_{t-1} = e_t$$

si $Z_t - Z_{t-1} = W_t \qquad \Rightarrow \qquad W_t = e_t$

Ahora W_t será un nuevo modelo, que en este caso en particular resultó ser RUIDO BLANCO, y sí es estacionario.

En general, el número de diferencias para hacer el modelo estacionario, variará dependiendo de cada caso. Para regresar al modelo original, unicamente se realiza la operación inversa (suma). Del ejemplo anterior

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

que es el modelo original.

Estos modelos son AUTORREGRESIVOS INTEGRADOS Y DE MEDIAS MOVILES que se denotan por ARIMA(p,d,q) (d=diferencias).

Continuando con los modelos no estacionarios, los modelos ARIMA(p,d,q) se expresan como:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

donde W_t se obtiene por medio de la diferencia de orden "d" de Z_t.

$$w_t = \begin{cases} Z_t , & d=0 \\ \nabla^d Z_t , & d>0 \end{cases}$$

También se puede expresar el modelo utilizando el operador de salto hacia atrás. Sea

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

si se separan los términos autorregresivos de los de medias móviles

$$w_t - \phi_1 w_{t-1} - \phi_2 w_{t-2} - \dots - \phi_p w_{t-p} = e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

$$\Rightarrow \quad \phi(B)W_t = \theta(B)e_t \qquad \text{ARIMA } (p,d,q)$$

donde $B\ W_t = W_{t-1}$ (OPERADOR DE SALTO HACIA ATRAS)

Si no se está trabajando con desviaciones con respecto a la media, el modelo ARIMA será

$$\phi(B)\, w_t = \delta + \theta(B) e_t$$

δ no es necesariamente igual a la media, ya que

$$E(w_t) = E\left(\frac{\delta + \theta(B) e_t}{\phi(B)}\right) = E\left(\phi(B)^{-1}(\delta + \theta(B) e_t)\right)$$

δ será igual a la media, unicamente cuando $\phi p=0$ para toda p, es decir cuando el modelo sea una ARIMA(0,d,q) ó IMA(d,q). NOTA: Cualquier modelo puede expresarse usando la notación explicada en este tema.

AUTOCORRELACIONES MUESTRALES Y AUTOCORRELACIONES PARCIALES MUESTRALES

Hasta ahora solamente se han considerado las funciones de autocorrelación y autocorrelación parcial TEORICAS que describen un proceso estocástico conceptual. En la práctica se tiene una serie de tiempo finita $z_1, \ldots, z_N$, N observaciones, de la cual sólo se pueden obtener ESTIMADORES DE LAS AUTOCORRELACIONES; por esta razón unicamente se llegan a conocer las autocorrelaciones MUESTRALES y autocorrelaciones parciales MUESTRALES.

El estimador más satisfactorio de la k-ésima autocorrelación es

$$\hat{\rho}_k = r_k = \frac{c_k}{c_0}$$

donde

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z}) \quad , \qquad k = 0, 1, \ldots, K$$

(z_t =desviaciones)

Obsérvese que Ck es el estimador de la autocovarianza y $\bar{z}$ es la media de la serie.

ERROR ESTANDAR DE LA AUTOCORRELACION ESTIMADA

Para identificar un modelo para una serie de tiempo en particular, es necesario tener un buen diagnóstico en cuanto a que ρ_k sea o no cero en algún período. Para este propósito se puede hacer uso de la siguiente expresión, que es una aproximación de la varianza de los coeficientes de autocorrelación estimados, de un proceso estacionario NORMAL dada por Bartlett:

$$\mathrm{var}(r_k) \simeq 1/N \left\{ \rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2 \right\}$$

Para cualquier proceso en el cual las autocorrelaciones son cero (v > q), todos los términos exepto el primero son cero cuando k > q. Así ,para la varianza de las autocorrelaciones estimadas

$$\widehat{\mathrm{var}}(r_k) \simeq 1/N \left\{ 1 + 2\sum_{v=1}^{q} \rho_v^2 \right\} \quad , \quad k > q$$

La raíz cuadrada de esta expresión es el ERROR ESTANDAR de para una muestra grande, puede entonces construirse un estadístico

$$t_{r_k} = \frac{r_k}{S_{r_k}} = \frac{r_k}{(\frac{1}{N}(1+2\sum_{v=1}^{k}\hat{\rho}_v^2))^{1/2}}$$

el cual se distribuye asintóticamente como t de Student con N grados de libertad.

Se ha venido hablando de si las autocorrelaciones son significativas o no, pero no se ha fijado cuando se consideran significativas y cuando no.

Se desea probar	Regla de Decisión
Ho: $\rho_k=0$ H_a: $\rho_k \neq 0$	Aceptar Ho con $\alpha = 0.05$ si $/t_{r_k}/ <= 2$

En lugar de considerar cada una de las autocorrelaciones pueden considerarse las primeras K autocorrelaciones. Box y Pierce mostraron que un proceso ruido blanco con $\rho_k=0, \forall k$, el estadístico

$$Q(K) = n(n+2)\sum_{k=1}^{K}\frac{1}{n-k} r_k^2$$

se distribuye como X_k^2

	Regla de Decisión:
Ho: $\rho_k=0$ k=1,2,....,K H : $\rho_k \neq 0$ al menos para alguna k	Aceptar Ho con α=0.05 si $Q(K) <= X^2_k$.

si se acepta esta hipótesis nula quiere decir que existe ruido blanco, si no habrá que buscar algún otro modelo.

En base a lo anterior se calculan las autocorrelaciones muestrales y se decide si son estadísticamente significativas o no. Pero es necesario recordar que es muy importante el juicio personal de quien esté analizando algún fenómeno.

1.2.3 MODELOS ESTACIONALES

Como se comentó en la introducción de este capítulo, uno de los componentes de una serie de tiempo es la VARIACION ESTACIONAL, ésta puede estar combinada con los otros patrones de una serie (tendencia, ciclos). Cuando una serie es estacional, se refiere a patrones anuales (12 meses), trimestrales (3 meses), etc.. En el caso de trimestrales, por ejemplo, se observarà que cada tres meses se repite un comportamiento.

Los modelos estacionales se analizan de la misma manera que los modelos ya explicados, unicamente tendrà que especificarse el grado de estacionalidad. En la pràctica, no es fàcil detectar el grado de estacionalidad de la serie si se observa unicamente la grafica de los datos originales y menos aûn si existe tendencia. Si, por otro lado, la serie no es estacionaria son necesarias diferencias ordinarias y/ó estacionales ($\nabla_s^D Y_t$).

Los modelos estacionales pueden ser autorregresivos y/ó medias móviles, por lo que al predecir con patrones estacionales estarân incluidos parâmetros estacionales (Φi representa una AR estacional y Θ, representa un MA estacional). El orden el tipo del proceso pueden ser identificados por el exponente del operador de salto hacia atrâs B y los parâmetros respectivamente. Para datos estacionarios, la estacionalidad puede ser encontrada identificando aquellos coeficientes de autocorrelación de mâs de 2 ó 3 periodos que son significativamente distintos de cero (recuerde que cualquier autocorrelaciôn que distinta de cero de manera significativa implica la existencia de un patrôn en los datos).

La idea que maneja Box-Jenkins de ir filtrando la serie original hasta obtener unicamente la variación aleatoria e , es tambiên aplicable en este tipo de modelos. Una vez obtenida la serie estacionaria, se puede ignorar la parte estacional del modelo y buscar un modelo ARMA "tentativo", ésto equivale a un primer filtro entonces se observarâ un comportamiento estacional en las autocorrelaciones de los residuales (que corresponde a lo que faltô considerar en el modelo tentativo). Una vez detectado el grado de estacionalidad y el tipo de modelo , se pasa por un segundo filtro para obtener, por último,un comportamiento de los residuales de ruido blanco.

A) MODELOS AUTORREGRESIVOS ESTACIONALES SAR(P)

Estos modelos se aplican si el valor actual Z_t, puede expresarse como una función lineal del valor de la serie "s" periodos atrâs, Z_{t-s}, y el choque aleatorio e_t. Se denotan por SAR(P) ó AR(P)s.

MODELOS AUTORREGRESIVO ESTACIONALES DE PRIMER ORDEN SAR(1)

SAR(1) se expresa como

$$Z_t = \Phi_1 Z_{t-s} + e_t \qquad ó \qquad (1 - \Phi_1 B^s) Z_t = e_t$$

donde Φ_1 es el parámetro autorregresivo estacional.

La función de autocorrelación del modelo SAR(1) será:

$$\gamma_0 = E(Z_t^2) = E[(\Phi_1 Z_{t-s} + e_t)^2] = E(\Phi_1^2 Z_{t-s}^2 + 2\Phi_1 Z_{t-s} e_t + e_t^2)$$

$$\gamma_0 = \Phi_1 E(Z_{t-s}^2) + E(e_t^2)$$

$$\gamma_0 = \Phi_1 \gamma_0 + \sigma^2$$

$$\gamma_0 = \frac{\sigma^2}{1 - \Phi_1^2}$$

de este resultado se obtiene una condición

$$\Phi_1^2 < 1$$

entonces la CONDICION DE ESTACIONARIDAD es $|\Phi_1| < 1$

$\gamma_1 = E(Z_t Z_{t-1})$ será igual con cero(s=1) ya que el modelo es una función lineal de "s" períodos atrás, por lo que no existe correlación con Z_{t-1}.

$$\gamma_1 = E(Z_t Z_{t-1}) = E[(\Phi_1 Z_{t-s} + e_t)(\Phi_1 Z_{t-s-1} + e_{t-1})]$$

$$= E(\Phi_1^2 Z_{t-s} Z_{t-s-1} + \Phi_1 e_t Z_{t-s-1} + e_t^2 + \Phi_1 Z_{t-s} e_{t-1})$$

$$\gamma_1 = \Phi_1^2 \gamma_1 \qquad , \qquad \underline{\gamma_1 = 0}$$

Ahora el coeficiente de autocorrelación ρ_s será

$$\gamma_s = E(Z_t Z_{t-s}) = E[(\Phi_1 Z_{t-s} + e_t)(Z_{t-s})]$$

$$\gamma_s = \Phi_1 \gamma_0 \qquad\qquad \underline{\rho_s = \frac{\gamma_s}{\gamma_0} = \Phi_1}$$

Si se comparan estos resultados con los del modelo AR(1), se observa que son semejantes

AR(1)	SAR(1)
$Z_t = \phi_1 Z_{t-1} + e_t$	$Z_t = \Phi_1 Z_{t-s} + e_t$
$\rho_1 = \phi_1$	$\rho_s = \Phi_1$

De ésto se deduce que la función de autocorrelación se comporta igual que en el modelo AR(1), unicamente que los valores distintos de cero serán los múltiplos de "s".

Y en la función de autocorrelación parcial, el valor significativo estará en "s", lo cual es de gran ayuda para identificar el orden y el tipo del modelo. Las figuras 1.20 y 1.21 ilustran el comportamiento de la ACF y PACF de estos modelos.

MODELOS AUTORREGRESIVOS ESTACIONALES DE ORDEN MAYOR SAR(P)

SAR(P) se expresa como:

$$Z_t = \Phi_1 Z_{t-s} + \Phi_2 Z_{t-2s} + \ldots + \Phi_P Z_{t-Ps} + e_t$$

D= orden de las diferencias estacionales, d= orden de las diferencias ordinarias, s= longitud del ciclo estacional

Por lo obtenido en SAR(1) con respecto a la comparación con AR(1), sucede exactamente lo mismo con los modelos de orden mayor. Tomando SAR(2), por ejemplo, las restricciones para los parámetros serán

$$|\Phi_2| < 1$$
$$\Phi_1 + \Phi_2 < 1$$

En general la función de autocorrelación irá decayendo a cero con valores significativos en "s" y múltiplos de "s" y la fn. de autocorrelación parcial tendrá valores significativos en s,2s,....,Ps.

EJEMPLO:

SAR(1) $Z_t = 0.8 Z_{t-12} + e_t$

s=12

ACF:

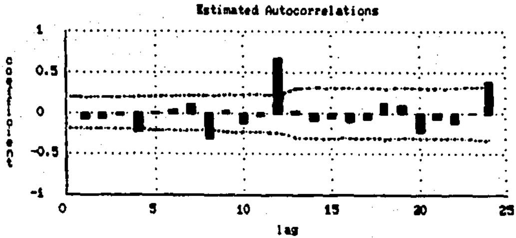


figura 1.20
ACF SAR(1)

PACF:

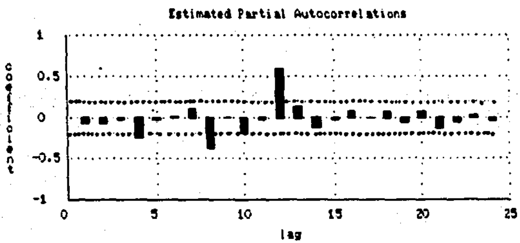


figura 1.21
PACF SAR(1)

B) MODELOS DE MEDIAS MOVILES ESTACIONALES SMA(Q)

Este modelo se usa si el valor de Z_t puede representarse como el choque aleatorio e_t y el choque ocurrido exactamente "s" intervalos atrás (e_{t-s}). Se denota como SMA(Q) ó MA(Q)s.

MODELOS ESTACIONALES DE MEDIAS MOVILES DE PRIMER ORDEN SMA(1)

SMA(1) se expresa como:

$$Z_t = e_t - \Theta_1 e_{t-s}$$

donde Θ_1 es el parámetro de medias móviles estacional.

Al igual que en los modelos de medias móviles no estacionales, existen condiciones de invertibilidad, pero ahora el parámetro es estacional.

MA(1)	SMA(1)
$Z_t = e_t - \theta_1 e_{t-1}$,	$Z_t = e_t - \Theta_1 e_{t-s}$
$/ \theta_1 / < 1$	$/ \Theta_1 / < 1$

Para los coeficientes y la función de autocorrelación:

$$\gamma_0 = Var(Z_t) = E(Z_t^2) = E[(e_t - \Theta_1 e_{t-s})^2] = E(e_t^2 - 2\Theta_1 e_t e_{t-s} + \Theta_1^2 e_{t-s}^2)$$

$$\gamma_0 = \sigma^2 + \Theta_1^2 \sigma^2$$

$$\gamma_0 = \sigma^2(1 + \Theta_1^2)$$

$$\gamma_1 = E(Z_t Z_{t-1}) = E(e_t e_{t-1} - \Theta_1 e_t e_{t-s-1} - \Theta_1 e_{t-1} e_{t-s} + \Theta_1^2 e_{t-s} e_{t-s-1})$$

$$\gamma_1 = 0 \quad \Rightarrow \quad \rho_1 = 0$$

$$\gamma_s = E(Z_t Z_{t-s}) = E(e_t e_{t-s} - \Theta_1 e_t e_{t-2s} - \Theta_1 e_{t-s}^2 + \Theta_1^2 e_{t-s} e_{t-2s})$$

$$\gamma_s = -\Theta_1 \sigma^2 \quad \Rightarrow \quad \rho_s = \frac{\gamma_s}{\gamma_0} = \frac{-\Theta_1 \sigma^2}{\sigma^2(1+\Theta_1^2)} = \frac{-\Theta_1}{(1+\Theta_1^2)}$$

$$\rho_s = \begin{cases} \dfrac{-\Theta_1}{(1+\Theta_1^2)} , & k = s \\ 0 , & k \neq s \end{cases}$$

Comparando con la fn de autocorrelación de MA(1), unicamente varía en que el valor significativo se encuentra en "s". Lo mismo se puede decir de la función de autocorrelación parcial (pacf) la cual irá disminuyendo a cero, con valores significativos en múltiplos de "s". Las figuras 1.22 y 1.23 muestran el comportamiento de la ACF y la PACF de estos modelos.

EJEMPLO:

SMA(1) $Zt = e_t + 0.8e_{t-s}$

s=12

ACF:

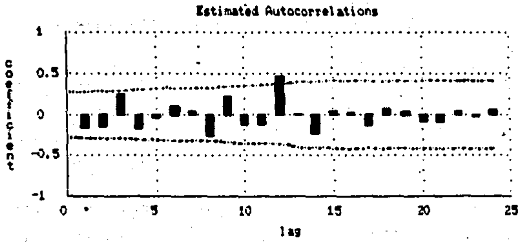


figura 1.22
ACF SMA(1)

PACF:

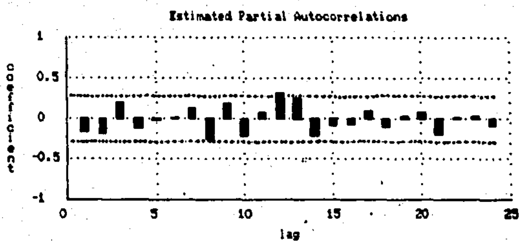


figura 1.23
PACF SMA(1)

MODELOS ESTACIONALES DE MEDIAS MOVILES DE ORDEN MAYOR SMA(Q)

SMA(Q) se expresa como

$$Z_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - \dots - \Theta_Q e_{t-Qs}$$

Conforme se va aumentando el orden de estos modelos, todos los càlculos se van complicando pero es de gran ayuda saber que su comportamiento se asemeja al modelo MA(q), q > 1, variando unicamente en la longitud del ciclo estacional. Por lo anterior, las condiciones de invertibilidad para un SMA(2) serân:

MA(2)	SMA(2)
$Z_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$	$Z_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s}$
$\theta_1 + \theta_2 < 1$	$\Theta_1 + \Theta_2 < 1$
$\theta_2 - \theta_1 < 1$	$\Theta_2 - \Theta_1 < 1$
$/\theta_2/ < 1$	$/\Theta_2/ < 1$

Los coeficientes de la función de autocorrelación son

$$\gamma_0 = E(Z_t^2) = E[(e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s})(e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s})]$$

$$= E(e_t^2 - \Theta_1 e_t e_{t-s} - \Theta_2 e_t e_{t-2s} - \Theta_1 e_t e_{t-s} + \Theta_1^2 e_{t-s} + \Theta_1 \Theta_2 e_{t-s} e_{t-2s}$$

$$- \Theta_2 e_t e_{t-2s} + \Theta_1 \Theta_2 e_{t-s} e_{t-2s} + \Theta_2^2 e_{t-2s}^2)$$

$$\gamma_0 = \sigma^2 + \Theta_2^2 \sigma^2$$

$$\gamma_0 = \sigma^2 (1 + \Theta_2^2)$$

$$\gamma_1 = E(Z_t Z_{t-1}) = E[(e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s})(e_{t-1} - \Theta_1 e_{t-s-1} - \Theta_2 e_{t-2s-1})]$$

$$\gamma_1 = 0$$

$$\gamma_s = E(Z_t Z_{t-s}) = E[(e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s})(e_{t-s} - \Theta_1 e_{t-2s} - \Theta_2 e_{t-3s})]$$

$$= E(e_t e_{t-s} - \Theta_1 e_t e_{t-2s} - \Theta_2 e_t e_{t-3s} - \Theta_1 e_{t-s}^2 + \Theta_1^2 e_{t-s} e_{t-2s} + \Theta_1 \Theta_2 e_{t-s} e_{t-3s}$$

$$- \Theta_2 e_{t-s} e_{t-2s} + \Theta_1 \Theta_2 e_{t-2s}^2 + \Theta_2^2 e_{t-s} e_{t-3s})$$

$$\gamma_s = -\Theta_1 \sigma^2 + \Theta_1 \Theta_2 \sigma^2$$

$$\gamma_s = \Theta_1 \sigma^2 (\Theta_2 + 1) \quad , \quad \rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\Theta_1 (\Theta_2 + 1)}{(1 + \Theta_2^2)}$$

La función de autocorrelación se trunca despuès de Qs y la de autocorrelación parcial decrece con valores en múltiplos de "s".

C) MODELOS ESTACIONALES MEZCLADOS

Este tipo de modelos tendrà tèrminos autorregresivos estacionales y de medias mòviles estacionales. Se conocen como modelos SARMA(P,Q) ó ARMA(P,Q)s. El modelo SARMA(P,Q) se expresa como

$$Z_t = \Phi_1 Z_{t-s} + \Phi_2 Z_{t-2s} + \dots + \Phi_P Z_{t-Ps} + e_t - \Theta_1 e_{t-s} - \dots - \Theta_Q e_{t-Qs}$$

ó

$$\Phi(B^s) Z_t = \Theta(B^s) e_t$$

donde

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}$$

Las caracterìsticas y restricciones para esta clase de modelos varìan en "s" con respecto a los modelos ARMA(p,q). Entonces las funciones de autocorrelación y autocorrelación parcial decaen, las dos, a cero en forma exponencial teniendo valores significativos en intervalos mùltiplos de "s". Esto se puede observar en las figuras 1.24 y 1.25 del siguiente ejemplo.

EJEMPLO:

SARMA(2,1) $Z_t = -0.8Z_{t-s} - 0.6Z_{t-2s} + e_t + 0.8e_{t-s}$

$s=12$

ACF:

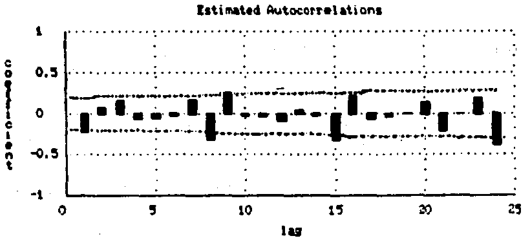


figura 1.24
ACF SARMA(2,1)

PACF:

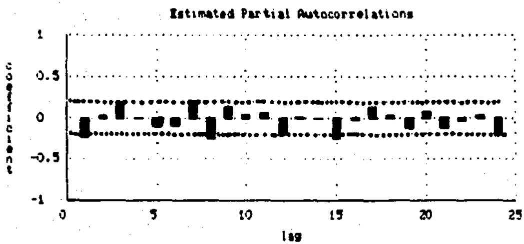


figura 1.25
PACF SARMA(2,1)

1.2.4 MODELOS GENERALES MULTIPLICATIVOS ESTACIONALES

Se puede decir que esta clase de modelos es la más complicada ya debido a que son muchas las combinaciones que se pueden tener, estos modelos constan de 2 partes:

1) la parte regular ordinaria no estacional

2) algunos parámetros estacionales

De ésto se desprenden algunos casos:

AR(p) x SAR(P)
MA(q) x SMA(Q)
ARMA(p,q) x SARMA(P,Q)

La forma general de estos modelos es ARIMA(p,d,q) x (P,D,Q)s y se expresan como

$$\phi(B)\,\Phi(B^s)\,z_t = \theta(B)\,\Theta(B^s)\,e_t$$
$$Z_t = \nabla^d \nabla_s^D y_t$$

A continuación se presentan algunos casos particulares de manera detallada.

---> ARIMA (1,0,0) x (1,0,0)s

Sea

$$\phi(B) = (1 - \phi_1 B)$$
$$\Phi(B) = (1 - \Phi_1 B)$$

este modelo se expresa como

$$(1 - \phi_1 B)(1 - \Phi_1 B^s) Z_t = e_t$$

desarrollando el lado izquierdo de la igualdad:

$$(1 - \Phi_1 B^s - \phi_1 B + \phi_1 B \Phi_1 B^s) Z_t = e_t$$
$$(1 - \Phi_1 B^s - \phi_1 B + \phi_1 \Phi_1 B^{s+1}) Z_t = e_t$$
$$Z_t - \Phi_1 Z_{t-s} - \phi_1 Z_{t-1} + \phi_1 \Phi_1 Z_{t-s-1} = e_t$$

$$Z_t = \underbrace{\phi_1 Z_{t-1}}_{\text{TERMINO AR, NO ESTACIONAL}} - \underbrace{\Phi_1 Z_{t-s}}_{\text{TERMINO AR, ESTACIONAL}} + \underbrace{\phi_1 \Phi_1 Z_{t-s-1}}_{\text{COMO EL MODELO ES AUTORREGRESIVO, } Z_t \text{ ESTA EN TERMINOS DE } Z_{t-s} \text{ Y DE UNO ANTERIOR, } Z_{t-s-1}} + e_t$$

Este modelo podría tomarse como un AR(s+1), con los coeficientes intermedios ($\phi_1, \phi_2, \ldots, \phi_{s-1}$) iguales con cero lo cual indica que AR(s+1) NO es el modelo adecuado. Si se modelara en forma tentativa con SAR(1)

$$Z_t = \Phi_1 Z_{t-s} + a_t \quad , \quad (1 - \Phi_1 B^s) Z_t = a_t \quad \ldots \quad (1)$$

se encontrará que a_t se comporta como un modelo autorregresivo AR(1)

$$(1 - \phi_1 B) a_t = e_t \leftarrow \text{RUIDO BLANCO}$$

que es precisamente lo que faltó considerar al modelar.

Multiplicando (1) por ($1-\phi_1 B$)

$$(1 - \phi_1 B)(1 - \Phi_1 B^s) Z_t = \underbrace{(1 - \phi_1 B) a_t}_{e_t}$$

y se tiene como resultado el modelo multiplicativo

$$(1 - \phi_1 B)(1 - \Phi_1 B^s) Z_t = e_t$$

La funciones de autocorrelación autocorrelación parcial se comportan como sigue:

ACF:
decrece con picos en múltiplos de "s".

PACF:
se trunca después de s+1

Este comportamiento se observa en las figuras 1.26 y 1.27

EJEMPLOS:

AR(1) x SAR(1)

span=12

$Zt = 0.8Z_{t-1} - 0.8Z_{t-s} + e_t$

ACF:

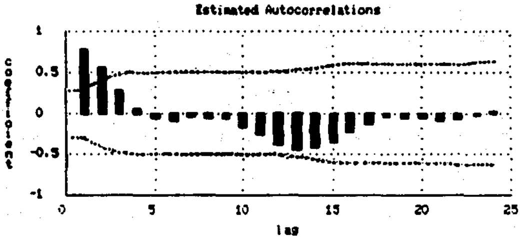


figura 1.26
ACF AR(1)xSAR(1)

PACF:

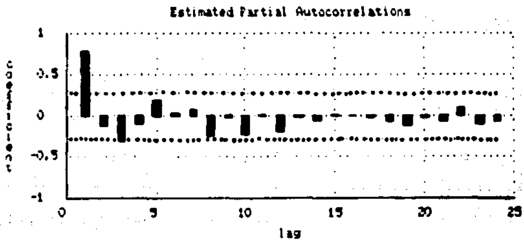


figura 1.27
PACF AR(1)xSAR(1)

---> ARIMA(0,0,1) x (0,0,1)s

$$\theta(B) = (1-\theta_1 B)$$
$$\Theta(B^s) = (1-\Theta_1 B^s)$$

Este modelo se expresa como

$$Z_t = (1-\theta_1 B)(1-\Theta_1 B^s)e_t$$
$$Z_t = \theta(B)\Theta(B^s)e_t$$

desarrollando el lado derecho de la igualdad

$$Z_t = (1-\Theta_1 B^s-\theta_1 B+\theta_1\Theta_1 B^{s+1})e_t$$

$$Z_t = 1-\underbrace{\theta_1 e_{t-1}}_{\text{TERMINO MA NO ESTACIONAL}}-\underbrace{\Theta_1 e_{t-s}}_{\text{TERMINO MA ESTACIONAL}}+\theta_1\Theta_1 e_{t-s-1}$$

haciendo lo mismo que en el modelo anterior, se podría tomar a éste último modelo como un MA(s+1) con $\theta_2, \theta_3, \ldots. \theta_{s-1}$ no significativos lo cual indica que MA(s+1) NO es el modelo adecuado.

Las funciones de autocorrelación y autocorrelación parcial de este modelo se comportan de la siguiente manera:

ACF:
Se trunca después de s+1, con picos en 1,s-1,s,s+1
PACF:
Decrece con valores significativos en s-1,2s-1,...

Obsérvese el siguiente ejemplo, las figuras 1.28 y 1.29 ilustran el comportamiento de las funciones de autocorrelación y autocorrelación parcial.

EJEMPLO:

MA(1) x SMA(1)
span=12

ACF:

$$Zt = e_t - 0.8\, e_{t-1} + 0.8\, e_{t-s}$$

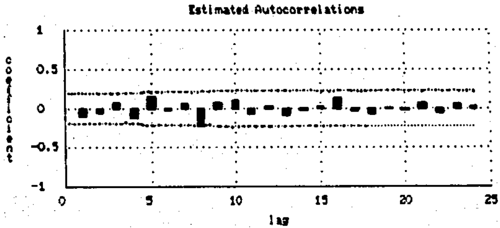


figura 1.28
ACF MA(1)xSMA(1)

PACF:

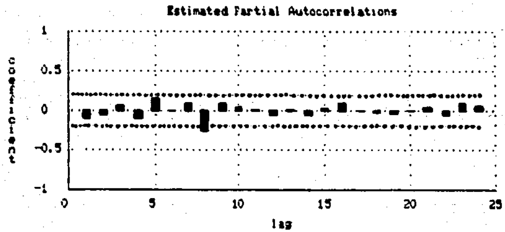


figura I.29
PACF MA(1)xSMA(1)

Generalizando las ACF's y PACF's para los modelos multiplicativos el comportamiento de las funciones se resume en la siguiente tabla.

MODELO	A C F	P A C F
AR(p) X SAR(P)	decrece con picos en múltiplos de s	se trunca después de p+Ps
MA(q) x SMA(Q)	se trunca después de q+Qs	decrece con picos en múltiplos de s
MEZCLADOS	decrece con picos en múltiplos de s	decrece con picos en múltiplos de s

tabla B

1.3 ESTIMACION DE MODELOS

Una vez que se ha identificado un modelo tentativo de la forma

$$\phi(B)W = \theta(B)e_t \quad , , , \quad (1)$$

es decir, una vez establecido p,d,q ; se requiere estimar los p+q coeficientes $\phi_1, \phi_2, \ldots, \phi_p$ y $\theta_1, \theta_2, \ldots \theta_q$. Estos deberán minimizar la suma de cuadrados de los residuales (1), los cuales serán los ESTIMADORES de los errores e ,que están dados por

$$e_t = \theta(B)^{-1}\phi(B)\,\omega_t \quad , , , \quad (2)$$

$$\hat{e}_t = \hat{\theta}(B)^{-1}\hat{\phi}(B)\,\omega_t$$

Por otro lado sean
$\phi = (\phi_1, \phi_2, \ldots, \phi_p)$ el vector de parámetros autorregresivos y
$\theta = (\theta_1, \theta_2, \ldots, \theta_q)$ el vector de parámetros de medias móviles

que minimicen la suma de cuadrados:

$$S(\phi,\theta) = \sum_{t=1}^{N} e_t^2 \quad , , , \quad (3)$$

cuyos estimadores son los vectores $\hat{\phi}$ y $\hat{\theta}$, entonces la suma de cuadrados a minimizar queda como

$$S(\hat{\phi},\hat{\theta}) = \sum_{t=1}^{N} \hat{e}_t^2 .$$

A) ESTIMACION DE LOS PARAMETROS DE UN MODELO AUTORREGRESIVO

Asumiendo que la muestra de los datos que se tiene es generada por un proceso autorregresivo y que se conoce el orden p del modelo, se necesitan estimar los valores de los parámetros $\phi_1, \ldots, \phi_p$ que describan este proceso.

(1) los residuales son la diferencia entre el valor observado y el timado $(Z_i - \hat{Z}_i)$.

La ecuación

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{t-p} + e_t$$

tiene la forma lineal de un modelo estadístico de regresión múltiple, los regresores $W_{t-1}, \ldots, W_{t-p}$ son variables estocásticas y el golpe aleatorio e_t es añadido al sistema en el tiempo t y es independiente de las variables aleatorias en puntos previos en el tiempo (los regresores son independientes del término error). Así, se puede estimar ϕ por el método de mínimos cuadrados.

Tómese el caso más sencillo, el AR(1)

$$w_t = \phi_1 w_{t-1} + e_t$$
$$e_t = w_t - \phi_1 w_{t-1}$$
$$e_t = (1 - \phi_1 B) w_t$$

se desea estimar ϕ_1 de modo que la suma sea mínima, por lo que el procedimiento es el siguiente

$$S(\phi) = \sum_{t=1}^{N} e_t^2 = \sum_{t=1}^{N} (w_t - \phi_1 w_{t-1})^2$$

$$\frac{dS(\phi)}{d\phi_1} = -2 \sum_{t=1}^{N} (w_t - \phi_1 w_{t-1}) w_{t-1} = 0$$
$$= -2 \sum_{t=1}^{N} w_t w_{t-1} - \phi_1 w_{t-1}^2 = 0$$

$$\sum_{t=1}^{N} w_t w_{t-1} - \phi_1 \sum_{t=1}^{N} w_{t-1}^2 = 0$$

$$-\phi_1 \sum_{t=1}^{N} w_{t-1}^2 = -\sum_{t=1}^{N} w_t w_{t-1}$$

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{N} w_t w_{t-1}}{\sum_{t=1}^{N} w_{t-2}^2} = \hat{\rho}_1$$

Se tienen N observaciones disponibles, $W_1, W_2, \ldots, W_N$, por lo que cuando t=1 se va a requerir de un valor desconocido W_0.

B) ESTIMACION DE LOS PARAMETROS DE UN MODELO DE MEDIAS MOVILES

Ahora se asume que la muestra de datos que se tiene genera un proceso de medias mòviles de orden q, y que la media de la muestra es cero (para efectos pràcticos). La minimización de la suma de los cuadrados de los residuales queda representada como

$$S(\theta)=\sum_{t=1}^{N} e_t^2=\sum_{t=1}^{N}(\omega_t-\theta_1 e_{t-1}-\dots-\theta_q e_{t-q})^2 \quad , \; , \; , \quad (1)$$

y provee estimadores asintòticamente eficientes si e ~ N . Los valores de e_t son inobservables por lo que tendràn que ser reemplazados por los observados, es decir los residuales.

Considèrese el proceso MA(1)

$$\omega_t=e_t-\theta_1 e_{t-1} \quad , \; , \; , \quad (2)$$

con / θ_1 / < 1

El procedimiento para minimizar la suma de cuadrados de los residuales es el siguiente

$$S(\theta) = \sum_{t=1}^{N} e_t^2 = \sum_{t=1}^{N}(\omega_t+\theta_1 e_{t-1})^2$$
$$= \sum_{t=1}^{N}\left[\omega_t-(-\theta_1 e_{t-1})\right]^2 \rightarrow \text{NO SE CONOCE } e_{t-1} \text{ HABRÁ QUE ESTIMARLO}$$

$$\omega_t = e_t - \theta_1 e_{t-1}$$
$$e_t = \omega_t + \theta_1 e_{t-1}$$
$$e_{t-1} = \omega_{t-1} + \theta_1 e_{t-2}$$

$$S(\theta) = \sum_{t=1}^{N}\left[\omega_t-(-\theta_1\omega_{t-1}-\theta_1^2 e_{t-2})\right]^2$$
$$\vdots$$
$$S(\theta)=\sum_{t=1}^{N}\left[\omega_t+\theta_1\omega_{t-1}+\theta_1^2\omega_{t-2}+\dots+\theta_1^N\omega_{t-N}\right]^2$$

$$\frac{dS(\theta)}{d\theta_1} = 2\sum_{t=1}^{N}(\omega_t+\theta_1\omega_{t-1}+\dots+\theta_1^N\omega_{t-N})(\omega_{t-1}+2\theta_1\omega_{t-2}+\dots+N\theta_1^{N-1}\omega_{t-N})$$

Se observa que además de requerir un valor desconocido "e_0", la ecuación NO ES LINEAL, lo cual se generaliza para procesos de orden superior. Por otro lado se requieren, para t=1 , $W_{t-1}, \dots, W_{t-w}$ desconocidos.

Para este caso, que es el más simple, con un solo parámetro que se sabe que su valor estará entre 1 y -1, se pueden probar diferentes valores de θ_1, dentro de este rango, calcular $S(\theta)$ y escoger el valor de θ_1^* que de como resultado la mínima suma de cuadrados de los residuales; ésto también se puede hacer con algún método de búsqueda. Lo anterior es un procedimiento experimental, y entonces la dificultad es mayor aplicando los modelos ARMA , ya que el procedimiento experimental es iterativo y por lo tanto los cálculos computacionales son considerables, conforme más grande sea el orden del modelo éstos se incrementan ;por lo que si el número de parámetros excede a 2, es preferible utilizar un algoritmo numérico de minimización.En nuestros días. existen paquetes de estimación para modelos ARMA (por ejemplo, IMSL,FORECAST PLUS. STAT GRAPHIC,SPSS), disponibles en varios centros de cómputo lo cual simplifica considerablemente el problema de estimación.

De todo lo anterior se observa que:

1. Van a ser necesarios valores que se desconocen, como por ejemplo "W_0" y "e_0".

2. La estimación se complica si el modelo incluye términos de medias móviles ya que la minimización de $S(\phi,\theta)$ requiere de métodos de optimización no lineales.

1.3.1 EL PROCEDIMIENTO DE ESTIMACION

Considérese el problema de estimación del modelo ARMA(p,q), dados $W_1, \dots, W_w$. Si $E(W_t) \neq 0$, W_t es reemplazado por $\tilde{W}_t = W_t - E(W_t)$ y para series moderadamente largas W_t puede ser substituida por $E(W_t)$.

Los choques aleatorios estimados están dados por

$$e_t = w_t - \phi_1 w_{t-1} - \dots - \phi_p w_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q}$$

en donde los et's necesitan ser calculados. Para obtener e_1, se requiere conocer $e_0, e_{-1}, \dots, e_{1-q}$ y $W_0, W_{-1}, \dots, W_{1-p}$, (θ^*, ω^*).

lo cual no está disponible; por lo que se hacen necesarios los VALORES INICIALES de e^* y W^* .

Se proponen dos posibilidades:

1. Tomar W_0, W_{-1}, ..., e_0, e_{-1} .. iguales a sus VALORES ESPERADOS INCONDICIONALES. Los valores esperados incondicionales de e_0, e_{-1} ...e_{-q} son todos cero. Si el modelo carece de término independiente (las Wt son desviaciones con respecto a la media), los valores esperados incondicionales de W_0, W_{-1}, ... serán también cero. Esto da una aproximación mas o menos correcta si los valores reales de ϕ_1 ....., ϕ_p no son cercanos a 1, y si el número de observaciones es grande con respecto a p y q.

2. Existe un procedimiento para encontrar los valores ESPERADOS CONDICIONALES para W_0, W_{-1}, ... estos valores son condicionales con respecto a los valores observados W_1, W_2, ...W_N y los valores estimados a_1, ...a_T . Esencialmente, pueden inicializarse W_0, W_{-1}... y a_0, a_{-1} ... con ceros, estimar el modelo minimizando $S(\phi,\theta)$ condicionada a estos valores cero, y entonces PRONOSTICAR

HACIA ATRAS (back forecasting) con el modelo para generar [illegible] para W_0, W_{-1}, ... (solamente se pueden generar valores para [illegible] y no para los at's ya que at se estima del valor observado y [illegible] el valor dado por el modelo).

Como la serie W_0, W_{-1}, ...W_N es estacionaria el proceso ARMA puede regresarse en el tiempo. Sea F el operador de "salto hacia adelante" (forward shift operator):

$$FW_t = W_{t+1}$$

Recuérdese que

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots$$

ahora

$$\phi(F) = 1 - \phi_1 F - \phi_2 F - \dots$$

Entonces un proceso ARMA puede expresarse como

$$\phi(F)W_t = \theta(F)e_t \quad \dots (3).$$

EJEMPLO:

ARMA(2,1) $(1-\phi_1 F-\phi_2 F^2)\omega_t = (1-\theta_1 F)e_t$

$$\omega_t-\phi_1\omega_{t+1}-\phi_2\omega_{t+2} = e_t-\theta_1 e_{t+1}$$

$$\omega_t = \phi_1\omega_{t+1}+\phi_2\omega_{t+2}+e_t-\theta_1 e_{t+1}$$

Obsérvese que pueden encontrarse los valores de $W_0, W_{-1}, \ldots$

$$\omega_0 = \phi_1 \underset{\text{OBSERVADO}}{\omega_1} + \phi_2 \underset{\text{OBSERVADO}}{\omega_2} + \underset{\substack{\text{SE USA EL}\\ \text{VALOR ESPERADO}\\ \text{CONDICIONAL}}}{e_0} - \theta_1 \underset{\text{ESTIMADO}}{e_1} \; ; \; , \; , \qquad (3)$$

$$\omega_{-1} = \phi_1 \underset{\substack{\text{ESTIMADO}\\ \text{EN (3)}}}{\omega_0} + \phi_2 \underset{\text{OBSERVADO}}{\omega_1} + \underbrace{e_{-1} - \theta_1 e_0}_{\substack{\text{VALORES}\\ \text{INCONDICIONALES}}}$$

Ahora, a partir de (3)

$$\omega_t = \phi(F)^{-1}\theta(F)\, e_t \qquad (4)$$

Esta ecuación puede usarse para encontrar valores ESTIMADOS de W_0, W_{-1} ....en términos de los valores ESTIMADOS de e_1, e_2, ...e_n. Estos últimos son los residuales. Ahora pueden obtenerse nuevos estimadores de mínimos cuadrados para ϕ y θ minimizando $S(\phi,\theta)$ CONDICIONADA a $\hat{W}_0, \hat{W}_{-1}, \ldots$. Se repite el proceso hasta que los estimadores sean convergentes (hasta que $\hat{\phi}$ y $\hat{\theta}$ dejen de cambiar de forma significativa). Desafortunadamente no hay garantía de que el proceso sea convergente, si no lo es se toman los estimadores iniciales, i.e. los incondicionales.

1.3.2 ESTIMACION NO LINEAL DE LOS PARAMETROS DEL MODELO.

Recuerdese que el problema es encontrar valores de $\emptyset$ y θ que minimicen

$$S(\emptyset,\theta)=\sum_{t=1}^{N} e_t^2=\sum_{t}^{N}[e_t/\emptyset,\theta,\omega]^2 \quad . \; . \; . \quad (b)$$

donde W = vector que contiene los valores de Wt.

Sea β el vector que contiene los p+q parametros de $\emptyset$ y θ, entonces la minimización con respecto a de la suma de cuadrados dada por

$$\sum_{t=1}^{N}\{f_t(\beta)\}^2$$

se simplifica considerablemente si cada $f_t(\beta)$, t=0,1,....,N , es una función lineal de los parámetros de . Esto difiere del modelo puramente autorregresivo al modelo de medias móviles.

Un proceso puramente autorregresivo tiene la forma de un modelo de regresión lineal múltiple (2), entonces

$$\emptyset(B)\omega_t=e_t$$

ó

$$\omega_t=\phi_1\omega_{t-1}+\phi_2\omega_{t-2}+\dots+\phi_p\omega_{t-p}+e_t$$

$$\hat{\emptyset}=(\hat{\phi}_1,\hat{\phi}_2,\hat{\phi}_3,\dots,\hat{\phi}_p)^T=(X^TX)^{-1}X^TY$$

donde

$$X=\begin{bmatrix}\omega_p & \omega_{p-1} & \dots & \omega_1\\ \omega_{p+1} & \omega_p & \dots & \omega_2\\ \vdots & & & \\ \omega_{T-1} & \omega_{T-2} & \dots & \omega_{T-p}\end{bmatrix}, \quad Y=\begin{bmatrix}\omega_{p+1}\\ \omega_{p+2}\\ \vdots\\ \omega_T\end{bmatrix}_{(T-p)}$$

$$\omega_{p+1}=\phi_1\omega_p+\phi_2\omega_{p-1}+\dots+\phi_p\omega_1$$

si el modelo incluye medias móviles no puede resolverse de la forma anterior. En ese caso

(2) Referencia (5)

$$\phi(B)\,\omega_t = \theta(B)\,e_t$$

$$e_t = \phi(B)\,\theta^{-1}(B)\,e_t$$

Esta última ecuación es no lineal, pero puede transformarse en una ecuación lineal aproximada utilizando la expansión en serie de Taylor y truncando.

N O T A :

SERIE DE TAYLOR:

$$f(x) = f(x_0) + \frac{(x-x_0)}{1!} f'(x_0) + \frac{(x-x_0)^2}{2!} f''(x_0) + \dots$$

$$f(x_1, x_2) = f(x_{01}, x_{02}) + \frac{(x_1 - x_{01})}{1!} \frac{\partial f(x_1, x_2)}{\partial x_1}\Big|_{\substack{x_1 = x_{01} \\ x_2 = x_{02}}} + (x_2 - x_{02}) \frac{\partial f(x_1, x_2)}{\partial x_2} + \dots$$

X_{01} → VALOR CERCANO A x_1, X_{10} → VALOR INICIAL DE X_1

X_{02} → VALOR CERCANO A x_2 X_{20} → VALOR INICIAL DE X_2

Para observar el proceso iterativo se utiliza el vector β definido anteriormente. Entonces se desea encontrar los valores de que minimicen

$$S(\beta) = \sum_{t=1}^{T} [e_t / \omega, \beta]^2 = \sum_{t=1}^{T} [e_t]^2$$

Para expander esta ecuación en serie de Taylor, sea β_0 una aproximación de β, i.e. una aproximación inicial de los parámetros

$$[e_t] \simeq \underbrace{[e_t / \omega, \beta_0]}_{\text{RESIDUALES}} + \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0}) \frac{\partial [e_t]}{\partial \beta_i}\Big|_{\beta = \beta_0} + \dots$$

SE TRUNCA LA SERIE

(NO SE PUEDEN SACAR RESIDUALES SIN UNA 1ª APROX.)

sean

$$X_{i,t} = \frac{-\partial [e_t]}{\partial \beta_i}\Big|_{\beta = \beta_0}, \quad \underbrace{[e_{t,0}]}_{\text{RESIDUALES CUANDO SE TOMA LA PRIMERA APROX.}} = [e_t / \omega, \beta_0]$$

entonces

$$[e_t] \approx [e_{t,0}] - \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0})\, x_{i,t}$$

$$\approx [e_{t,0}] - \sum_{i=1}^{p+q} \beta_i x_{i,t} + \beta_{i,0} x_{i,t}$$

$$[e_t] + \sum_{i=1}^{p+q} \beta_i\, x_{i,t} = [e_{t,0}] + \sum_{i=1}^{p+q} \beta_{i,0}\, x_{i,t}$$

reescribiendo

$$\underbrace{[e_{t,0}] + \sum_{i=1}^{p+q} \beta_{i,0} x_{i,t}}_{\text{VARIABLE DEPENDIENTE "COMPUESTA"}} \approx \sum_{i=1}^{p+q} \beta_i\; x_{i,t} + [e_t]$$

El lado derecho de la ecuación será

$$\beta_1 X_{1,t} + \beta_2 X_{2,t} + \ldots + \beta_{p+q} X_{p+q,t} + e_t \; . . . \quad (C)$$

donde las Xi't's corresponden a las variables independientes del modelo de regresión lineal múltiple (recuérdese que en realidad es una misma variable). Sea

$$y_{i,t} \approx [e_{t,0}] + \sum_{i=1}^{p+q} \beta_{i,0}\; x_{i,t}$$

entonces

$$y_{i,t} \approx \sum_{i=1}^{p+q} \beta_i x_{i,t} + [e_t]$$

donde t = 1,2,...,T (queda un sistema con T ecuaciones).

Observe que (5) es semejante a un modelo de regresión lineal múltiple. Con este modelo se calculan los nuevos valores de , que serán . Ahora se utiliza como aproximación para la expansión de Taylor, se linealizan y se encuentran nuevos valores para ; así sucesivamente, hasta que 2 estimaciones sucesivas sean practicamente iguales

$$\beta_k - \beta_{k-1} \simeq 0$$

ó

$/ \beta_k - \beta_{k-1} / <= \epsilon$, ϵ=cierto error de tolerancia designado de acuerdo a cada problema a "k" se le llama NUMERO DE CONVERGENCIA. Al final queda el modelo

$$y_{i,t} \approx \sum_{i=1}^{p} \beta_i x_{i,t} + [e_t]$$

con

$$x_{i,t} = -\frac{\partial [e_t]}{\partial \beta_i} \Bigg|_{\beta=\beta_k}$$

$[e_{i,t}]$, indica que van cambiando los residuales

$\beta_{i,k}$, la k'ésima observación

Este modelo no es el modelo exacto ARMA(p,q), sino que es una linealización del modelo estimado $\phi(B)W_t = \theta(B)e$.

Los errores estandar y los estadísticos se calculan a partir de la última linealización ya que es un modelo de regresión lineal, pero no tienen la misma relevancia (3).

1.3.3 OBTENCION DE UN VALOR INICIAL PARA LOS PARAMETROS -

Sea el proceso AR(p), cuya función de autocorrelación está dada por

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}$$

por lo que se generan "p" ecuaciones lineales simultâneas

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

llamadas ECUACIONES DE YULE-WAKER.

En el caso de este modelo el sistema puede resolverse sin necesidad de valores iniciales. Tómese ahora el caso del proceso de

(3) NOTA: Antes de tomar decisiones en base a los errores estandar y los estadísticos debe usarse el criterio.

medias móviles MA(1):

$$\rho_1 = -\frac{\theta_1}{(1+\theta_1^2)}$$

$$\rho_1(1+\theta_1^2) = -\theta_1$$

$$\rho_1 + \theta_1^2\rho_1 + \theta_1 = 0$$

$$\theta_1 = \frac{-1 \pm \sqrt{1-4\rho_1^2}}{2\rho_1} \quad ,$$

$$1 - 4\rho_1^2 \geq 0$$

$$\rho_1^2 \leq \frac{1}{4}$$

$\theta_1 \rightarrow$ PUEDE SER IMAGINARIO

$$-\frac{1}{2} \leq \rho_1 \leq \frac{1}{2}$$

LO CUAL NO SIEMPRE SUCEDE

Si la solución es real, tenemos dos soluciones para θ_1, y solamente una puede ser posible:

Por ejemplo

Si $\hat{\rho}_1 = 0.4$

$$\theta_1 = \frac{-1 \pm \sqrt{1-.64}}{.8} = \frac{-1 \pm 0.6}{0.8} = \begin{cases} -0.5 \\ -2 \end{cases}$$

de estas dos soluciones, se concluye que θ_1 = -0.05 ya que cumple con la condición de invertibilidad del modelo MA(1) ($/ \theta_1 / < 1$).

Si el proceso de orden mayor se obtiene un conjunto de ecuaciones simultáneas NO LINEALES y para encontrar las aproximaciones iniciales pueden linealizarse. Box-Jenkins proporciona tablas para valores iniciales; algunos autores sugieren usar 0.1 como valor inicial, y otros sugieren usar cero. Por último, como ejemplo se analiza el proceso de estimación para el modelo ARMA(1,1)

Para efectos prácticos, se utiliza la última ecuación. Lo que se sugiere es minimizar la suma de cuadrados de los errores, observe que en la ecuación (4) sólo se tiene un e lo cual hace más sencillo el procedimiento.

DESPEJANDO e_t de (6)

$$e_t = (1-\theta_1 B)^{-1}(1-\phi_1 B)\omega_t \quad \text{ECUACION NO LINEAL.}$$

$$X_{i,t} = -\frac{\partial[e_t]}{\partial\beta_i}\Big|_{\beta_i=\beta_{i,0}} \quad , \quad \beta\,[\underset{\beta_1}{\underset{\uparrow}{\phi_1}},\ \underset{\beta_2}{\underset{\uparrow}{\theta_1}}] \quad , \beta_{1,0}=\phi_{1,0}$$

$$[e_t] \simeq [e_{t,0} - \sum_{i=1}^{p+q}(\beta_i-\beta_{i,0})X_{i,t}$$

$$X_{1,t} = -\frac{\partial[(1-\theta_1 B)^{-1}(1-\phi_1 B)\omega_t]}{\partial\beta_1}\Big|_{\beta_i=\beta_{i,0}} = (1-\theta_1 B)^{-1}(-B)\omega_t\Big/_{\substack{\theta_{1,0}\\ \phi_{1,0}}}$$

$$X_{1,t} = (1-\theta_{1,0}B)^{-1}(B)\omega_t \ \ldots \ (7)$$

$$X_{2,t} = \frac{-\partial[(1-\theta_1 B)^{-1}(1-\phi_1 B)\omega_t]}{\partial\beta_2}\Big|_{\substack{\theta_{1,0}\\ \phi_{1,0}}} = -(B-\phi_{1,0}B^2)\omega_t(1-\theta_{1,0}B)^{-2} \ldots (8)$$

DE (7)

$$B\omega_t = (1-\theta_{1,0}B)X_{1,t}$$
$$\omega_{t-1} = (1-\theta_{1,0}B)X_{1,t}$$
$$\omega_{t-1} = X_{1,t} - \theta_{1,0}X_{1,t-1}$$
$$X_{1,t} = \omega_{t-1} + \theta_{1,0}X_{1,t-1} \ \ldots \ (9)$$

La ecuación (9) es una ecuación recursiva, equivalente a (7), unicamente que de (7) no se puede calcular X, ya que B es un operador que necesita estar evaluado, lo cual en (7) no sucede.

Haciendo

$X_{1,0} = 0$

$W_0 = 0$

pueden resolverse las ecuaciones recursivas (9) en forma sucesiva:

$$X_{1,1} = W_0 + \theta_{1,0}X_{1,0} = 0$$
$$X_{1,2} = W_1 + \theta_{1,0}X_{1,1} = W_1$$
$$X_{1,3} = W_2 + \theta_{1,0}X_{1,2}$$
$$X_{1,4} = W_3 + \theta_{1,0}X_{1,3}$$
$$\vdots$$

de la ecuación (8):

$$X_{2,t} = -(\omega_{t-1} - \phi_{1,0}\omega_{t-2})(1-\theta_1 B)^{-2}$$

$$(1-\theta_1 B)^2 X_{2,t} = \phi_{1,0}\, \omega_{t-2} - \omega_{t-1}$$

$$(1-2\theta_1 B + \theta_1^2 B^2)\, X_{2,t} = \phi_{1,0}\, \omega_{t-2} - \omega_{t-1}$$

$$X_{2,t} - 2\theta_1 X_{2,t-1} + \theta_1^2 X_{2,t-2} = \phi_{1,0}\, \omega_{t-2} - \omega_{t-1}$$

$$X_{2,t} = 2\theta_1 X_{2,t-1} - \theta_1^2 X_{2,t-2} + \phi_{1,0}\, \omega_{t-2} - \omega_{t-1} \quad \ldots (10)$$

(Ecuación recursiva equivalente a (8))

Haciendo

$X_{a,0} = 0$, $X_{a,-1} = 0$

$W_0 = 0$, $W_{-1} = 0$

$$X_{a,1} = 2\theta_1 X_{a,0} - \theta_{1,0}^2 X_{a,-1} + \phi_{1,0}\, \omega_{-1} - \omega_0 = 0$$

$$X_{a,2} = 2\theta_1 X_{a,1} - \theta_{1,0}^2 X_{a,0} + \phi_{1,0}\, \omega_0 - \omega_1 = -\omega_1$$

$$X_{a,3} = 2\theta_1 X_{a,2} - \theta_{1,0}^2 X_{a,1} + \phi_{1,0}\, \omega_1 - \omega_2$$
$$= -2\phi_{1,0}\, \omega_1 + \theta_{1,0}\, \omega_1 - \omega_2$$

Los valores iniciales $\phi_{1,0}$, y $\theta_{1,0}$, se obtienen de las ecuaciones

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-2\theta_1\phi_1+\theta_1^2} \qquad \rho_2 = \phi_1\, \rho_1$$

$$\phi_{1,0} = \frac{\rho_2}{\rho_1}$$

Este valor se substituye en la primera ecuación la cual, a su vez, puede resolverse por algún método numérico. Para calcular los residuales obtenidos en el modelo inicial $[e_{t,0}]$

$$[\,e_{t,0}] = \omega_t - (\phi_{1,0}\, \omega_{t-1} - \theta_{1,0}\, e_{t,0})$$

la serie se inicializa con $e_{*,0} = 0$, $W_0 = 0$. La ecuación linealizada será

$$\underset{\text{RESIDUALES}}{[e_{t,0}]} + \underbrace{\phi_{1,0} X_{1,t} + \theta_{1,0} X_{2,t}}_{\text{APROXIMACIONES}} = \underbrace{\phi_1 X_{1,t} + \theta_1 X_{2,t}}_{\text{PARAMETROS DESCONOCIDOS}} + \underset{\text{ERROR ALEATORIO}}{[e_t]}$$

N O T A : Los corchetes indican que el valor depende de los valores de ϕ, θ, W, por lo que van cambiando cada aproximación, por lo que ϕ y θ son estimados cada iteración y W son observaciones.

Este nuevo modelo se resuelve por Regresión Lineal Múltiple (bivariada) (4) para encontrar $\phi_{1,1}, \theta_{1,1}$. El proceso se repite hasta encontrar los $\phi_{1,k}$ y $\theta_{1,k}$ adecuados.

(4) Referencias (4) y (5)

1.4 DIAGNOSTICO DEL MODELO

Para entrar en esta tercera etapa, el modelo ya ha sido identificado y estimado. Ahora toca analizar si realmente el modelo describe el comportamiento de los datos, para ésto se deberán examinar los dos puntos siguiêntes:

1. Las autocorrelaciones de los residuales. Estas no deberán denotar ningún patrón si el modelo es el adecuado.

2. El periodograma integral. Se detecta periodicidad en los residuales.

por otro lado, se utilizarán las siguientes estadísticas:

a. Valor del estimador.
b. Estimador del error estandar del estimador.
c. El estadístico T (cociente del estimador sobre su error estañdar)
d. Límites al 95% de confianza.

y como medidas de bondad de ajuste:

e. Suma de los errores al cuadrado.

$$R^2 = 1 - \frac{\sum \hat{e}_t^2}{\sum y_t}$$

f. Estimador de la desviación estándar de los residuales.

Para mostrar un caso de un modelo inadecuado, supôngase que las ACF y PACF, en la fase I se interpretaron de manera incorrecta, y el modelo, en vez de un MA(1) se identificó como un ARMA(1,1). En este caso las autocorrelaciones de los residuales indican claramente un patrón, lo cual sugiere que el modelo NO es el adecuado y será necesario regresar a la fase I para tratar de identificar un mejor modelo.

Supóngase ahora que el error hecho en la fase I fué el seleccionar un MA(2) en vez de un MA(1). esto no es tan malo como el caso anterior. El valor estimado del segundo parámetro del modelo (0) será cercano a cero, lo cual hace que el modelo MA(2) sea equivalente a MA(1) . El problema aquí es el de la PARQUEDAD ó PARSIMONIA (4); el usar más parámetros de los necesarios significa más grados de libertad y más cálculos de los necesarios ya que este segundo parámetro no aporta nada al modelo.

(4) Utilizar el menor número de parámetros de tal forma que el delo represente el comportamiento del fenómeno.

ANALISIS DE ESTACIONARIDAD.

Es importante que se obtenga la estacionaridad desde la etapa de identificación. Hay que recordar que si los datos no son estacionarios se pueden hacer diferencias para que lo sean cuidando de no sobrediferenciar ya que el modelo se puede complicar.

ANALISIS DE LOS RESIDUALES

Para hacer este análisis, lo primero es observar la gráfica de la ACF de los residuales en busca de estacionaridad, estacionalidad, puntos discrepantes y/o influyentes. Supóngase que el modelo

$$\phi(B)\omega_t = \Theta(B)a_t$$

con $W_t = \nabla^d Z$

ha sido resultado de la etapa I y $(\hat{\phi},\hat{\theta})$ ya han sido estimados y minimizan la suma de cuadrados de los residuales. Si el modelo se ajusta a los datos, el comportamiento de la ACF se acerca al ruido blanco. Es decir, la ACF deberá ser sin picos, con valores todos cercanos a cero. Cada autocorrelación será o no significativamente distinta de cero de acuerdo a su estadístico T.

$$t_{r_k} = r_k \;/\; s_{r_k} = r_k$$

(donde r_k es la autocorrelación muestral y s es el error estandar del estimador).

T se distribuye asintóticamente como una t de Student con "N" grados de libertad y se acepta que r =0 si / t /<=2 (ver autocorrelación muestral). Por otro lado en vez de considerar cada uno de los r 's se pueden considerar, por ejemplo, las primeras 20 autocorrelaciones que indican que el modelo no es el adecuado. Supóngase que se tienen las primeras K autocorrelaciones, r ,(k=1,2,3,...,K), de cualquier modelo ARIMA(p,d,q), el modelo será adecuado si

$$Q(K) = n(n+2)\sum_{k=1}^{K}\frac{r_k^2(\hat{a}_t)}{n-k}$$

K= No. de autocorrelaciones

n= No. de datos disponibles

se distribuye aproximadamente como . Entonces la hipótesis de que los errores sean ruido blanco se acepta si

$$Q(K) < X^2_{k-p-q}$$

para reafirmar la prueba de hipótesis pueden obtenerse las primeras diferencias de los residuales

$$a_t = e_t - e_{t-1}$$

el resultado se comportará como un MA(1) con r_k cercano a $\rho_1 = -0.05$, ésto si los residuales son ruido blanco.

NOTA: los residuales de los modelos mal especificados pueden usarse para alterar el modelo en forma lógica.

Supóngase que se usa el modelo MA(1)

$$W_t = (1 - \theta_1 B)e_t \ldots (10)$$

y al observar las ACF y PACF de los residuales, siguen un patrón, por ejemplo, MA(1), ésto indica que el modelo está mal. Es decir que los residuales se comportan como

$$e_t = (1 - \lambda_1 B)a_t$$

Sustituyendo en (10)

$$\begin{aligned} W_t &= (1 - \theta_1 B)(1 - \lambda_1 B)a_t \\ &= (1 - \lambda_1 B - \theta_1 B + \theta_1 \lambda_1 B^2)a_t \end{aligned}$$

$$W_t = a_t - \lambda_1 a_{t-1} - \theta_1 a_{t-1} + \theta_1 \lambda_1 a_{t-2}$$

$$W_t = a_t - (\lambda_1 + \theta_1)a_{t-1} + \theta_1 \lambda_1 a_{t-2} \ldots (11)$$

Obsérvese que (11) corresponde a un modelo MA(2), ésto indica que se le debe aumentar al modelo un parámetro Suponga ahora que los residuales se comportan como un modelo AR(1), es decir

$$e_t = \beta_1 e_{t-1} + a_t$$

Sustituyendo en (10)

$$W_t = (1 - \theta_1 B)(\beta_1 e_{t-1} + a_t)$$

$$W_t = \beta_1 e_{t-1} + a_t - \theta_1 \beta_1 e_{t-2} - \theta_1 a_{t-1}$$

$$W_t = \beta_1 (e_{t-1} - \theta_1 e_{t-2}) + a_t - \theta_1 a_{t-1}$$

el modelo que ahora se obtuvo es un ARMA(1,1) por lo que habrá que agregar un têmino autorregresivo al modelo MA(1). De estos dos ejemplos se puede concluir que de lo residuales depende la forma en que se va a modificar el modelo.

PERIODOGRAMA INTEGRAL ACUMULATIVO

En algunas situaciones, particularmente en el ajuste de series de tiempo estacionales, se puede temer que no se han tomado en cuenta adecuadamente las caracterîsticas PERIODICAS de la serie, asî que se estarâ alerta de la periodicidad de los residuales. La funciôn de autocorrelaciôn puede no ser un muy buen indicador con respecto a la aleatoriedad ya que los efectos periôdicos se diluirân entre ciertas autocorrelaciones.

EL PERIODOGRAMA INTEGRAL provee de un medio efectivo para detectar alguna periodicidad no aleatoria. El periodograma integral se define como

$$C(f_j) = \frac{\sum_{i=1}^{j} R^2(f_i)}{ns^2}$$

donde fi es la frecuencia y se define como un dispositivo para correlacionar los e_t 's con los caminos de seno y coseno a diferentes frecuencias. Esta frecuencia estâ dada por

$$f_i = i/n$$

Por otro lado, sean

$$\text{perîodos} = 1/f_i$$
$$\text{errores} = e_t$$
$$s^2 = \hat{\sigma}^2$$
$$\text{nâmero de observaciones disponibles} = n$$

La funciôn R^2 esta dada por

$$R^2(f_i) = \frac{2}{n}\left[\left(\sum_{t=1}^{n} e_t \cos 2\pi f_i t\right)^2 + \left(\sum_{t=1}^{n} e_t \operatorname{sen} 2\pi f_i t\right)^2\right]$$

y mide quê tanto se ajusta la funciôn trigonomêtrica con frecuencia fi a los residuales. Ahora, si el modelo es el adecuado y los paràmetros se conocen, los e pueden ser calculados de los datos que producirân RUIDO BLANCO y la gràfica del periodograma integral para las series ruido blanco està dada por la figura 1.30

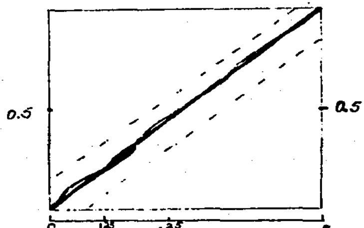


figura 1.30
Periodograma Integral
para series Ruido Blanco

Si el modelo es inadecuado se producen e_t's no aleatorios y el periodograma mostrarâ desviaciones que salen de la banda de confianza. Si se observan saltos regulares, por ejemplo 12,24,36,... se tendrà indicio de estacionalidad, ò por otro lado , si se tienen valores altos a frecuencias bajas, es un indicio de que que existe TENDENCIA en los datos. Este comportamiento se representa en la figura 1.31.

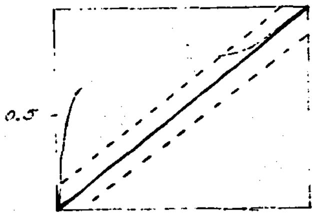


figura 1.31
Periodograma Integral
para series con TENDENCIA

(a la mitad de las observaciones se completa un periodo)

Recuèrdese que en la pràctica no se conocen los valores exactos de los paràmetros, sòlo los estimados y que no se tiene e, unicamente los estimados (residuales) a. Sin embargo para muestras grandes, el periodograma para a tiene propiedades muy similares a las de e. Asì el inspeccionar mas detenidamente el periodograma de a provee de màs informaciòn ùtil para el diagnòstico del modelo, màs aùn cuando èste no fuè ajustado adecuadamente.

MODELOS SOBREESPECIFICADOS.

Una tècnica que puede ser usada para el diagnòstico es el SOBREESPECIFICAR (o sobre ajustar) el modelo. Tambièn puede suceder que el modelo que se tiene identificado (pensando que es el correcto) se encuentre sobre ajustado, es decir tiene màs paràmetros. Los paràmetros redundantes pueden localizarse usando el estadìstico "t" y el estimador de las correlaciones entre los coeficientes. El estadìstico "t" se obtiene como

$$t = \frac{\hat{\beta}_i - \beta_i}{\sigma_{\hat{\beta}_i}} \sim t^{\alpha}_{(n-p-q)}$$

PRUEBA DE HIPOTESIS

$$H_0 : \beta_i = 0$$
$$H_a : \beta_i \neq 0$$

SE ACEPTA H_0 SI $|t| < t_{(n-p-q)}$
$\Rightarrow \exists$ EL MODELO PUEDE SIMPLIFICARSE

Obsèrvese el siguiente ejemplo

EJEMPLO:

Sea el modelo AR(3)

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + e_t$$

Supòngase que se encuentra que $\phi_3 = 0$, entonces el modelo se reduce a un modelo AR(2)

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t$$

este modelo AR(2), no se creò unicamente quitando ϕ_3 del modelo anterior, sino que se volvieron a estimar los paràmetros, y de este modo se reafirma si AR(2) es el adecuado, y observando si el error no se incrementa significativamente (i.e. la suma de cuadrados de los errores). Al quitar un paràmetro aumenta la suma de los errores pero no de manera significativa.

Si en el mismo modelo AR(3) se encuentra que ϕ_2 es cero, se deberán observar los siguientes dos puntos:

1. Si ϕ_2 y ϕ_3 están correlacionados el modelo nuevo será AR(2)

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t$$

2. Si ϕ_2 y ϕ_3 no están correlacionados se elimina a ϕ_2 y se tiene indicio de estacionalidad.

En general, si el parámetro insignificante es el de mayor orden, se le suprimirá y se reestimará el modelo sin él. Es importante comparar el nuevo modelo con el anterior. Si el parámetro insignificante no es el de mayor orden, se deberá analizar la matriz de correlaciones de los parámetros estimados y:

i. Si el parámetro insignificante está fuertemente correlacionado se eliminará el de mayor orden y se reestima el modelo.
ii. Si esta correlación no es fuerte (o no hay) deberá eliminarse el parámetro que resultó aproximadamente cero y esto indica estacionalidad.

MODELOS SUBESPECIFICADOS

Recuérdese que el método de Box-Jenkins maneja el principio de parquedad (parsimony), entonces hay que verificar si el modelo tentativo contiene el número adecuado de parámteros. El caso de modelos subespecificados, es el caso contrario al anterior, y se examinará si faltan parámetros en vez de que sobren, por lo que podrá incluirse un parámetro adicional y ver si se mejora el modelo. También es conveniente analizar las funciones de autocorrelación y autocorrelación parcial muestrales enlistando todos los posibles modelos. Después se procede a ajustar empezando con el modelo MAS SENCILLO , y de uno en uno se van agregando los parámetros extra, hasta encontrar el mejor modelo. Es recomendable ir agregando parámetros uno por uno ya que existe el riesgo de la REDUNDANCIA de parámetros. Esto generalmente ocurre cuando se agregan parámetros al mismo tiempo. Para ilustrar ésto obsérvese el siguiente ejemplo.

EJEMPLO:

Sea el modelo AR(2)

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t \quad \ldots \quad (1)$$

$$Z_{t-1} = \phi_1 Z_{t-2} + \phi_2 Z_{t-3} + e_{t-1} \quad \ldots \quad (2)$$

restando (1) - (2)

$$Z_t - Z_{t-1} = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t - \phi_1 Z_{t-2} - \phi_2 Z_{t-3} - e_{t-1}$$

$$Z_t = \underbrace{(1+\phi_1)}_{\phi_1} Z_{t-1} + \underbrace{(\phi_2 - \phi_1)}_{\phi_2} Z_{t-2} - \underbrace{\phi_2}_{\phi_3} Z_{t-3} + e_t - \underbrace{e_{t-1}}_{\theta_1}$$

ahora el modelo es ARMA(3,1). Todos los parámetros son significativos pero al observar la matriz de correlacion se observarà que ϕ_2 y ϕ_3 estân muy correlacionados ya que ambos estân en funciôn de ϕ_2.

Matriz de Correlaciôn de Parâmetros

$$\begin{array}{c|cccc} & \phi_1 & \phi_2 & \phi_3 & \theta_1 \\ \hline \phi_1 & 1 & & & \\ \phi_2 & & 1 & \approx 1 & \\ \phi_3 & & \approx 1 & 1 & \\ \theta_1 & & & & 1 \end{array}$$

El problema de la redundancia de parâmetros puede ejemplificarse como sigue

Sea

$$(1 - \phi_1 B) Z_t = e_t \ldots (3)$$

multiplicando (3) por $(1-\theta_1 B)$

$$(1-\phi_1 B)(1-\theta_1 B)Z_t = (1-\theta_1 B)e_t$$
$$(1-\theta_1 B-\phi_1 B+\phi_1\theta_1 B^2)Z_t = (1-\theta_1 B)e_t$$
$$Z_t = \theta_1 Z_{t-1} + \phi_1 Z_{t-1} + \phi_1\theta_1 Z_{t-2} + e_t - \theta_1 e_{t-1}$$
$$Z_t = \underbrace{(\phi_1+\theta_1)}_{\phi_1} Z_{t-1} + \underbrace{\phi_1\theta_1}_{\phi_2} Z_{t-2} + e_t - \theta_1 e_{t-1}$$

ϕ_1 y ϕ_2 están correlacionados, los dos contienen la misma información. Por lo tanto se está reflejando el mismo comportamiento que con AR(1), sólo que se está utilizando un modelo más complicado.

En resumen, se puede decir que en esta etapa, se analizan los modelos "tentativos", ya estimados, buscando el modelo que represente mejor el comportamiento del fenómeno tratando de que sea, el mismo tiempo, el más sencillo posible.

1.5 PRONOSTICOS

Habiendo considerado algunas propiedades de los modelos ARIMA, se muestra como pueden ser usados para pronosticar valores futuros de una serie de tiempo observada. Una vez que se ha encontrado el modelo adecuado puede ser utilizado para generar pronósticos. Se llamarâ **ORIGEN** al período actual (Yn, dato original) y se desea pronosticar el valor de Y para "h" períodos hacia adelante, i.e., se desea conocer el valor de Yn+h. Al valor de "h" se le llama **HORIZONTE**.

NOTACION:

Yn(h) = valor pronosticado de Yn+h. Por lo tanto no es una variable aleatoria sino un valor que se conoce.

Yn+h = Variable aleatoria cuya función de distribución es condicional a los datos pasados y presentes, así como al modelo que rige al fenómeno.

A medida que transcurre el tiempo las Yn+h se vuelven conocidas y pueden compararse con las Yn(h).

1.5.1 FUNCION DE COSTO CUADRATICA

El error al hacer un pronóstico puede definirse como:

$$\epsilon_n(h) = Y_{n+h} - Y_n(h)$$

este error ocasiona un costo $C(\epsilon)$. Este costo usualmente se estima como

$C(\epsilon) = \alpha\,[\epsilon\, n(h)]^2$ ...FUNCION DE COSTO CUADRATICA

Donde α =costo unitario de cada error al cuadrado.

Esta función se usa ya que existe costo tanto si el pronóstico es errôneo por exceso como por defecto. Graficamente:

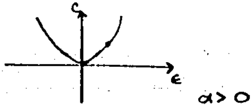


figura 1.32

También podría estimarse como:

$$C(\epsilon) = \begin{cases} \alpha \epsilon_n(h) & \text{si} \quad \epsilon_n(h) > 0 \\ 0 & \text{si} \quad \epsilon_n(h) = 0 \\ -\beta \epsilon_n(h) & \text{si} \quad \epsilon_n(h) < 0 \end{cases}$$

$\alpha, \beta > 0$

Graficamente se comporta como

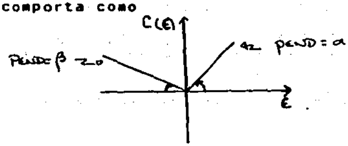


figura 1.33

En el caso de la función de costo cuadrática se tiene la propiedad de que la media de la distribución del pronóstico minimiza el costo es decir

SEA $m_h = E(Y_{n+h})$, $d = m - m_h$
$m = m_h + d$ = CUALQUIER PRONOSTICO

$$E[C(\epsilon)] = E[\alpha(\epsilon_n(h))^2] = E[\alpha(Y_{n+h} - Y_n(h))^2]$$
$$= E[\alpha(Y_{n+h}^2 - 2Y_{n+h}Y_n(h) + Y_n^2(h))]$$
$$= \alpha[E(Y_{n+h}^2) - 2E(Y_{n+h}, Y_n(h)) + E(Y_n^2(h))]$$
$$E[C(\epsilon)] = \alpha[E(Y_{n+h}^2) - 2Y_n(h)m_h + Y_n^2(h)]$$

Ahora, tomando como pronóstico a la media:

$$E[C(\epsilon)] = \alpha[E(Y_{n+h}^2) - 2m_h^2 + m_h^2]$$

$$E[C(\epsilon)] = \alpha[E(Y_{n+h}^2) - m_h^2] \quad \ldots \quad (1)$$

y si por el otro lado, se toma como pronóstico a cualquier otro punto

$$Y_n(h) = mh + d$$

$$
\begin{aligned}
E[c(\epsilon)] &= \alpha\left[E(Y_{n+h}^2) - 2Y_n(h)mh + Y_n^2(h)\right] \\
&= \alpha\left[E(Y_{n+h}^2) - 2(mh+d)mh + (mh+d)^2\right] \\
&= \alpha\left[E(Y_{n+h}^2) - 2m^2h - 2mhd + m^2h + 2mhd + d^2\right]
\end{aligned}
$$

$$E[c(\epsilon)] = \alpha\left[E(Y_{n+h}^2) - m^2h + d^2\right] \ldots \quad (2).$$

Comparando (1) con (2) el costo mínimo se obtiene con (1).

Para calcular E(Yn+h):

Sea Yt un proceso estacionario e invertible ARIMA(p,d,q)

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}.$$

ó

$$\phi(B) Z_t = \theta(B) e_t$$

donde

$Z_t = \nabla^d Y_t$

entonces, si t=n+h

$$Z_t = \phi_1 Z_{n+h-1} + \phi_2 Z_{n+h-2} + \ldots + \phi_p Z_{n+h-p} + e_{n+h} - \theta_1 e_{n+h-1} - \ldots - \theta_q e_{n+h-q}$$

efectuar los siguientes puntos

1. Reemplazar el error actual y los pasados, i.e., los en+j, j<=0, por los residuales.

2. Reemplazar cada error futuro, en+j, j>=1 por su valor esperado (cero).

3. Reempazar Zn+j, j<=0, por los valores observados.

4. Pronosticar en forma sucesiva los >n+j, j>=1, hasta Zn+h-1 y sustituir sus valores.

Una vez hecho esto, es necesario aplicar las sumas correspondientes para regresar a la variable Yt. Para ilustrar lo anterior véase el siguiente ejemplo.

EJEMPLO:

$$Z_{t-\mu} = \phi_1 (Z_{t-1-\mu})\, e_t \qquad ; \quad \mu = E(Z_t)$$
$$Z_t = (1-\phi_1)\mu + \phi_1 Z_{t-1} + e_t$$

PRONOSTICOS

$$Z_n(1) = E(Z_{n+1}) = E\left[(1-\phi_1)\mu + \underset{\text{OBSERVADO}}{\phi_1 Z_n} + e_{n+1}\right]$$

$$Z_n(1) = (1-\phi_1)\mu + \phi_1 Z_n$$
$$Z_n(2) = E(Z_{n+2}) = E\left[(1-\phi_1)\mu + \phi_1 Z_{n+1} + e_{n+2}\right]$$
$$= (1-\phi_1)\mu + \phi_1 Z_n(1)$$
$$= (1-\phi_1)\mu + \phi_1\left[(1-\phi_1)\mu + \phi_1 Z_n\right]$$
$$= (1-\phi_1)\mu + \phi_1(1-\phi_1)\mu + \phi_1^2 Z_n$$
$$Z_n(2) = (1-\phi_1)\mu(1+\phi_1) + \phi_1^2 Z_n$$
$$Z_n(3) = E(Z_{n+3}) = E\left[(1-\phi_1)\mu + \phi_1 Z_{n+2} + e_{n+3}\right]$$
$$= (1-\phi_1)\mu + \phi_1 Z_n(2)$$
$$= (1-\phi_1)\mu + \phi_1\left[(1-\phi_1)\mu(1+\phi_1) + \phi_1^2 Z_n\right]$$
$$= (1-\phi_1)\mu + \phi_1(1-\phi_1)\mu(1+\phi_1) + \phi_1^3 Z_n$$
$$Z_n(3) = (1-\phi_1)\mu(1+\phi_1+\phi_1^2) + \phi_1^3 Z_n$$
$$\vdots$$
$$Z_n(h) = (1-\phi_1)\mu\underbrace{(1+\phi_1+\phi_1^2+\dots+\phi_1^{h-1})}_{\text{SERIE GEOMETRICA DECRECIENTE}} + \phi_1^h Z_n$$

COMO

$$\lim_{h\to\infty}(1+\phi_1+\phi_1^2+\dots+\phi_1^{h-1}) = \frac{a}{1-r} = \frac{1}{1-\phi_1}$$

RESULTA

$$\lim_{h\to\infty} Z_n(h) = (1-\phi_1)\frac{1}{1-\phi_1}\mu + 0 = \underline{\underline{\mu}}$$

En general, cuando el valor de "h" es muy grande el pronóstico tiende a la media de la serie (véase figura 1.34). Debe de ser claro, intuitivamente, que mientras más se aleja el horizonte se pierde confiabilidad y el mejor pronóstico es la media, el intervalo de confianza aumenta. No tiene caso hacer pronósticos lejanos.

MODELO MA(1)

$$Z_{n+1} = \mu + e_{n+1} - \theta_1 e_n$$

$$Z_n(1) = E(Z_{n+1})$$

$$= \mu + E(e_{n+1}) - \theta_1 E(e_n)$$

$$= \underline{\underline{\mu}}.$$

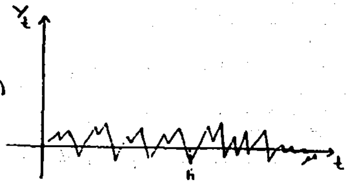


figura 1.34

1.5.2 PRONOSTICO DEL ERROR MINIMO CUADRATICO MEDIO Y SUS PROPIEDADES

En algunas ocasiones se requiere hacer un pronóstico puntual y se desea minimizar el error del pronóstico por lo que se utiliza la esperanza condicional que tiene la propiedad de tener el error mínimo cuadrático medio del pronóstico. Esto es, si el modelo es correcto no hay otro pronóstico que produzca errores cuyos cuadrados tengan un menor valor esperado.

Se denota por $\hat{Y}_t(\ell)$ el pronóstico condicional esperado de $Y_{t+\ell}$ dados los datos hasta el tiempo t y a los datos $y_t, y_{t-1}, \ldots$ por Ht. Si F es otro pronóstico de $Y_{t+\ell}$, entonces el error mínimo cuadrático medio de F es

$$E[(Y_{t+\ell} - F)^2 / H_t] = E\{[Y_{t+\ell} - \hat{Y}_t(\ell) + g]^2 / H_t\}$$

donde g es la diferencia entre F y el pronóstico condicional esperado $\hat{Y}_t(\ell)$

$$E[(Y_{t+\ell} - F)^2 / H_t] = E\{[Y_{t+\ell} - (\hat{Y}_t(\ell) + g)]^2 / H_t\}$$
$$= E[Y_{t+\ell} - (\hat{Y}_t(\ell))^2 / H_t] + g^2$$

el cual es el error cuadràtico medio para la esperanza condicional del mejor pronòstico ; g es mayor que cero si F difiere de $\hat{Y}_t(l)$. Para minimizar el error cuadràtico, hàgase g=0 y se concluye que $\hat{Y}_t(l)$ tiene el error mìnimo cuadràtico medio del pronòstico.

1.5.3 CALCULO DE LA ESPERANZA CONDICIONAL

Considèrese el càlculo de la esperanza condicional del pronòstico de la forma de Ecuaciòn en Diferencias del modelo ARIMA(!) . Supòngase que se està en el tiempo t y se desea pronosticar Y (1) que es la esperanza condicional E (Y_{t+1}/ H_t). Y_{t+1} està dado por

$$Y_{t+1} = \Phi_1 Y_t + \Phi_2 Y_{t-1} + \ldots + \Phi_{p+d} Y_{t-(p+d)+1} + \delta + e_{t+1} - \theta_1 e_t - \ldots - \theta_q e_{t-q+1}$$

donde Φ_i està determinado por ϕ_i y el valor de d. Como E (e_{t+1}/ H_t) es cero, $\hat{Y}_t(1)$ està dado por

$$\hat{Y}_t(l) = E(Y_{t+1} | H_t) = \Phi_1 Y_t + \ldots + \Phi_{p+d} Y_{t-(p+d)+1} + \delta + e_{t+1} - \theta_1 e_t - \ldots - \theta_q e_{t-q+1}$$

por otro lado

$$Y_{t-1} - \hat{Y}_{t-1}(1) = e_t \qquad , \; i = 0, 1, 2, \ldots$$

Considèrese el problema de generar pronòsticos para un modelo MA(2). Entonces

$$\hat{Y}_t(1) = \delta - \theta_1 \hat{e}_t - \theta_2 e_{t-1}$$

para poder evaluar $\hat{Y}_t(1)$ se necesitan los valores de e_1 y e_0. POr otro lado el pronòstico anterior fue $\hat{Y}_{t-1}(1)$ y el error de pronòstico fuè

$$\begin{aligned}[Y_t - \hat{Y}_{t-1}(1)] &= (\delta + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}) - (\delta - \theta_1 e_{t-1} - \theta_2 e_{t-2}) \\ &= e_t\end{aligned}$$

Similarmente està dado por $[Y_{t-1} - \hat{Y}_{t-2}(1)]$ entonces se puede evaluar $\hat{Y}_t(1)$

$$\hat{Y}_t(1) = \delta - \theta_1 E\underbrace{(Y_t - \hat{Y}_{t-1}(1))}_{e_t} - \theta_2 E\underbrace{(Y_{t-1} - \hat{Y}_{t-2}(1))}_{e_{t-1}}$$

(!) Referencia (1)

Para calcular el primer pronóstico se debe hacer un supuesto sobre el valor de la perturbación previa. El recurso más intuitivo es darle su valor esperado ($e_t = 0$). Ahora se puede pronosticar para ℓ períodos adelante de t.

$$\hat{Y}_t(\ell) = E(Y_{t+\ell} \mid H_t)$$

$$= \phi_1 E(Y_{t+\ell-1} \mid H_t) + \ldots + \phi_{p+d} E[(Y_{t+\ell-(p+d)} \mid H_t]$$

Como la esperanza de los errores pasados son cero desaparecen de la ecuación si $\ell > q$. Entonces

$$\hat{Y}_t(\ell) = \phi_1 \hat{Y}_t(\ell-1) + \ldots + \phi_{p+d} \hat{Y}_t(\ell-p-d)$$

donde $\hat{Y}_t(\ell-i)$ es la observación $Y_{t+\ell-i}$ $i <= \ell$

1.5.4 ACTUALIZACION DE PRONOSTICOS

Algunas veces es conveniente, cuando ya se ha hecho el pronóstico, actualizarlo de período a período sin realizar todos los cálculos. Para actualizar los pronósticos se utiliza la forma Golpe Aleatorio (1)

$$Y_{t+\ell} = \mu + e_{t+\ell} + \Psi_1 e_{t+\ell-1} + \Psi_2 e_{t+\ell-2} + \ldots$$

El pronóstico $\hat{Y}_t(\ell)$ en el tiempo t se expresa como

$$\hat{Y}_t(\ell) = \mu + \Psi_\ell e_t + \Psi_{\ell+1} e_{t-1} + \ldots$$

Puesto que el pronóstico corresponidente de $Y_{t+\ell}$ que fuè hecho en el período t-1 es

$$\hat{Y}_{t-1}(\ell+1) = \mu + \Psi_{\ell+1} e_{t-1} + \Psi_{\ell+2} e_{t-2} + \ldots$$

la diferencia $\hat{Y}_t(\ell)$ y $\hat{Y}_{t-1}(\ell+1)$ es la actualización del pronóstico

$$\hat{Y}_t(\ell) - \hat{Y}_{t-1}(\ell+1) = \Psi_\ell e_t$$

(1) Referencia (1)

pero e_t es el error del pronóstico de Y_t y es la diferencia entre Y_t y $\hat{Y}_{t-1}(1)$, entonces

$$e_t = Y_t - \hat{Y}_{t-1}(1)$$

Una vez conocido e_t, se puede agregar el pronóstico Y_{t+1}. Este error e_t también tiene una varianza y una desviación estándar. De la forma golpe aleatorio, el error del pronóstico [$Y_{t+\ell} - \hat{Y}_t(\ell)$] denotado por E $e_t(\ell)$, está dado por

$$e_t(\ell) = e_t + \psi_1 e_{t-1} + \ldots + \psi_{\ell-1} e_{t-\ell+1}$$

el error tiene esperanza cero, puesto que cada perturbación tiene esperanza cero

$$\begin{aligned} Var[e_t(\ell)] &= \sigma^2 + \psi_1^2 \sigma^2 + \psi_{\ell-1}^2 \sigma^2 \\ &= \sigma^2 (1 + \psi_1^2 + \ldots + \psi_{\ell-1}^2) \end{aligned}$$

la desviación estándar

$$S.D.[e_t(\ell)] = \sigma (1 + \psi_1^2 + \ldots + \psi_{\ell-1}^2)^{1/2}$$

de acuerdo a èsto las medidas de disperción de $e_t(\ell)$ se calculan en base a ψ y σ^2. Por último se tienen 2 opciones de actualizar el pronóstico:

1. PRONOSTICOS ACTUALIZADOS EN FORMA SECUENCIAL. Usando el punto adicional se reestiman los parâmetros.

2. PRONOSTICOS ADAPTATIVOS. Se dejan los valores de los parâmetros pero se cambia el origen al nuevo valor conocido. Si el modelo va variando quiere decir que no existe un patrôn constante y hay que revisar sistemâticamente el ajuste.

CAPITULO 2

DESCRIPCION DEL FENOMENO

El objetivo de este capítulo es explicar detalladamente el fenómeno que va a ser analizado posteriormente. Se presenta todo lo que comprende a los Ingresos por Turismo en México. Se toman como punto de partida las siguientes definiciones:

TURISMO: Afición a viajar por gusto de recorrer un país. Organización de los medios conducentes a facilitar estos viajes.

TURISTA: Se considera que es turista cualquier persona que visita temporalmente un país distinto al de su residencia habitual (comprendiendo tanto a los extranjeros como a los nacionales no residentes en su país de origen), por cualquiera de los siguentes motivos:

a. Por recreo, descanso, razones de familia o por motivos de salud.

b. Por asistencia a convenciones o reuniones de toda clase (asuntos científicos, administrativos, religiosos, deportivos, etc.).

c. Por negocios.

d. En tránsito, cuando permanezcan por lo menos 24 horas en el país.

e. Los estudiantes que viajan por cuenta propia y/o becados por entidades no residentes en el país.

Quedan excluidos del concepto de turista los siguientes tipos de viajeros:

f. Los visitantes que permanecen en las poblaciones fronterizas sin salir de los límites de ellas.

g. Las personas que llegan, con o sin contrato de trabajo, a desempeñar un empleo, a ejercer una profesión ó a residir en el país.

h. Los diplomáticos.

i. Los estudiantes que viajan becados por entidades residentes en el país.

j. Los visitantes que pasan por el país con permanencia menor de 24 horas.

INGRESOS POR TURISMO: Actividad económica en la que se efectúa una prestación de servicios y una oferta de bienes por parte de los residentes de un país a los de otros. La contraprestación de estos servicios y de esta adquisición de bienes se efectúa mediante una entrada de divisas al extranjero. El turismo constituye un caso especial de exportación de bienes y servicios sin desplazamiento en su mayor parte. Es conveniente señalar que èstos son medidos en dòlares y el anàlisis que se harà posteriormente respetarà esta medida.

IMPORTANCIA ECONOMICA DEL TURISMO EN MEXICO

El turismo se ha convertido en una actividad fundamental para el desarrollo económico de algunos países; la importancia del mismo es debido a razones infraestructurales y a otras susceptibles a una acción económica pública ó privada. En ambos aspectos el continente europeo es el que se encuentra en la situación más privilegiada. Europa, gran centro turístico mundial, se encuentra con la corrientes turísticas intraeuropeas que cada vez adquieren mayor importancia, y la americana (basicamente la estadounidense).

En Mèxico, las divisas generales que dejan los visitantes ayudan mucho a nivelar la balanza de pagos al exterior y a mantener las reservas en dòlares.

El turismo es una de las industrias más productivas y satisfacorias que puede fomentar un país. La moneda turística constituye una de las mejores formas de la extensa distribución de los ingresos nacionales que se conocen, ya que nadie tiene el monopolio de este negocio.

En tèrminos económicos y sociales puede decirse que el turismo ha sido una actividad con recuperación, no sòlo significativa sino creciente; no ùnicamente en cuanto a la aportación de divisas sino a la oportunidad que ha tenido Mèxico ante tantos cientos de millares de visitantes de mostrar su expresión cultural y civilizada.

2.1 MARCO TEORICO

ANTECEDENTES

La evolución histórica del turismo en México es compleja y se puede dividir en cuatro etapas fundamentales:

a) Etapa del Nacimiento del Turismo moderno (1920-1940)
b) Etapa del Desarrollo del Turismo Moderno (1940-1953)
c) Etapa de Tecnificación del Turismo en México (1958-1975)
d) Etapa de Consolidación del Turismo Moderno (1975-a la fecha)

PRIMERA ETAPA: En esta etapa se crea la Comisión Mixta Proturismo para acrecentar el movimiento turístico hacia el país. El gobierno decide promulgar la Ley Orgánica que crea la comisión Nacional de Turismo, encomendándole, entre otras cosas:

a) Fomentar el turismo nacional e internacional.

b) Promover el desarrollo de la industria hotelera

c) Vigilar que los turistas gocen de facilidades y garantías en la aplicación de las leyes de sanidad, migración y aduanas.

d) Realizar propaganda en el interior y en el exterior de materia de turismo.

En los primeros años de los 30's empieza a resaltar a la vista nacional la hotelería turística, y por lo tanto la industria de servicios, se fundan las primeras agencias de viajes y se reglamentan los guías de turistas. A partir de 1933 se crean varios órganos que deben actuar en conjunto para dar gran impulso y creciente importancia al turismo, en los que interviene tanto la iniciativa privada como las dependencias gubernamentales con relación al turismo. Es por iniciativa del gobierno y apoyo del sector privado que se funda la Asociación Mexicana de Turismo en Marzo 25 de 1939, y en base a la cual se irían organizando y construyendo mas instituciones del ramo turístico.

SEGUNDA ETAPA:

Durante la administración de Manuel Avila Camacho y el período de Miguel Alemán Valdés, se crean las bases legales, instituciones y políticas que impulsan el desarrollo turístico de México. Se crea una nueva "Comisión Nacional de Turismo ", teniendo las siguientes funciones:

i. Reglamentar el funcionamiento de las agencias de viajes y clasificarlas en agencias de primera, segunda y tercera clase.
ii. Normar la presentación del oficio de guía de turistas.
iii. Fijar el precio de los transportes y precisar la de otros servicios turísticos.

El 15 de diciembre de 1949, la primera "Ley Federal de Turismo" norma la prestación de servicios turísticos y otorga al estado la facultad de aprobar tarifas. Un importante factor para el desarrollo turístico, es la creación del Fondo de Garantía y Fomento del Turismo (FONGATUR), cuya función básica, fue la de promover nuevos centros turísticos y el desarrollo de los ya existentes, así como favorecer las afluencias de las corrientes turísticas nacional y extranjeras. De 1940 a 1958, es la etapa de gran desarrollo ya que en ella se dan las bases legales, institucionales y promocionales que habían de conducir al país al desarrollo turístico con que hoy se cuenta.

TERCERA ETAPA:

En esta etapa, los esfuerzos oficiales y privados se encaminan a buscar una transformación técnica de la actividad turística para convertirla en una auténtica industria con bases estables y progresivas. La labor de tecnificación turística y de promoción institucional, caracterizan los objetivos en esta etapa del desarrollo turístico de México.

A principios de la década de los sesentas se empieza a trabajar en la reorganización de la Ley Federal de Turismo. Se promulgó la Segunda Ley Federal de Turismo, la cual otorga actividades importantes al Departamento de Turismo, entre las cuales destacan las siguientes:

a) Fomentar el turismo tanto en el plano nacional como internacional.
b) Supervisar los servicios turísticos.
c) Crear o autorizar escuelas dedicadas a la capacitación de los prestadores de servicios.
d) Dirigir la propaganda oficial en materia de turismo.
e) Organizar y registrar las Cámaras de Trusimo en su confederación, organismos que deben agrupar a servidores turísticos.

f) Formar un Catâlogo Turîstico Nacional.
g) Cooperar con las demâs dependecias del Ejecutivo, centralizadas y paraestatales, asî como con los gobiernos de los estados federados, en todo aquello que tenga relaciôn con el turismo.

Es en esta etapa cuando es creado el Instituto Mexicano de Investigaciones Turîsticas, por la Secretarîa de Hacienda y Crêdito pûblico, contando con el apoyo y patrocinio del Consejo Nacional de Turismo, del Departamento de Turismo, de la Secretarîa de gobernaciôn y de la Universidad Nacional Autônoma de Mêxico, como un organismo de caracter têcnico que tendrâ por objeto estudiar en forma sistemâtica y permanente el fenômeno turîstico, dar asesorîa têcnica especializada al sector turismo y dar acceso a su acervo documental turîstico al estudiante, profesor, investigador y demâs personas que requieran informaciôn de esta actividad.

Desde principios de la dêcada de los años sesentas, Mêxico empieza a incrementar su turismo receptivo, pasando en 1960 de 760,338 turistas a 1,062,972 en 1963. En esta misma dêcada, se promulgan importantes disposiciones jurîdico - administrativas como son los reglamentos de guîas de turistas, de guîas de choferes y similares, asî como un nuevo reglamento de agencias de viajes; la tercera Ley Federal de Turismo, el Plan Nacional de Desarrollo Turîstico, etc..

Son pues, las actividades de este perîodo asî como las de los primeros años de los setentas, las que dan el eslabôn perfecto para que el turismo vaya tomando el matiz de consolidaciôn.

CUARTA ETAPA:

Dentro del sexenio de Luis Echeverrîa A. y del de Josê Lôpez Portillo se caracterizan los siguientes puntos dentro del turismo:

i) La elevaciôn del Departamento Autônomo de Turismo a rango de Secretarîa de Estado.
ii) La importancia decisiva del turismo en la actividad econômica del Estado.
iii) La promulgaciôn de dos leyes federales dentro del turismo.
iv) La mayor presupuestaciôn, programaciôn y planeaciôn de que es objeto el turismo.
v) La internacionalizaciôn y maduraciôn de la imâgen polîtica de Mêxico.
vi) Ser el paîs sede, permanente y eventual, de importantes actividades turîsticas tanto en el campo de tecnificaciôn como en el cultural.
vii) La realizaciôn de convenios internacionales en diversos paîses.

La Secretaría de Turismo es la encargada de formular la programación de la actividad turística nacional y organizar, coordinar, vigilar y fomentar su desarrollo protegiendo los medios que proporcionan los servicios al turista y a las demás funciones a las que se refiere la Ley Federal de Turismo. Por otro lado, se acordó que la Secretaría de Turismo coordinara sus funciones con otras entidades públicas para que el turismo en México fuera más viable.

En 1980, se consolidan los esfuerzos institucionales y son: La Nueva Ley Federal de Turismo y el Plan Nacional de Turismo la cual tiene por objeto:

1) Planeación y programación de la actividad turística.
2) Promoción de la demanda exterior e interior, así como el fomento y desarrollo de la oferta en materia de turismo.
3) Creación, conservación , mejoramiento, protección y aprovechamiento dde los recursos turísticos de la nación.
4) Regularización y control de los servicios turísticos principales y conexos.

PLANES MAESTROS PARA UN DESARROLLO TURISTICO

Durante el perìodo 1975-1977, FONATUR desarrollò los planes maestros(2) en San Josè del Cabo y Loreto-Nopolò, a un nivel tal que permitieran cubrir todos los requerimientos tècnicos, econòmicos y financieros establecidos por el Banco Mundial y, por lo tanto, desembocar en la concertaciòn del crèdito con que se està financiando parcialmente su realizaciòn.

En 1978, ademàs de la creaciòn de los fideicomisos respectivos, se procediò a revisar la capacidad de respuesta de esos planes maestros, ante los lineamientos generales del Sistema Nacional de Planificaciòn Turìstica y el Plan Nacional de Turismo de Desarrollo Urbano.

Los conceptos bàsicos introducidos en los planes maestros que ahora rigen las mùltiples acciones del programa de obras son:

1.- Relacionar y ubicar los nuevos desarrollos turìsticos dentro de una visiòn del conjunto de la regiòn, constituìda por el Estado y por diversas subregiones de desarrollo turìstico predominante; resultantes del Sistema Nacional de Planificaciòn Turìstica y de los planes nacional y estatal de desarrollo urbano.

2.- Convertir a los poblados urbanos actuales en el punto de partida de los proyectos, facilitando la integraciòn de la poblaciòn permanente a la actividad turìstica y econòmica que se presentarà.

3.- Diversificar aùn màs la oferta turìstica, buscando un conjunto de amenidades que permitan una mejor ocupaciòn del tiempo libre; que resulten acordes con los patrones que rigen los gustos y preferencias del mercado turìstico y se aprovechen los atractivos turìsticos de que se dispone.

4.- Lotificar las zonas habitacionales de tal manera que se logre la amplitud y el espacio requeridos por modos de vida propios de la provincia.

5.- Remodelar el centro històrico turìstico de los poblados actuales, proporcionando, ademàs, auxilio tècnico que facilite el proceso de cambio que vaya haciendo congruente el desarrollo urbano con el desarrollo turìstico.

6.- Realizar y promover el equipamiento urbano y turìstico necesarios.

7.- Propiciar la conservaciòn y mejoramiento del medio ambiente.

(2) La informaciòn de estos planes maestros data aproximadamente de 1971-1981.

8.- Encauzar, mediante las instalaciones específicas, el aprovechamiento de las energías libres.

DIVERSIFICACION DE LA OFERTA

México es poseedor de una extensa y enorme variedad de recursos turísticos, sin embargo no han sido explotados con igual intensidad trayendo como consecuencia una tendencia preferencial, por parte del turismo extranjero, e incluso nacional, a incidir en unos pocos centros turísticos, interiores ó de costa.

Es lógico que esta tendencia preferencial haya producido desequilibrios en el desarrollo así como crecimientos de los centros turísticos, que deben ser combatidos mediante un adecuado fomento y explotación de aquellos centros que ó están en un estado de potencialidad turística ó han iniciado ya su proceso de desarrollo, logrando así un crecimiento turístico más armónico.

DIVERSIFICACION DE LA DEMANDA

Por razones coyunturales, todo el mercado turístico internacional hacía posible la vigencia de la ley de la proximidad geográfica como reguladora de la demanda turística. La demanda turística mexicana en el extranjero se localizaba en E.U.A. A partir de 1961, gracias a la gran promoción turística de México en el exterior se han dado los primeros pasos a fin de lograr una diversificación de la demanda. En opinión del IMIT(3), la diversificación de la demanda debe mantenerse dentro de los siguientes criterios:

a) Mantener el ritmo de crecimiento de la corriente turística extranjera total; contando con que cada día será más difícil mantener ese ritmo.

b) Localización y explotación de nuevos mercados extranjeros dentro de los cuales destacan el europeo, canadiense, centro y sudamericano.

c) Localización de grandes núcleos de lo que convencionalmente se ha designado como turismo "tipológico", es decir, aquellas grandes agrupaciones que ejercitan algún departamento, tienen alguna marcada preferencia por el desarrollo de alguna actividad. Tal sería el caso,por ejemplo, de las asociaciones de deportistas

(3) IMIT= Instituto Mexicano de Investigaciones Turísticas.

naûticos.

d) Aprovechamiento de los movimientos turísticos extraordinarios para lograr en ellos una difusión sencilla de los recursos turísticos.

INCREMENTO DEL TURISMO NACIONAL

Por razones circunstanciales que no solamente atienden a factores turísticos, sino que se encuentran radicadas en la problemática del actual estado de desarrollo socioeconómico; el turismo nacional, en referencia al turismo internacional, no ha sido aprovechado en toda su potencialidad, sin embargo, resulta urgente promover la circulación intensa de las corrientes turísticas nacionales, que no solamente juegan un papel de equilibrio en las épocas deficitarias del turismo extranjero, sino que en algún momento permitirán desviar la demanda mexicana (ó atenuarla) en el extranjero; lográdose con ello que la balanza turística resulte siempre más favorable.

Para tal efecto urge localizar en forma técnica y objetiva los mercados nacionales y potenciales del turismo con base en indicadores generales que permitan la promoción realizada para tal efecto y se produzcan sus máximos rendimientos.

ABATIMIENTO DE TEMPORADAS

Esta demostrado que es una tendencia natural del fenómeno turístico el de incidir en las llamadas curvas de estacionalidad; pero también está probado que dichas curvas pueden ser atenuadas e incluso abatidas. Un caso puede ser que en 1965 fué posible abatir las curvas de temporalidad en el flote de las corrientes turísticas que visitaron la ciudad de México, Acapulco; hasta el punto que se ha caído en el problema inverso de la saturación.

En opinión del IMIT, la estacionalidad está ligada con la promoción, la cual debe ser intensificada durante épocas deficitarias, así como las actividades, diversificación de demanda, pues se trata de factores concurrentes que propician mayor densidad y fluidez en la corriente de visitantes.

TRANSFORMACION DEL TURISMO FRONTERIZO

De todos es conocido, que determinados centros ubicados en la frontera norte, reciben una afluencia más que considerable de visitantes extranjeros de muy escasa temporalidad, ello trae como consecuencia que en los puntos mismos de la República se localicen fuertes núcelos de corrientes turísticas, suceptibles de ser promovidas para que visiten el interior de la república, en especial determinados centros o lugares vecinos a la frontera capaces de satisfacer la motivación de descanso que les provoca.

Es por ello que el IMIT considera que mediante el establecimiento de determinados centros turísticos puente, en atención a que jugarían tal papel para transformar el turismo fronterizo en turismo al interior del país se obtendría, como resultado inmediato, una mayor estancia del turista , un mejor rendimiento económico, así como una mayor estabilidad y control de dichos visitantes.

2.2 ESTADISTICAS GENERALES

Entre las labores más importantes que tiene a su cargo la Subdirección de Investigación Económica y Bancaria del Banco de México, S.A., se encuentra, justamente, la de elaborar la balanza de pagos del país. La necesidad de obtener materiales confiables para su elaboración es una preocupación constante que se traduce en una sostenida mejoría de los sistemas de captación y compilación de datos.

En lo referente a los ingresos por turismo provenientes del exterior, es la propia Subdirección de Investigación Económica y Bancaria la que viene realizando una encuesta en los puertos de salida de los visitantes. Para quien observa el desarrollo del muestreo y sus aplicaciones en los últimos años, lo más sorprendente es el rápido aumento del número y tipos de encuestas realizadas por muestreo. A grandes rasgos las encuestas por muestreo pueden clasificarse en dos categorías principales , DESCRIPTIVAS y ANALITICA. En una encuesta descriptiva el único objetivo es obtener cierta información respecto a grandes grupos. En una encuesta analítica se hacen comparaciones entre varios grupos de una población para averiguar si existen ciertas diferencias entre ellos y formular ó verificar hipótesis sobre sus causas.

La distinción entre encuestas analíticas y descriptivas no es perfectamente clara, muchas encuestas proporcionan datos que sirven para ambos propósitos (descriptivo y analítico). Junto con el aumento del número de encuestas descriptivas se ha notado un aumento de encuestas realizadas con fines analíticos, particularmente las relacionadas con el estudio del comportamiento humano y la salud.

Los objetivos de encuesta de turismo están encauzados hacia:

a) La obtención de información necesaria para integrar el renglón de ingresos por turismo de la balanza de pagos.

b) La captación de información adicional para el estudio y análisis del comportamiento del fenómeno.

Es conveniente cerciorarse que todos los datos son pertinentes a la encuesta y que no se omiten datos esenciales. Particularmente en presencia de poblaciones humanas,existe la tendencia a hacer un número excesivo de preguntas que no se analizan posteriormente. Un cuestionario demasiado largo produce una baja general de la calidad de las respuestas,tanto a las preguntas importantes como a las otras.

Los resultados de una encuesta por muestreo están siempre sujetos a cierta incertidumbre porque sólo se mide una parte de la población y por los errores en las mediciones realizadas. Esta falta de certeza se puede reducir al tomar muestras más grandes y emplear mejores dispositivos de medición. Pero ésto suele costar tiempo y dinero, en consecuencia la especificación del grado de precisión deseado,es un paso importante en la preparación de la encuesta.

Para documentar la información sobre ingresos por concepto de turismo, es indispensable captar los datos del gasto realizado en su viaje al interior del país, de los turistas residentes en el exterior, que comprende la suma de gastos por: transporte local, hospedaje, alimentación, diversiones, compras de artículos para uso personal y de regalo, atención médica, etc.. Quedan excluidos los gastos por concepto de compra de mercancía para fines comerciales, inversiones, etc., es decir, todos aquellos gastos que no estén relacionados con la actividad propia del turista.

Respecto a la información adicional para el estudio y análisis del fenómeno, es necesario captar otros datos: la residencia, la permanencia, el nivel de ingreso de los turistas, objetivo del viaje, medio de transporte utilizado, etc.

2.2.1 DELIMITACION DEL CAMPO DE ESTUDIO

La corriente turística está formada por la suma de desplazamientos temporales efectuados por personas que viajan con metas de recreo, negocios, salud ó actividades científicas, deportivas ó religiosas no remuneradas ni lucrativas. Dichos desplazamientos pueden ser nacionales ó internacionales. Los desplazamientos nacionales constituyen el turismo nacional y los desplazamientos internacionales el turismo extranjero. Cabe señalar que las transacciones llevadas a cabo por el turismo extranjero dan lugar a movimientos de divisas que afectan la balanza de pagos, no así, las transacciones realizadas por el turismo nacional.

En el "Manual de la Balanza de Pagos" del Fondo Monetario Internacional se establece que los gastos efectuados por los visitantes residentes en el exterior por motivos de viaje en el país compilador se consideran como créditos en la Cuenta de Viajeros; y los que efectúan los residentes por motivos de viaje en sus visitas al extranjero, se registran como débitos de la misma cuenta. Los gastos por concepto de pasajes internacionales que efectúan los turistas no se computan en esta cuenta, sino en la cuenta de Transportes Diversos. Dentro de la corriente turística que recibe México del exterior se manifiestan 2 tipos de visitantes:

1.- Los que visitan el interior y.
2.- Los que visitan exlcusivamente las ciudades fronterizas.

Las características y el comportamiento de ambos presentan diferencias notables lo que hace necesario que su cómputo se realice separadamente.

INFORMACION CONSEGUIDA PARA ENCUESTAS FUTURAS

Cuanta màs informaciòn se tenga inicialmente, màs fàcil serà el diseño de una muestra que proporcione estimaciones exactas. Toda muestra obtenida es un guía potencial de futuros muestreos, por los datos que revela sobre las medias, las desviaciones estàndar, la naturaleza de la variabilidad de las medidas principales, así como sobre los costos de obtenciòn de datos. La pràctica de muestreo avanzarà màs rapidamente si se prevê lo necesario para reunir y registrar este tipo de informaciòn.

POBLACION BAJO MUESTREO

Actualmente existe una gran variedad de planes para seleccionar una muestra. Por cada plan considerado se pueden hacer, a groso modo,estimaciones del tamaño de la muestra partiendo del conocimiento del nivel deseado.

La poblaciòn muestreada debe coincidir con la poblaciòn sobre la cual se desea informaciòn (la poblaciòn objetivo). En ocasiones, por razones de factibilidad ò simple conveniencia, la poblaciòn muestreada es màs restringida que la poblaciòn objetivo. De ser así debe recordarse que las conlusiones extraídas de la muestra son aplicables a la poblaciòn muestreada, y habrà que recurrir a las fuentes de informaciòn para decidir hasta que grado se aplican estas conclusiones a la poblaciòn objetivo. Toda la informaciòn que se obtenga respecto a las diferencias entre ambas poblaciones serà de utilidad.

Para cuantificar los ingresos que cada país recibe del exterior por concepto de turismo al interior, la poblaciòn està formada por el conjunto de turistas residentes en el exterior que visitan el interior del país.

2.2.2 METODO DE CAPTACION

Dada la naturaleza de la informaciòn requerida se determinò que èsta debe obtenerse del turista ("sujeto de turismo") atravès de la entrevista directa al momento de abandonar el país porque su posibilidad de gastar en el interior del mismo ha terminado. La figura 2.1 muestra el cuestionario utilizado.

El tratar de captar las características de la corriente turística en el total de las plazas de salida del país es un procedimiento que implica un alto costo y considerables recursos. Toda informaciòn, debe reunir requisitos de calidad, oportunidad y bajo costo. De los mètodos de producciòn de informaciòn se concluyò que el de muestreo es el que mejor se adapta a los requisitos anteriores. Efectivamente, el tomar una muestra de una poblaciòn con ciertos requerimientos probabilísticos resulta ser econòmico, muy opotuno y ademàs con una precisiòn controlada que puede llegar a ser tan buena como la de un

CUESTIONARIO UTILIZADO

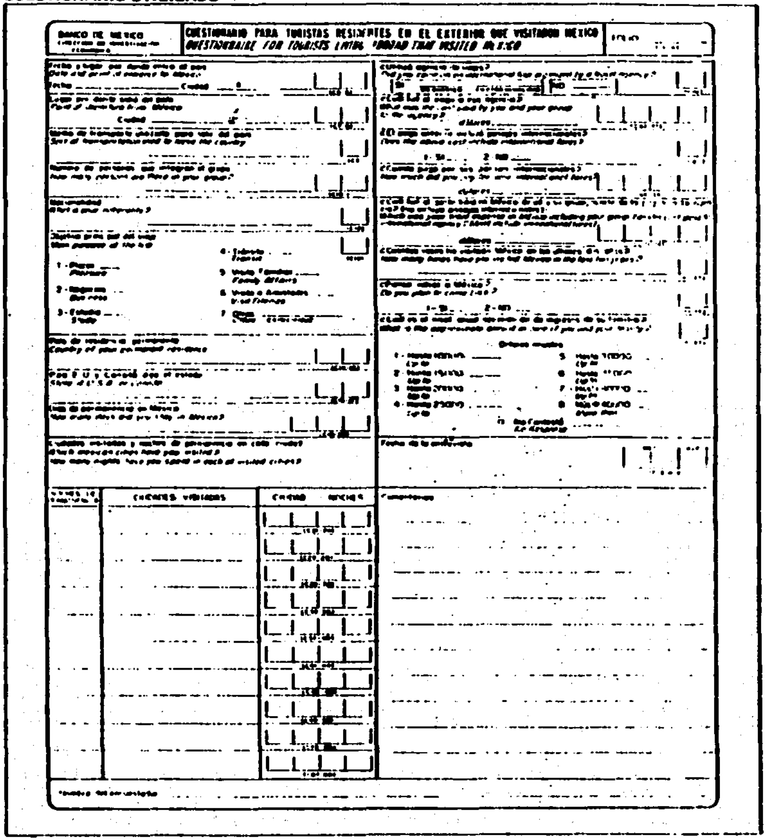


FIGURA 2.1

censo ó registro contable.

En el muestreo simple aleatorio la varianza del estimador depende, además del tamaño de la muestra, de la variabilidad de "Y" en la población. Si la población es muy heterogénea y las consideraciones de costo limitan el tamaño de la muestra, podría ser imposible obtener un estimador lo sificientemente preciso tomando una muestra simple aleatoria de toda la población. Si la población no es homogénea cualquier estimación hecha en base a una muestra aleatoria directa estará sometida a fluctuaciones muy grandes de muestreo.

Supóngase que la población puede dividirse en partes ó ESTRATOS, donde los elementos de cada estrato tienen las mismas características; y se hace ahora un muestreo aleatorio de cada estrato.

Deberá ser posible obtener una mejor estimación de los promedios de los estratos lo que a su vez ayudará a obtener una mejor estimación del promedio de la población.

Resumiendo lo anterior, si puede encontrarse una forma de subdividir la población de tal manera que se reduzca considerablemente la variación de y con respecto a x puede hacerse una mejor estimación de la población total.

De acuerdo a las consideraciones anteriores y a que es posible dividir una población es subpoblaciones, cuyas varianzas sean menores a la varianza de la población original, se puede emplear una técnica de MUESTREO ESTRATIFICADO MULTIETAPICO(4), para obtener una estimación más precisa de la media de la población que la media muestral de una muestra aleatoria.

En el muestreo estratificado, la población de N unidades se divide primero en subpoblaciones de N_1 ,N_2 ,...,N_L unidades respectivamente. Estas subpoblaciones no se traslapan y en su conjunto comprenden a toda la población

$$N_1 + N_2 + \dots + N_L = N$$

Estas subpoblaciones se denominan ESTRATOS. Para obtener todo el beneficio de la estratificación, los valores de los Nh deben ser conocidos. Una vez determinados los estratos, se extrae una muestra de cada uno, las extracciones deben hacerse independientemente en los diferentes estratos.

Esta técnica está orientada a obtener una selección con igual probabilidad (MESIP), con una fracción general fija e igual a 1/100.

FORMACION DE ESTRATOS

La consideración básica implicada en la formación de estratos es que éstos deden ser HOMOGENEOS INTERNAMENTE. Si se van a seleccionar

(4) Para mayor información consultar referencia (6)

unidades al azar dentro de los estratos, la varianza de éstos para el rango que se está estimando debe ser tan pequeña como sea posible. Esto puede lograrse asignando a un mismo estrato unidades que se cree son similares. De este modo puede utilizarse cualquier conocimiento previo, intuición personal y juicio para producir semejanza dentro de un mismo estrato. La situación ideal es aquella en la que se tiene disponible la distribución de Y ; entonces se crearán los estratos estableciendo los límites de la distribución en los puntos adecuados. Si se carece de esta información se buscará la distribución de Y tal como se obtiene de un censo reciente ó la de un rasgo de X, estrechamente correlacionado con Y.

EL MARCO MUESTRAL

Antes de seleccionar la muestra, la población debe ser dividida en partes llamadas "unidades de muestreo ó unidades", éstas deben cubrir la totalidad de la población y no traslaparse en el sentido de que todo elemento de la población pertenezca a una y solamente una unidad. Algunas veces la unidad apropiada es obvia, en otras ocasiones existe la posibilidad de escoger lo que será la unidad de muestreo.

El marco muestral ó de referencia está integrado por el conjunto de plazas de salida del país, es decir, las ciudades fronterizas y los aeropuertos internacionales.

Los estratos establecidos fueron catorce (1978)(nueve para vía aérea y cinco para vía terrestre) y a cada estrato le corresponde una ó más plazas de salida.

Las etapas y unidades de muestreo que se adoptaron por cada vía de transporte fueron las siguientes:

ETAPAS	PRIMERA	SEGUNDA	TERCERA	CUARTA		QUINTA	ULTIMA
UNIDADES	PLAZAS	DIAS	TURNOS	VUELOS	HORAS	MINUTOS	G. TURIS
AEREA	X	X	X	X			X
TERRESTRE	X	X	X		X	X	X

tabla 2a

SELECCION DE LAS UNIDADES DE MUESTREO

En la primera etapa se seleccionaron 14 de las 78 unidades primarias de muestreo (UPM) en la siguiente forma:

Se eligieron 5 UPM que por su importancia tuvieron una probabilidad de selección en primera etapa igual a la unidad, a las que se denominó: unidades primarias autorrepresentadas (UPAR). Estas plazas formaron, cada una por separado, un estrato independiente.

Las 73 plazas restantes fueron agrupadas en 9 estratos tomando en consideración por orden de importancia, las siguientes variables: localización geográfica, tipo de unidad primaria (aérea ó terrestre), y medida estimada de tamaño (MT) de cada unidad, representada por el número de turistas que reporta cada plaza.

Dentro de cada uno de estos estratos, se seleccionó una UPM con probabilidad de cada UPM respecto a su estrato. Las plazas resultantes para cada estrato fueron las siguientes:

ESTRATO	TIPO	NOMBRE DE LA U.P	PROBABILIDAD DE SELECCION EN LA PRIMERA ETAPA
I	A	Acapulco, Gro.	1.00 Autorrepresentada
II	A	Cancún, Q.R.	1.00 Autorrepresentada
III	A	Distrito Federal	1.00 Autorrepresentada
IV	A	Guadalajara, Jal.	1.00 Autorrepresentada
V	T	Nvo. Laredo, Tam.	1.00 Autorrepresentada
VI	T	Cd.Juárez Chih.	.5106
VII	T	Nogales, Sonora.	.3613
VIII	T	Reynosa, Tamaul.	.5147
IX	T	Talismán, Chis.	.3635
X	A	Pto Vallarta,jal.	.8427
XI	A	Mazatlán Sinal.	.6459
XII	A	Mérida Yucatan.	.3387
XIII	A	Monterrey Nuev.L.	.9522
XIV	A	Hermosillo Sonor.	.7390

Una vez formados los 14 estratos y seleccionadas las 14 plazas o unidades primarias, se prosiguió a elegir, manteniendo la estructura apropiada de probabilidades, las unidades de segunda etapa en adelante. Esto se realizó, seleccionando en las plazas terrestres: días del mes en los que se llevan a cabo las entrevistas; turnos en cada día; horas en cada turno, minutos en cada hora, hasta llegar a la última etapa en la que se selecciona la unidad básica de muestreo ó "unidad última de muestreo" (UUM), que es el grupo turístico. En las plazas aéreas el procedimiento varía a partir de la cuarta etapa, en donde una vez distinguidos los días y turnos del levantamiento, se listan los vuelos de salida y se seleccionan con igual probabilidad. Una vez determinados los vuelos se lleva a cabo la última etapa que consiste en seleccionar las unidades de muestreo, a partir de un MUESTREO SISTEMATICO(5). Supóngase que las N unidades de la población se numeran de 1 a N en cierto orden. Para elegir una muestra de n unidades, se toma una unidad al azar entre las k primeras y luego se toman las subsecuentes a intervalos de k. La selección de la primera unidad determina toda la muestra que se denomina MUESTRA DE TODAS LAS k-ésimas UNIDADES.

Se puede esperar que la muestra sistemática sea tan precisa como la muestra aleatoria estratificada correspondiente con UNA unidad por estrato. La diferencia es que con la muestra sistemática, las unidades ocurren en la misma posición relativa del estrato, mientras que con el muestreo aleatorio estratificado, la posición dentro del estrato se determina separadamente por aleatorización dentro de cada estrato. La muestra sistemática se reparte más uniformemente

La unidad básica de muestreo es el grupo turístico que puede consistir en una ó más personas y tiene como característica la condición de construir una unidad de gasto. Así dos familiares que viajan juntos y hacen sus gastos separadamente se consideran dos grupos turísticos distintos. Las entrevistas a los grupos turísticos se efectúan por encuestadores adiestrados en la diferentes plazas seleccionadas y están sujetas a programas de trabajo de campo. El programa de trabajo se establece en forma precisa para no dar margen al encuestador de aplicar las encuestas a su criterio, lo que por otra parte facilita la supervisión del trabajo.

RESUMEN Y ANALISIS DE LOS DATOS

El primer paso después de realizar la encuesta es el editar los cuestionarios obtenidos con la esperanza de corregir errores ó cuando menos desechar los datos que obviamente están equivocados. Habrá que tomar ciertas decisiones respecto al procedimiento de cálculo en los casos de omisión de respuesta de quienes responden ó de eliminación de datos en el proceso de edición. después se realizarán los cálculos que conduzcan a las estimaciones.

2.2.3 INFERENCIA DEL UNIVERSO

Dado el diseño de la encuesta, se requiere información complementaria del censo de turistas residentes en el exterior que visitan el interior del país con el fin de hacer inferencia al universo y así obtener el monto total de los ingresos por turismo. La información deberá estar desglozada de la siguiente manera:

(5) referencia (6)

NUMERO MENSUAL DE TURISTAS RESIDENTES EN EL EXTERIOR QUE VISITARON EL INTERIOR DEL PAIS, POR PLAZA DE REGISTRO SEGÚN EXTRANJEROS Y NACIONALES.

PLAZA	EXTRANJEROS	NACIONALES	TOTAL
Acapulco	X	X	X
Agua Prieta	X	X	X
.	.	.	.
.	.	.	.
.	.	.	.
Zihuatanejo	X	X	X
T O T A L	X	X	X

FUENTES: Dirección de Servicios Migratorios, Secretaría de Gobernación, Dirección General de Estadística, Secretaría de Programación y Presupuesto.

ESTIMADORES SIMPLES INSESGADOS Y SU ERROR DE MUESTREO

Una de las ventajas más importantes del diseño probabilístico que se ha utilizado en este caso en particular, es el trabajar con estimadores insesgados de una sencillez excepcional para su cálculo.

Los estimadores insesgados tienen la propiedad de que el valor esperado ó el valor medio de muestras repetidas sea igual al valor que se obtendría aplicando los procedimientos de recopilación de información al total de la población muestral; únicamente podrán obtenerse bajo la base de un muestreo probabilístico.

La muestra de turismo no llega a ser totalmente una muestra probabilística de la población delimitada por la encuesta, èsto es debido a que existen problemas tales como: no respuesta; errores de conteo de los entrevistadores, etc., que hacen imposible determinar exactamente la probabilidad de selección de todas las unidades de la población.

Sin embargo, una encuesta como ésta en donde todas las acciones se han encaminado a mantener mínimo el distanciamiento de un muestro probabilístico, es fundamentalmente diferente a una operación de recopilar información en donde se tolera el uso del muestreo de juicio o por cuota. En el primer caso es posible, con cierto riesgo, interpretar los resultados bajo el supuesto que provienen de una muestra probabilística; en el segundo caso no existe una base firme para evaluar la precisión de los estimadores.

VARIANZA DE LOS ESTIMADORES

Los errores muestrales en los estratos que tienen unicamente una unidad primaria de muestreo (UPM) y los de dos ó más, se calculan en forma separada. Las varianzas estimadas de los dos grupos de estratos se suman adecuadamente para proporcionar un estimador total para la encuesta.

ESTIMADORES INSESGADOS DE VARIANZA MINIMA

Sea $x_1, x_2, \ldots, x_n$ una muestra aleatoria de $f(x;\theta)$ y $\hat{\theta}=d(x_1, \ldots, x_n)$ un estimador de θ tal que:

a) $E(\hat{\theta}) = \theta_0$; i.e, $\hat{\theta}$ es insesgado

b) Var $(\hat{\theta})$ es menor que la varianza de cualquier otro estiamdor isesgado.

En tales condiciones, $\hat{\theta}$ es el estimador insesgado de la varianza mínima de θ.

VARIANZA ESTIMADA PARA LOS ESTRATOS CON UPM NO AUTORREPRESENTADA.

El método para la estimación de los errores de muestreo en los estratos que contienen dos ó más UPM, hace uso del hecho de que la varianza de un estimador basado en una muestra compleja multietápica puede, bajo ciertas condiciones, ser estimada bajo un proceso relativamente sencillo en el cual la información requerida para los cálculos, son los totales muestrales de las unidades denominadas "conglomerados últimos". El "conglomerado último" es el conjunto de unidades incluidas en la muestra de una unidad primaria.

El estimador de la contribución de la varianza de las UPM no autorrepresentadas es complicado debido a que únicamente se selecciona una UPM por estrato. Esta dificultad se supera formando grupos , por ejemplo 4, tres de ellos con 2 estratos y el otro con 3 estratos, manteniendo cada grupo tan homogéneo como sea posible. El criterio para conformar grupos homogeneos es el mismo que el usado en los estratos originales. Los grupos se arreglan bajo la base de las características del estrato completo y sin el conocimiento de cual UPM fue seleccionada en cada estrato.

Mediante la modificación anterior se obtiene el estimador de la varianza aplicada, el cual sería el apropiado si todas las UPM de la muestra de un grupo fueran seleccionadas del mismo estrato. Este método es conocido como LA TECNICA DE ESTRATOS AGRUPADOS O ESTRATOS COLAPSADOS, resultando una ligera sobreestimación en la varianza. La estimación mediante estratos colapsados de la varianza relativa de un estimador insesgado del total para los estratos no autorrepresentados está dada por:

$$V^2_{NAR} = \frac{1}{Y^2_{NAR}} \sum_{g=1}^{4} \frac{Lg}{Lg-1} \sum_{h}^{Lg} \left(Y_{gh} - \frac{MT}{MTg} Y_g \right)^2$$

en donde:

$$V^2 Y_{NAR} = CV^2 Y_{NAR} = CV^2(Y_{NAR}) = \frac{var(Y_{NAR})}{Y^2_{NAR}}.$$

$CV\, Y_{NAR}$ = Coeficiente de variación de Y_{NAR}

Ygh = Estimación del total, para el estrato no autorrepresentado h en el grupo de estratos g de la variable en estudio Y.

MTgh = Medida de tamaño estimada del estrato no autorrepresentado h en el grupo de estratos colapsados g.

Lg = Número de estratos en el g-ésimo grupo de estratos. g={1,...,4}

Yg = $\sum_{h}^{Lg}$ Ygh = Estimacion del total, de los estratos no autorrepresentados pertenecientes al grupo g.

MTg = $\sum_{h}^{Lg}$ MTgh = Medida estimada de tamaño del grupo g.

Y = $\sum_{g=1}^{4}$ Yg = Estimación del total, para los estratos no autorrepresentados.

Varianza ESTIMADA PARA LOS ESTRATOS CON UPM AUTORREPRESENTADAS.

Para las UPM autorrepresentadas se utiliza el método denominado "grupo aleatorio" para estimar los errores del muestreo. Este consiste en dividir la muestra en submuestras aleatorias y estimar la varianza a partir de las desviaciones cuadráticas de los estimadores de estas submuestras, respecto a la media.

La estimación por este método de la varianza relativa de un estimador insesgado del total para los estratos autorrepresentados está dada por:

$$V^2_{NAR} = \frac{1}{Y_{AR}} \frac{1}{d-1} \sum_{h} \left[d \sum_{j}^{d} Y_{jh} - \left(\sum_{j}^{d} Y_{jh} \right)^2 \right]$$

en donde:

Vjh = es el estimador del total de la variable en estudio, para el j-ésimo día en estrato h.

d = Número de unidades de 2a. etapa en muestra.

Y = $\sum_{h}^{5} \sum^{d}$ Vjh = Estimador del total de los estratos autorrepresentados.

ESTIMADORES DE VARIANZA DEL TOTAL POBLACIONAL.

Para estimar la varianza del total para cualquier variable, se requiere 9 totales de UPM (para los estratos con UPM no autorrepresentadas) y 75 totales diarios (d=5)para los estratos con UPM autorrepresentadas.

La varianza relativa combinada del estimador simple insesgado para el total poblacional se calcula de la siguiente forma:

$$V^2Y = \frac{Y_{AR}^2 \, V^2Y_{AR} + Y_{NAR}^2 \, V^2Y_{NAR}}{(Y_{AR} + Y_{NAR})^2}$$

2.2.4 ERRORES QUE NO SON DE MUESTREO

Es fácil encontrar ejemplos para mostrar que cuando se realiza una encuesta hay errores de medida, de observación ó de respuesta. Además de estar sujetas a los errores de respuesta, las encuestas pueden estar sometidas a errores de cubrimiento, de procesamiento, etc.

TRATAMIENTO DE LA NO RESPUESTA

La no respuesta se presenta con certeza en casi todas las encuestas referentes a poblaciones humanas. No se conoce hasta la fecha un método insesgado ó al menos consistente para ajustar la no respuesta por tanto las magnitudes del sesgo introducido por el ajuste de la no respuesta será desconocido.

El procedimiento para ajustar la no respuesta, en el caso de utilizar los estimadores en una muestra autoponderada, es el siguiente:

1. El entrevistador debe asegurarse si el grupo que no responda pertenece o no a la población. Si el grupo pertenece a la población será un caso de no respuesta y es importante clasificarlo como tal, para ésto es conveniente llenar un cuestionario con una clave especial de no respuesta.

Si el grupo no pertenece a la población será un caso que se denominará "blanco" y no es necesario tomarlo en consideración.

2. Para cada unidad secundaria se calculará el cociente:

$$r = \frac{\text{Grupos entrevistados + grupos no entrevistados}}{\text{grupos entrevistados}}$$

donde: grupos entrevistados= grupos con no respuesta.

Este cociente se aplicará a la totalidad de la información muestral correspondiente a los grupos entrevistados dentro de cada unidad secundaria, excepto en las unidades de la segunda etapa. En donde el cociente exceda de 1.5 es recomendable que el proceso de agrupación sea mayor, es decir que se trabaje a nivel de unidad primaria.

ERRORES DE RESPUESTA

Es útil suponer que hay un valor verdadero yi, que corresponde a la unidad Ui en la población. Cuando un entrevistador se dirige a la unidad para obetener información sobre algún "item" se supone que la respuesta que obtiene es una observación sobre una variable aleatoria con cierta distribución. Diferentes entrevistadores producirán diferentes distribuciones, dependiendo de su habilidad, la interacción entre el entrevistador y el informante y así sucesivamente. Cuando se llega al caso de que dos unidades diferentes son entrevistadas por la misma persona no se puede suponer que las respuestas obtenidas no estén correlacionadas. Otro punto que hay que recordar es que la distribución de las respuestas producidas por un entrevistador dependerá de lo que se llama las CONDICIONES ESENCIALES de la encuesta. En una encuesta muy detallada que cuente con recursos considerables en la cual se presta gran atención a los problemas de capacitación, entrevistas, etc., la distribución será diferente de una en la cual todo lo que se considera importante es elaborar un cuestionario y ordenar a algunas personas que obtengan datos en un momento dado. De este modo al hablar de las variables aleatorias de que se trate, se debe tener siempre en mente las condiciones esenciales de la encuesta que determinan eas distribuciones.

SESGO EN LAS RESPUESTAS

La razón de tratar al entrevistador para el estudio de los errores de respuesta, es que las encuestas modernas, en gran escala, son realizadas por lo general, con ayuda de entrevistadores capacitados especialmente para ese fin, con el objeto de obtener buenos resultados. Supongase que para la encuesta está disponible un gran número de M entrevistadores. La respuesta obtenida por el entrevistador de la unidad j es una variable aleatoria que tiene una distribución

$$E\,(X_{rjk}) = \overline{X}_{rjk} \;\; y \;\; V_k\,(X_{rjk}) = S^2_{rj}\,.$$

El promedio de respuestas obtenidas por el entrevistador i sobre todas las unidades N en la población será

$$\sum_{j} \frac{X_{ij}}{N}$$

y el promedio obtenido por todos los entrevistadores, M, disponibles para la encuesta será

$$\sum_{i} \frac{\overline{X}_i}{M} = \overline{X}$$

A éste se le puede llamar valor esperado de la encuesta, y el valor verdadero será

$$\sum_{i} \frac{X_i}{N} = \overline{Y}$$

El objetivo es estimar Y y por lo tanto la diferencia X-Y, entre el valor esperado de la encuesta y el valor verdadero es llamada el SESGO de la respuesta. El sesgo de la respuesta dependerá obviamente de los procedimientos de la entrevista, el cuestionario y la capacitación del personal. A menos que puedan crearse procedimientos adecuados que garanticen un pequeño sesgo de respuesta, no valdrá la pena continuar con la encuesta.

Como la respuesta dependerá de quién entreviste a quién, debe haber procedimientos adecuados para hacer aleatoria la asignación de los entrevistadores para la muestra (seleccionados entre los M disponibles) entre las unidades de la muestra (seleccionadas entre las N unidades de la población).

2.2.5 SERIE HISTORICA DE DATOS

Todo lo explicado anteriormente se resume en cifras que son precisamente los datos de interès con los que se trabajarà en esta tesis. La informaciòn presentada en este capìtulo asì como estos datos fueron proporcionados por la Biblioteca y Hemeroteca del Banco de Mèxico. Las tablas (a),(b),(c),(d), (e) presentan toda esta informaciòn. La tabla (f) presenta la serie històrica de ingresos por turismo que se obtuvo de multiplicar la columna de Nòmero Total de Turistas(miles) por la columna correspondiente al gasto medio total (en dòlares).

OBSERVACIONES DE INGRESOS POR TURISMO

TURISMO RECEPTIVO *

Continuación...

Meses	Número de turistas 1/ ([illegible]) Total	Vía Aérea	Vía Terrestre	[illegible] Total	Vía Aérea	Vía Terrestre	[illegible] Total	Vía Aérea	Vía Terrestre	[illegible] Total	Vía Aérea	Vía Terrestre
1976												
Enero	232	124	117	300.10	380.00	191.61	24.01	38.71	14.38	11.8	10.4	13.4
Febrero	270	150	111	317.13	407.75	187.05	25.00	38.10	12.87	12.2	10.4	14.8
Marzo	267	142	125	304.36	380.57	200.52	24.58	38.15	13.94	12.4	10.2	14.8
Abril	210	134	144	272.35	354.81	165.50	22.00	37.00	17.53	17.3	9.3	15.1
Mayo	236	106	120	279.30	380.32	194.10	23.05	38.52	14.56	11.5	9.4	13.3
Junio	256	112	142	281.67	374.75	171.31	26.28	38.92	17.94	8.9	10.2	9.8
Julio	288	125	163	290.07	380.81	170.80	23.87	36.05	17.51	10.1	10.0	10.1
Agosto	258	116	143	283.17	374.57	172.50	24.28	34.81	15.00	10.8	10.8	10.8
Septiembre	175	80	85	236.77	347.40	143.40	22.47	30.40	14.82	10.5	11.4	9.8
Octubre	231	123	107	252.36	346.10	142.06	24.94	35.38	12.00	10.2	9.6	11.0
Noviembre	[illegible]	127	111	226.76	312.06	133.77	22.73	33.87	12.10	10.1	9.7	11.1
Diciembre	[illegible]	160	160	253.84	353.50	161.10	19.76	32.30	11.33	12.8	10.0	14.5
1977												
Enero	236	140	98	380.21	380.88	176.65	25.30	35.00	12.40	12.1	10.8	14.1
Febrero	260	170	90	387.40	380.81	144.78	28.71	35.06	12.43	10.0	9.1	11.8
Marzo	260	160	101	387.82	373.97	172.14	29.94	38.11	12.73	11.2	9.8	13.5
Abril	280	157	131	277.30	381.80	148.09	23.47	40.77	11.76	10.9	9.4	12.0
Mayo	222	108	113	251.73	381.36	145.00	28.51	40.17	14.65	9.5	9.8	10.0
Junio	257	128	128	240.11	328.00	171.15	22.27	[illegible]	14.24	11.1	10.1	12.0
Julio	320	151	177	363.70	366.53	144.06	23.00	[illegible]	14.00	10.2	10.0	10.3
Agosto	280	143	148	[illegible]	352.21	176.00	24.00	32.51	13.30	11.0	10.5	11.5
Septiembre	197	98	100	233.34	323.75	144.40	24.84	38.80	14.30	9.5	8.8	10.1
Octubre	230	122	113	380.30	370.12	140.00	30.14	40.47	15.00	9.2	9.1	9.4
Noviembre	267	146	116	253.00	342.90	142.00	28.00	38.00	12.01	9.5	8.8	10.2
Diciembre	401	182	200	272.30	375.60	176.01	26.04	41.00	14.75	10.5	9.1	11.8
1978												
Enero	300	187	104	301.30	418.36	176.00	30.30	44.11	17.40	11.3	9.4	14.2
Febrero	312	211	100	303.75	401.30	160.34	31.58	44.00	12.00	11.2	8.8	14.6
Marzo	365	216	148	361.94	421.50	225.00	30.75	41.00	13.40	12.8	10.0	15.0
Abril	277	167	160	304.10	380.57	200.27	30.70	[illegible]	12.00	11.6	8.3	16.1
Mayo	270	140	132	275.00	380.40	187.37	30.30	42.00	14.00	9.7	8.7	10.9
Junio	300	150	140	232.16	330.00	135.33	34.32	38.00	13.36	8.7	8.3	10.2
Julio	360	190	162	300.46	350.00	180.00	27.00	38.40	17.43	10.1	9.0	10.0
Agosto	300	187	140	277.24	370.00	140.00	25.00	38.21	13.07	10.7	10.7	10.7
Septiembre	237	135	101	273.10	360.00	161.00	30.74	37.07	14.30	10.2	9.4	11.3
Octubre	300	140	160	300.77	414.36	132.77	31.30	40.00	14.00	9.4	8.7	8.8
Noviembre	310	160	130	300.15	381.30	148.00	30.00	46.00	13.77	9.3	8.4	10.0
Diciembre	440	230	211	307.00	401.53	158.07	30.30	44.94	13.00	10.5	10.1	11.0

* [illegible]
1/ [illegible]

Septiembre 1984

IV-[illegible]

figura a

TURISMO RECEPTIVO *

Continuación...

Meses	Número de turistas 1/ ([illegible]) Total	Vía Aérea	Vía Terrestre	[illegible] Total	Vía Aérea	Vía Terrestre	[illegible] Total	Vía Aérea	Vía Terrestre	[illegible] Total	Vía Aérea	Vía Terrestre
1979												
Enero	330	220	111	352.00	444.00	171.30	33.00	47.91	13.07	10.4	9.4	12.4
Febrero	357	247	110	372.07	418.81	177.20	33.22	48.50	13.37	10.6	9.4	13.1
Marzo	377	247	130	370.10	475.29	168.31	36.30	51.25	13.27	10.5	9.3	12.7
Abril	357	215	142	348.25	485.00	171.03	32.81	48.02	13.51	10.8	9.7	12.7
Mayo	304	183	119	327.40	420.30	160.87	33.02	51.15	13.00	9.8	8.2	12.3
Junio	351	199	152	317.14	470.81	157.73	33.20	49.72	15.79	9.5	9.0	10.3
Julio	360	205	153	332.36	452.77	171.97	36.04	48.22	18.30	9.3	9.4	9.3
Agosto	373	218	156	341.92	449.36	190.03	34.53	48.00	17.40	9.9	9.2	10.9
Septiembre	245	146	98	325.00	447.36	170.00	34.50	38.51	15.43	9.7	8.9	11.3
Octubre	290	172	117	340.44	473.30	150.94	36.01	38.30	15.66	9.2	8.8	10.2
Noviembre	334	188	125	350.34	485.47	162.10	37.63	38.26	16.00	9.3	8.8	10.1
Diciembre	466	240	210	370.43	516.19	221.01	32.32	51.21	16.36	11.7	10.1	13.8
1980												
Enero	363	240	114	432.50	523.16	213.30	36.01	50.40	15.77	11.1	10.8	13.5
Febrero	374	262	112	448.30	572.10	227.16	44.00	[illegible]	7.63	10.4	9.4	12.0
Marzo	407	264	142	432.03	543.00	230.42	39.30	57.30	13.00	11.9	9.5	16.4
Abril	321	191	130	416.00	526.25	253.40	36.36	56.00	15.00	12.1	9.1	18.8
Mayo	300	187	130	390.71	504.73	180.30	41.00	[illegible]	18.57	9.8	8.2	10.0
Junio	336	188	148	372.40	522.17	184.21	41.40	61.15	18.34	9.0	8.5	7.5
Julio	362	196	136	383.46	503.00	187.00	38.07	[illegible]	19.00	9.3	8.9	8.8
Agosto	363	202	160	386.72	504.51	194.31	38.00	[illegible]	20.00	9.4	8.5	9.3
Septiembre	230	138	98	416.70	564.42	222.01	40.76	[illegible]	17.00	10.2	8.8	12.3
Octubre	270	164	116	413.00	543.05	190.00	40.17	[illegible]	17.00	9.8	8.4	8.8
Noviembre	327	194	132	414.17	566.01	207.30	43.00	[illegible]	17.00	9.6	8.1	11.0
Diciembre	466	250	217	376.70	516.77	236.17	[illegible]	[illegible]	16.70	10.4	7.8	13.1
1981 p/												
Enero	340	210	130	407.00	643.00	237.57	40.07	77.90	17.20	10.0	8.1	13.2
Febrero	381	262	119	513.30	634.63	227.11	50.15	73.20	17.30	13.2	[illegible]	12.9
Marzo	360	232	124	486.46	581.21	245.41	48.61	66.70	16.76	10.9	9.8	14.8
Abril	350	196	150	430.00	581.30	237.00	40.30	[illegible]	14.51	12.0	[illegible]	14.3
Mayo	316	181	137	425.07	447.40	211.13	45.36	72.30	19.17	9.4	8.1	11.0
Junio	322	187	140	415.30	567.00	230.03	42.61	[illegible]	14.00	9.7	4.3	11.7
Julio	360	190	162	387.40	500.70	217.00	38.00	[illegible]	10.81	8.9	8.4	11.7
Agosto	330	180	153	383.70	515.70	225.70	38.07	[illegible]	10.30	10.4	8.7	12.3
Septiembre	230	130	100	412.00	567.73	225.60	38.00	67.74	16.00	10.6	8.4	13.6
Octubre	261	136	126	396.36	564.23	200.32	44.30	74.00	15.00	8.8	7.5	10.7
Noviembre	317	160	150	423.30	579.42	190.00	47.00	73.00	15.00	8.8	7.8	10.4
Diciembre	464	230	234	440.40	600.40	264.00	[illegible]	[illegible]	[illegible]	11.8	9.8	14.8

* [illegible]
1/ [illegible]
p/ [illegible]

IV-[illegible]

Septiembre 1984

figura b

TURISMO RECEPTIVO *

Continuación ..

AÑOS	NÚMERO DE TURISTAS 1/ (miles)			GASTO MEDIO (dólares)			GASTO MEDIO DIARIO (dólares)			PERMANENCIA MEDIA (días)		
	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre
[illegible]	3 787	2 172	1 [illegible]	373.15	518.[illegible]	177.79	[illegible].16	[illegible].04	14 28	10.3	8.0	12.9
Enero	320	203	117	517.[illegible]	[illegible]	207.[illegible]	48.04	73.40	18.43	10.6	9.5	12.7
Febrero	333	220	113	490.53	632.53	220.77	43.50	[illegible]	15.01	11.4	9.6	15.0
Marzo	344	213	131	474.23	620.04	218.55	41.00	68.65	14.51	11.4	9.2	14.0
Abril	328	185	146	380.78	551.52	180.58	30.41	60.95	14.43	10.2	8.0	12.9
Mayo	290	160	130	330.57	480.41	174.07	36.50	64.11	14.45	9.6	7.0	12.1
Junio	300	167	133	314.00	423.19	170.86	33.06	53.20	15.97	9.3	7.9	11.1
Julio	349	188	160	340.29	483.30	182.19	33.52	56.00	14.62	10.4	8.7	12.5
Agosto	319	180	139	323.40	435.25	175.00	31.44	48.06	14.60	10.3	9.1	12.0
Septiembre	222	126	96	295.46	387.07	173.44	30.07	47.14	14.53	9.8	8.2	11.0
Octubre	257	146	111	300.25	420.19	128.22	32.03	53.04	12.22	9.5	8.1	11.3
Noviembre	272	150	122	320.40	477.01	140.50	31.53	54.04	11.70	10.4	8.8	12.4
Diciembre	470	232	267	295.84	441.79	130.20	28.32	47.27	11.80	10.4	9.3	11.6
[illegible]	4 748	2 990	1 757	342.08	468.90	131.15	37.05	54.40	12.01	9.2	8.6	10.4
Enero	350	240	110	344.12	440.79	127.90	37.93	54.97	11.84	9.1	8.2	10.8
Febrero	385	270	119	371.75	470.63	142.96	30.36	54.70	11.63	9.7	8.6	12.3
Marzo	486	313	173	240.46	458.84	143.79	36.55	54.71	11.80	9.7	8.4	12.2
Abril	393	251	142	320.54	422.04	157.40	32.50	40.[illegible]	12.47	10.0	8.8	12.0
Mayo	371	228	143	300.21	410.00	136.46	32.58	30.40	11.01	9.5	8.3	11.4
Junio	402	237	166	230.90	432.00	108.22	30.72	56.03	12.33	8.2	7.7	8.8
Julio	490	286	184	332.00	480.00	120.05	34.96	51.31	12.30	9.5	9.4	9.7
Agosto	383	237	146	321.23	436.50	130.50	33.96	40.97	13.73	9.5	9.3	9.7
Septiembre	281	167	114	314.92	447.62	120.46	37.34	55.36	13.41	8.4	8.1	9.0
Octubre	332	220	112	358.74	472.30	134.70	41.71	50.53	13.01	8.6	8.1	9.7
Noviembre	382	253	129	363.53	488.22	110.72	45.18	60.62	14.70	8.0	8.1	9.3
Diciembre	517	303	214	393.58	577.30	133.51	39.90	60.40	13.02	5.8	3.8	10.[illegible]
[illegible]	4 655	3 [illegible]	1 [illegible]	410.57	543.78	193.05	44.14	44.79	[illegible]	9.5	[illegible]	[illegible]
Enero	398	285	113	465.15	588.63	158.43	33.04	71.77	[illegible]	8.8	4.2	10.3
Febrero	447	325	122	513.21	644.11	163.80	46.68	64.43	11.00	11.0	[illegible]	[illegible]
Marzo	507	340	150	440.36	550.77	203.62	44.78	47.37	16.42	10.1	8.3	14.1
Abril	392	246	146	437.34	560.92	220.45	40.01	71.41	16.78	10.7	7.8	15.5
Mayo	358	234	124	425.14	537.13	214.35	40.71	77.43	17.72	8.7	8.0	12.1
Junio	371	220	151	400.00	555.91	188.81	42.92	71.20	15.72	9.5	7.8	11.0
Julio	378	220	150	384.21	517.50	188.28	42.55	50.20	20.31	9.0	8.9	9.3
Agosto	399	225	144	335.46	518.60	179.07	43.61	53.42	13.75	7.0	4.8	9.3
Septiembre	251	154	97	394.01	516.25	173.91	14.68	41.08	19.71	8.6	8.5	8.8
Octubre	290	150	106	316.63	408.54	211.55	39.41	55.64	18.24	7.6	8.4	11.6
Noviembre	375	246	129	378.38	478.25	191.35	43.78	60.63	18.32	9.5	7.9	10.5
Diciembre	520	307	212	380.79	518.27	207.34	40.12	50.38	18.91	8.7	8.8	11.5

* [illegible]
1/ [illegible]

figura c

TURISMO RECEPTIVO *

continuación . . .

AÑOS TRIMESTRES Y MESES	NÚMERO DE TURISTAS 1/ (miles)			GASTO MEDIO (dólares)			GASTO MEDIO DIARIO (dólares)			PERMANENCIA MEDIA (días)		
	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre
[illegible]	4 207	2 692	1 515	408.72	530.74	187.00	44.08	62.86	18.20	9.1	8.5	10.3
Enero	381	300	87	508.30	677.00	316.35	56.16	78.13	18.90	9.0	8.8	11.7
Febrero	410	311	99	540.11	683.01	230.50	50.55	81.00	17.03	9.2	8.0	12.9
Marzo	480	341	139	488.46	650.11	230.00	49.07	66.10	17.06	9.9	8.7	13.2
Abril	341	231	110	453.74	573.38	202.39	47.03	67.79	16.64	9.6	8.5	12.2
Mayo	323	200	120	388.61	497.79	214.63	43.21	66.00	18.46	9.1	7.6	11.6
Junio	346	200	136	385.63	511.27	190.47	42.53	58.34	20.14	9.1	8.8	9.8
Julio	361	200	161	351.85	480.38	157.78	40.56	57.50	17.07	8.8	8.9	8.8
Agosto	350	205	150	332.56	470.72	183.73	36.62	50.11	18.21	9.2	9.4	9.3
Septiembre	212	123	89	324.30	441.36	182.43	36.25	48.95	18.30	8.9	9.0	8.9
Octubre	247	140	107	303.97	413.75	180.73	30.72	50.30	20.30	7.7	7.4	7.9
Noviembre	329	190	130	300.60	410.62	154.06	34.62	52.72	19.46	7.9	7.9	7.3
Diciembre	480	200	200	331.08	443.81	187.86	33.04	47.83	17.07	9.8	9.3	13.5
[illegible]	4 825	2 950	1 875	387.41	501.30	188.50	39.04	51.00	17.88	9.9	9.7	10.4
Enero	382	253	129	453.04	570.97	178.35	42.20	53.45	18.43	10.7	10.7	10.8
Febrero	381	278	103	478.53	580.04	184.92	43.08	56.30	15.63	10.9	10.6	11.5
Marzo	470	315	161	458.70	578.44	222.70	41.37	59.00	18.30	11.1	9.8	12.6
Abril	341	229	112	435.94	507.73	197.93	41.33	58.00	18.94	9.8	9.4	10.7
Mayo	358	221	136	348.73	458.36	175.00	38.26	52.49	18.00	9.1	8.7	9.8
Junio	386	235	151	349.63	495.76	187.93	35.63	48.94	20.74	9.5	10.5	8.1
Julio	390	227	163	332.79	420.60	202.57	34.47	45.72	19.00	9.7	9.3	10.1
Agosto	410	249	161	337.74	451.95	182.26	35.00	43.86	18.87	9.6	10.3	8.8
Septiembre	251	153	98	332.94	438.87	180.40	35.55	47.16	17.50	9.4	9.3	9.5
Octubre	329	202	127	347.21	462.38	182.73	39.06	52.91	18.75	8.8	8.8	8.7
Noviembre	414	266	149	373.37	480.55	173.53	42.03	56.02	18.56	8.8	8.4	9.4
Diciembre	525	318	207	397.16	513.34	218.41	36.63	52.06	17.33	10.8	9.7	12.8

* [illegible]
1/ [illegible]

figura d

CUADRO IV-11

VIAJEROS AL INTERIOR

AÑOS TRIMESTRES Y MESES	NUMERO DE TURISTAS (miles)			GASTO MEDIO (dólares)			GASTO MEDIO DIARIO (dólares)			PERMANENCIA MEDIA (días)		
	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre	Total	Vía Aérea	Vía Terrestre
[illegible]	3 787	2 172	[illegible]	373.13	[illegible]	177.78	38.13	[illegible]	[illegible]	10.3		12.9
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	131.13	[illegible]	54.46	12.63	9.7	[illegible]	10.4
[illegible]	[illegible]	[illegible]	[illegible]	419.17	543.78	[illegible]	[illegible]	54.78	16.77	9.5	[illegible]	11.6
[illegible]	[illegible]	[illegible]	1 515	[illegible]	532.74	[illegible]	[illegible]	62.50	[illegible]	9.1	8.5	[illegible]
[illegible]	[illegible]	[illegible]	1 675	387.41	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.9	8.7	10.4
I	[illegible]	[illegible]	373	[illegible]	575.37	[illegible]	42.42	58.15	[illegible]	10.9	10.3	12.3
II	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.5	9.5	9.4
III	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	33.14	[illegible]	[illegible]	10.1	[illegible]	9.4
IV	[illegible]	[illegible]	[illegible]	376.42	[illegible]	[illegible]	38.12	54.11	[illegible]	9.6	9.0	10.6
[illegible]	[illegible]	[illegible]	1 772	[illegible]	[illegible]	[illegible]	43.31	[illegible]	[illegible]	9.7	9.9	11.4
I	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	211.15	[illegible]	[illegible]	[illegible]	9.9	9.9	12.3
II	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10.0	8.7	12.7
III	[illegible]	777	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.9	9.7	10.1
IV	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.1	9.3	10.7
[illegible]												
I	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10.4	9.3	13.3
II	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10.2	8.8	12.7
[illegible]												
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	45.72	[illegible]	9.7	9.3	10.1
[illegible]	410	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	251	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	299	[illegible]	127	[illegible]	[illegible]	[illegible]	[illegible]	52.51	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	145	[illegible]	[illegible]	173.58	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.4
[illegible]	[illegible]	318	[illegible]	387.16	513.94	216.61	[illegible]	[illegible]	[illegible]	[illegible]	9.7	12.9
[illegible]												
[illegible]	[illegible]	301	118	[illegible]	[illegible]	211.38	[illegible]	[illegible]	[illegible]	9.7	9.1	11.8
[illegible]	[illegible]	[illegible]	[illegible]	512.61	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	11.0
[illegible]	[illegible]	[illegible]	[illegible]	477.57	576.91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	431.46	516.74	230.15	[illegible]	65.13	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	41.51	52.52	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	252	[illegible]	[illegible]	[illegible]	231.73	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	319	[illegible]	115	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	230	[illegible]	[illegible]	[illegible]	197.98	47.36	61.76	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	467	297	151	[illegible]	488.72	266.75	45.53	63.15	19.75	8.7	7.8	10.4
[illegible]	[illegible]	304	277	[illegible]	[illegible]	715.38	37.53	[illegible]	[illegible]	10.0	8.7	11.0
[illegible]												
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	18.15	10.1	9.9	13.0
[illegible]	[illegible]	[illegible]	137	[illegible]	[illegible]	[illegible]	[illegible]	67.91	[illegible]	9.1	9.4	13.3
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	14.19	11.7	9.0	10.0
[illegible]	[illegible]	321	146	472.91	561.27	[illegible]	44.12	60.25	20.37	10.7	9.3	13.0
[illegible]	[illegible]	[illegible]	147	[illegible]	511.43	[illegible]	41.53	[illegible]	19.33	10.1	8.4	13.0
[illegible]	[illegible]	27C	[illegible]	[illegible]	580.34	[illegible]	42.11	58.57	21.17	9.6	8.6	11.3

[illegible]

figura e

FIGURA 6

DATA LISTING FOR FILE TURISMO.DAT

VARIABLE INGRESOS

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1978	96716.24	110370.0	125833.9	84235.70	74479.50	71148.00	100411.8	91489.20	64919.04	72242.56	91949.70	135502.40
1979	117162.80	132811.10	139561.60	124325.30	98036.96	111316.10	118981.30	127536.20	82244.05	99240.96	113575.00	176348.40
1980	157030.20	175140.50	176161.80	137914.50	120525.50	124381.60	127934.40	133487.80	94174.20	115458.60	135433.60	171439.50
1981	168993.60	185395.20	167100.10	151411.30	135172.30	138304.20	137555.40	129130.70	98267.82	111079.30	134202.00	204063.90
1982	165721.60	166010.50	163135.10	129712.40	103768.70	94218.00	120505.10	103164.60	65596.56	77937.82	89613.12	126915.40
1983	122506.70	146821.50	168350.40	128330.20	114716.90	120559.80	149787.00	123031.10	88492.52	119101.70	138868.50	203486.00
1984	185594.80	229404.90	227328.70	171437.30	152200.10	150626.00	142157.70	142234.70	96386.52	111482.50	141892.50	202384.00
1985	211973.20	225135.10	232380.10	154725.30	126813.00	133180.30	119991.10	121191.20	68751.59	75080.59	98140.80	152577.40
1986	164000.50	181557.90	218103.20	138289.10	125194.10	134957.20	129788.10	138555.40	83567.94	114232.10	154575	208509.00
1987	226153.20	242369.90	255022.40	195019.90	168534.40	161244.90	130990.00	176950.90	120025.80	158998.50	176153.80	216997.70
1988	261129.70	288812.90	315671.20	220849.00	171060.20	175754.70						

CAPITULO 3

ESTUDIO DE LOS PAQUETES

En èste capìtulo se muestran, por separado, los dos paquetes a utilizar en el anàlisis. Es necesario aclarar que solamente se pretende presentar lo correspondiente al Mètodo de Box-Jenkins ya que, el paquete STATGRAPHICS , en cuanto a contenido, comprende temas que el FORECAST PLUS no contempla y viceversa.. El objetivo es dar a conocer el funcionamiento de los paquetes y de este modo hacer màs clara la aplicaciòn y la comparaciòn de ellos.

FORECAST PLUS es un paquete estadìstico de pronòstico que maneja todas las etapas para el anàlisis de predicciòn. Permite analizar los datos de manera gràfica y el manejo de menùs es muy claro y sencillo ; sobre todo en el manejo de las etapas del mètodo.

El paquete STATGRAPHICS (Statistical Graphics System) es un paquete para computadoras personales (PC) integrado por una gran variedad de funciones estadìsticas con gràficas de color de alta resoluciòn. Permite el acceso a gràficas en la mayorìa de los procedimientos estadìsticos haciendo màs fàcil para el usuario explorar los datos de manera numèrica y gràfica.

3.1 FORECAST PLUS

Los planes del proyecto Forecast Plus empiezan en 1983, en este tiempo , Walonick Associates (compañìa creadora del paquete), comercializaba de manera satisfactoria con un paquete de anàlisis estadìstico llamado STATPAC. Walonick Associates se encontrò con la necesidad de crear un paquete de pronòsticos PROFESIONAL. Y es a finales de 1983 cuando se empiezan a desarrollar los programas de Forecast Plus; es obvio que el proyecto era demasiado complejo para que una sola persona trabajara en èl. Por lo que mucha gente trabajò en el proyecto de manera contìnua desde su inicio. Esta experiencia demostrò que un buen paquete de software nunca quedarà estàtico.

GENERALIDADES

Forecast Plus comprende 3 discos:

1. DATA MANAGEMENT. Contiene todos los programas necesarios para crear, editar y manipular archivos de datos. Comprende un conjunto de programas para la exploraciòn de datos.

2. MAIN SYSTEM. Contiene todos los programas para generar pronósticos.

3. BATCH DISK. Contiene los programas necesarios para crear y correr series de "forecasting tasks" en modo batch(1).

El primer paso en cualquier procedimiento de pronóstico es recolectar y asentar los datos en la computadora. El disco "DATA MANAGEMENT" permite manipular los datos según se necesite una vez que estén almacenados. Estos archivos de datos pueden contener tanto una variable (simple) como variables múltiples (series univariadas y series multivariadas).

El segundo paso es adquirir un conocimiento general de los datos. Existen diferentes técnicas de predicción apropiadas a diferentes tipos de patrones en los datos (estacionalidad vs no estacionalidad). El paquete de exploración en el disco de Data Management es una parte muy importante del FORECAST PLUS ya que da la opción de escoger la técnica adecuada. Una vez entendidos los datos, es posible escoger una técnica de predicción apropiada.

El paso final es modelar los datos con un método específico de pronóstico. FORECAST PLUS proporciona 13 métodos diferentes a escoger; varían en cuanto a sofisticación, técnicas, aplicabilidad a conjuntos muy específicos de datos. Como se ha venido mencionando, el método a utilizar será el Box - Jenkins.

Todos los programas del FORECAST PLUS están escritos en basic y luego compilados en modulos .EXE . Estos programas no están disponibles a examinarse ó modificarse.

FORECAST PLUS está dividido en 4 componentes que se presentan en el menú principal:

1. Manejo de datos (data management)
2. Paquete de exploración (exploratory Package)
3. Análisis de pronóstico (analysis forecasting)
4. Modo Batch (batch mode)
5. Fin de programa (end of program)

MANEJO DE DATOS (DATA MANAGEMENT)

Los programas contenidos en esta etapa sirven para crear, editar, transformar e imprimir archivos de datos. Una vez que se ha recolectado una serie histórica de datos puede usarse la opción de "manejo de datos" para almacenarlos. Esta etapa también contiene un menú:

(1) Modo Batch: No es un procesamiento contínuo,si no que espera a ser llamado para ser ejecutado.

DATA MANAGEMENT MENU

1. Proporcionar nuevos datos
2. Editar un archivo existente
3. Editar etiquetas ya existentes
4. Imprimir un archivo de datos
5. Transformar un archivo de datos
6. Cambiar el ajuste del dia
7. Reestructurar archivos de datos
8. Cambiar la tabla de parâmetros
9. Editor Batch

No se pretende analizar cada una de las opciones de c/u de los menô, unicamente se quiere presentar el camino para identificar, estimar, diagnosticar un modelo y pronosticar valores futuros. Por lo anterior solamente se verân con detalle las opciones 1,2,5.

1. PROPORCIONAR NUEVOS DATOS

Actualmente existen 2 programas de entrada diferentes dentro de este paquete, uno para archivos de una sola variable (single variable) y otro para archivos de variables mûltiples (multiple variables). Ambos programas permiten crear un archivo nuevo de datos ô añadir "records" a un archivo ya existente.

Para cada archivo de datos realmente existen dos archivos, uno contiene el renglôn actual de datos mientras que el otro contiene informaciôn de niveles (labeling information) sobre los datos. Es decir que si se crea un archivo llamado DATOS se crearâ un archivo de niveles llamado DATOS.LBL.

Cuando se va a crear un archivo nuevo de datos es necesario proporcionar cierta informaciôn. Esta informaciôn es la que queda almacenada en el archivo .LBL. La pantalla que se despliega es:

nômero de datos	1
Tipo de niveles	YEAR
nivel inicial	
observaciôn inicial	1
No. de observaciones entre cada nivel	
incremento de niveles	1

Note que 2 de estos parâmetros no tienen valor default. El nômero de variables se refiere a las variables que habrâ en el archivo de datos que en este caso en particular serâ una sola variable.

El tipo de nivel default es AÑO (YEAR) pudiendo cambiarse por MES, DIA ó NUMERADO. En el caso de tener un tipo de nivel AÑO se puden tener 12 observaciones entre cada año, cada observación representa un mes. En el caso de NUMERADO sólo irá numerando las observaciones. Para el caso del fenómeno de ingresos por turismo el tipo de nivel es año y con 12 observaciones entre cada nivel.

El nivel inicial es el nivel en el cual se encuentra el primer dato recolectado. Si se seleccionó el tipo de nivel año, el nivel inicial será el año en el que se recolectó el primer dato.

La observación inicial se refiere a la primera pieza que se recolectó que no tiene que ser el mismo valor del nivel inicial.

El número de observaciones entre cada nivel no tiene valor default por lo tanto tendrá que proporcionarse. En el caso de dar el valor de 1 significará que no existen observaciones entre cada nivel.

El incremento entre niveles será de 1 en la mayoría de los casos pero sí existen otras posibilidades. En el caso de MESES, por ejemplo, se puede tener información de cada dos meses, el incremento será de dos.

Una vez asignada la información anterior se procede a alamacenar los datos.

Como ya se mencionó se va a trabajar con una sola variable; el programa de entrada de datos para una variable permite insertar, borrar ó cambiar los datos facilmente. Es muy importante proporcionar los datos correctamente, ya que el dejar un espacio en blanco, por error, el paquete almacena un valor equivocado.

Por ejemplo:

VALOR ACTUAL	VALOR ALMACENADO
1	1
5	5
	7.5M
10	10
	12M
	14M
16	16

La letra M que aparece a la derecha del valor almacenado se refiere a "MISSING". Estos valores no se ven en pantalla sino que se crean cuando se guardan los datos en disco. Una solución a este problema es volver a cargar el archivo y volverlo a salvar en disco, así los valores erróneos serán sustituídos por la mediana.

Por otro lado, también se pueden corregir manualmente.

2. EDITAR UN ARCHIVO YA EXISTENTE

El programa pregunta el nombre del archivo a editar, el cual debe estar en línea durante la edición (NO REMOVER EL DISCO).

3. TRANSFORMAR UN ARCHIVO DE DATOS

El programa de transformación de datos es fácil de usar, permite hacer cualquier transformación en forma virtual (transformaciones algebraicas). se pueden crear nuevos valores a partir de los valores que ya existen.

En el análisis de series de tiempo, generalmente es necesario hacer transformaciones a los datos; por otro lado, ciertos procedimientos en FORECAST PLUS solamente trabajan con ciertos patrones y ésto hace necesario transformar los datos. Para hacer las transformaciones se pueden utilizar ciertos comandos, dependiendo de las necesidades:

NEW, SELECT, RECODE, IF-THEN RECODE, COMPUTE, IF-THEN COMPUTE, NORMALIZE, LAG, DIFFERENCE, DUMMY, SUM AND WRITE.

Cada uno de estos comandos se encuentran explicados con detalle en el manual del FORECAST PLUS(2)

PAQUETE DE EXPLORACION (EXPLORATORY PACKAGE)

Una vez cargado el archivo de datos a memoria y de elegir la opción 2 del menú principal, aparece un menú que expone 6 técnicas de exploración:

1. TIME PLOT
2. 4253HT ROBUST SMOOTHING (suavizamiento robusto)
3. BOX PLOT (para variación de tendencia/ciclo)
4. AGGREGATE BOX PLOT (para variación estacional)
5. SPREAD VS LEVEL PLOT (amplitud contra nivel)
6. AUTOCORRELATION FUNCTION (función de autocorrelación)

Estas opciones sirven para observar el comportamiento de los datos.

(2) Referencia (10)

Esto es muy ûtil en la etapa de identificaciôn. Este paquete de exploraciôn tiene la opciôn de realizar transformaciones a los datos. De las 6 opciones anteriores puede decirse que el TIME PLOT (opciôn 1. es el primer paso a efectuar. TIME PLOT presenta la grâfica de las observaciones con o sin transformaciones.

Existen 6 transformaciones ya predefinidas dentro del paquete de exploraciôn. Estas transformaciones son las mâs usadas para adquirir linealidad y estabilizar la varianza. Por ejemplo, la tranformaciôn logartimo se utiliza para cambiar de estacionalidad multiplicativa a aditiva. Es necesario notar que las transformaicones son temporales. es decir que se aplican solamente durante la sesiôn de trabajo.

NOTA: Si los datos contienen valores negativos, FORECAST PLUS, suma automâticamente una constante a los datos para poder realizar las transformaciones legalmente. Si los datos contienen valores negativos y la suma de la constante no es apropiada es necesario utilizar la opciôn 5 del menû de DATA MANAGEMENT (tranform a data file) para transformar los datos de una manera apropiada.

En seguida se presenta un resumen de cada una de estas têcnicas.

a) TIME PLOT

Como ya se dijo, esta têcnica puede ser el primer paso dentro de anâlisis de exploraciôn. Aquî se revela tendencia, estacionalidad. Se puede detectar el comportamiento de los datos originales. Dentro del TIME PLOT se tiene la opciôn de graficar datos transformados por lo tanto, si los datos mejoran se reflejarâ en la grâfica.

b) 4253HT ROBUST SMOTHING (suavizamiento robusto 4253)

Esta têcnica es un perfeccionamiento del TIME PLOT. Revela tendencia y patrones removiendo el ruido entorno a los datos. El têrmino "robusto" indica que esta têcnica no es particularmente sensitiva a la aleatoriedad de los datos, es decir, que se ignoran valores irregulares ô lejanos; la grâfica resultante imprime datos suaves, y existe la opciôn de graficar en el TIME PLOT , despuês de suavizar, para observar los valores âsperos ô irregulares (residuales despuês de suavizar).

El valor default en el menû es que SI se grafiquen estos datos se refiere al grado con el cual los datos no pueden ser considerados por esta têcnica. El paquete tiene la opciôn de guardar los datos suaves por un lado y por otro los irregulares ô residuales. El archivo que contenga los datos suaves puede ser usado para pronosticar y el otro para revelar irregualriedades en los datos.

c) BOX PLOT

El Box Plot es una técnica gráfica para resaltar algunos aspectos de los datos. Puede revelar la localización central (media), amplitud (spread), forma, dispersión de datos irregulares, etc. Esta técnica, como su nombre lo dice, grafica cajas (box) concentrando un cierto número de observaciones. Es necesario proporcionar este número de lo contrario el valor default depende del número de datos a ser analizado.

Basicamente, esta técnica despliega el valor medio de los datos sobre ciertos períodos de tiempo dados y una indicación de la variabilidad de los datos dentro de cada período. También revela la dirección de la tendencia y si los datos tienen estacionalidad multiplicativa ó aditiva.

El tipo de gráficas de cajas que maneja FORECAST PLUS es algo diferente a otras versiones; otros métodos de gráficas de cajas son más suceptibles a distorcionar datos. La técnica usada en este paquete es más estable ya que involucra el uso de estadísticos de mayor orden.

Una caja de datos que se distribuye normalmente puede aparecer como

figura 3.1

Una caja de datos no simétricos con datos discrepantes puede aparecer como

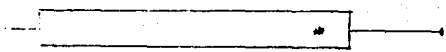


figura 3.2

Los extremos de la caja muestran en donde muchos de los datos decaen y las colas indican valores que no están cerca del centro pero que probablemente son datos discrepantes.

Existe también una variación del BOX PLOT que es el BOX PLOT AGREGADO. Se usa generalmente para hacer notar las diferencias entre mes y mes, por ejemplo, el patrón de estacionalidad y la variabilidad para cada mes. En este ejemplo el mes de ENERO de TODOS los años formará la primera caja, FEBRERO de TODOS los años la segunda y así sucesivamente.

La interpretación del BOX PLOT AGREGADO es la misma que la del BOX PLOT SIMPLE sólo que se manejan diferentes unidades de tiempo.

d) AMPLITUD VS. NIVEL (spread vs level)

La gráfica de amplitud vs nivel es un opción para detectar las relaciones entre el nivel de los datos y la variabilidad de los éstos. Este método es muy usado para detectar si la estacionalidad es aditiva ó multiplicativa.

Es necesario proporcionar el número de observaciones entre cada subconjunto de datos, el número mínimo de observaciones en cada subconjunto de datos deberá ser mayor que 3. Si los datos son estacionales es necesario proporcionar la longitud de la estacionalidad ó "span" ó un múltiplo de esta longitud como el número de observaciones en cada subconjunto. Si el número de observaciones se asigna como "automático (A)" éste será calculado en base al número total de observaciones de la serie:

Si N < 50 el número de observ. en cada caja es 4
N < 30 el número de observ. en cada caja es 5
N < = 500 el número de observ. en cada caja es N/10
N > 500 el número de observaciones en cada caja es N/20

El propósito de la gráfica de aplitud contra nivel es determinar si la variabilidad de los datos cambia con el nivel y si hay una transformación que pueda estabilizar la serie. Cuando los datos necesitan una transformación se despliega una línea diagonal en la gráfica (i.e, tiene una pendiente) fig. 3.3. Si una de las 6 opciones de transformaciones reduce la variabilidad, esta opción sugerirá una transformación apropiada. Si se corre por segunda vez la gráfica de amplitud vs nivel ya con la transformación antes sugerida y ésta efectivamente estabiliza la serie, se despliega una línea vertical (poca pendiente ó no hay pendiente) fig. 3.4.

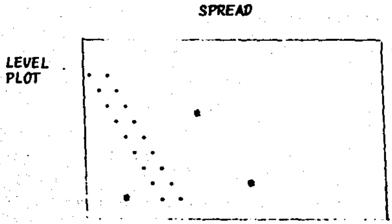


figura 3.3
la varianza es dependiente del nivel

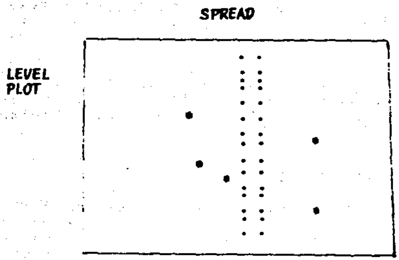


figura 3.4
La varianza no está relacionada con el nivel

e) FUNCION DE AUTOCORRELACION.

Como ya se explicó anteriormente, la gráfica de la función de autocorrelación es de gran utilidad para identificar la existencia de alguna relación entre los datos en uno ó varios periodos de tiempo. En el paquete de exploración se puede obtener la gráfica de la ACF de los datos originales y/ó con transformaciones; y se tienen, además 4 opciones:

1. Dar el grado de diferencias ordinarias.

2. Dar el grado de diferencias estacionales.

3. Dar la longitud de la estacionalidad (span).

4. Dar el # de períodos (lags) a imprimir.

Recuérdese que el diferenciar elimina tendencia y estacionalidad y que se debe tener cuidado con no sobrediferenciar ya que el modelo se complica. Esto se verá reflejado en la grafica de la función de autocorrelación, por lo que es muy recomendable graficar unicamente con primeras diferencias y observar la gráfica, luego graficar con segundas diferencias y observar los cambios de una gráfica a otra. En caso de diferencias estacionales puede suceder que con sólo las primeras diferencias es suficiente para que se elimine la estacionalidad.

El número de períodos de tiempo (lags) "default" a imprimir es el número de registros dividido entre 3, y éste es el número de autocorrelaciones a ser calculadas e impresas.

El comportamiento de la ACF cuando existe tendencia, estacionalidad, etc. se explicó con detalle en el capítulo 1. El mejor camino para aprender sobre la función de autocorrelación es graficar diferentes casos y observar el tipo de autocorrelaciones que se producen.

ANALISIS DE PRONOSTICO

La sección de pronóstico de este paquete contiene 13 técnicas que se pueden utilizar para predecir valores futuros. Algunas de estas técnicas se parecen; generalmente se toman en cuenta los resultados obtenidos en el paquete de exploración para escoger cual técnica utilizar. Cada una de éstas requiere de cierta información diferente a la ya proporcionada con los datos, estos valores dependerán del análisis de identificación y de la hipótesis generada. La meta al encontrar los mejores parámetros es que sean aquellos valores que minimicen la suma de cuadrados de los residuales (el error entre pronóstico y el valor real), es decir, aquellos valores que minimicen el error cuadrático medio (MSE), el porcentaje del error medio absoluto (MAPE) y el porcentaje del error medio (MPE). En otras palabras, se está buscando minimizar la diferencia entre pronósticos y valores reales.

Una vez aplicado el paquete de exploración es posible entender la varianza y estacionalidad de los datos y ésto es lo que determina que técnica es más conveninte usar. Si se está inseguro de cual técnica es más apropiada se pueden probar todas y ver cual produce el mejor pronóstico, el juicio personal juega un papel muy importante. El mejor camino para distinguir entre un BUEN pronóstico y un MAL pronóstico es el observar las gráficas. Si el pronóstico se vé bueno PROBABLEMENTE lo sea, si se vé malo, seguro lo es, es aquí en donde interviene el juicio y la intuición.

El método de Box - Jenkins es bastante accesible a cualquier tipo de datos, ya que se puede utilizar si se presenta cualquier tendencia, estacionalidad aditiva ó multiplicativa, etc. Posiblemente es una técnica difícil en cuanto a uso pero con la práctica se llega a convertir en una técnica muy noble.

METODO DE BOX - JENKINS

Después de escoger la opción 13 del menú de Análisis de Pronóstico aparece en menú correspondiente al método de Box - Jenkins:

1) Identificación
2) Estimación / Pronóstico
3) Regresar al menú de Análisis

1) Identificación.

Durante la fase de identificación es necesario descubrir las transformaciones que mejoren el modelo. Existe un menú de identificación en donde se tiene la opción de indicar cual transformación aplicar:

Transformación a la serie de datos originales	7
Grado de diferencias Ordinarias	0
Grado de diferencias Estacionales	0
Longitud del período estacional	12
Gráfica de función de autocorrelación	Y
Gráfica de Función de Autocorrelación Parcial	Y

Los valores que aparecen son los valores default. Posiblemente parece que todo es muy repetitivo, pero recuérdese que se esta en otra etapa del paquete y que todas las tranformaciones solamente duran mientras dura la sesión.

Una vez transformada la serie es recomendable desplegar las gráficas de ACF y de PACF para checar si los datos son estacionarios. Estos nos dicen si es necesario volver a transformar.

2) ESTIMACION

Cuando se elige esta opción es porque ya se tiene una serie estacionaria y algunos modelos tentativos identificados. El menú de estimación tiene todas las opciones necesarias para indicar correctamente el modelo. Este menú es el siguiente:

Transformación de la serie de datos original	7
Grado de diferencias ordinarias	0
grado de diferencias estacionales	0
longitud del período estacional	12
Términos autorregresivos ordinarios	0
Términos de medias móviles ordinarios	0
Términos autorregresivos estacionales	0
Términos de medias móviles estacionales	0
Incluye término constante	0
Número de pronósticos hacia atrás	0
Iteraciones máximas para la estimación de parámetros	50
Tolerancia de Convergencia	001
Estimación inicial (A/U)	A
Gráfica de función de Autocorrelación de residuales	Y

Los valores default para las transformaciones serán aquellos indicados en la fase de identificación.

El término constante es similar al de regresión, si se sacaron diferencias no deberá haber término constante, y si no existen diferencias sí deberá existir éste .

La técnica de Box -Jenkins usa un procedimiento de estimación no lineal y es un método iterativo, repite aproximaciones sucesivas en donde cada aproximación refina la estimación. El valor default para el máximo número de iteraciones es 50, ésto significa que el programa correrá por medio de su procedimiento de aproximaciones un máximo de 50 veces usando la estimación anterior para refinar la siguiente. Usualmente no se utilizan las 50 iteraciones para encontrar el mejor parámetro.

La tolerancia de convergencia de .001 indica que las iteraciones continúan hasta que el límite de iteraciones se alcanza ó hasta que el perfeccionamiento en la suma de cuadrados de los residuales es menor que .001.

FORECAST PLUS da la opción de empezar las iteraciones a partir de los parámetros dados por el usuario (una estimación inicial). Y por último la gráfica de la ACF de los residuales es lo que indica si el ajuste tuvo éxito ya que ajustando los parámetros poco a poco se van reduciendo las autocorrelaciones de los residuales al mínimo.

3) PRONOSTICO

Una vez encontrado el modelo se procede a generar el pronóstico. Los últimos parámetros encontrados en la etapa de estimación serán usados para producir el pronóstico. El menú que se despliega es el siguiente:

Nûm. de Pronôsticos	1
Origen del Pronôstico	(N)
Porcentaje del Intervalo de confianza del Pronôstico	95
Grâfica de datos originales pronôstico y error	Y
Tabla de datos originales, pronôstico y error	Y

El nûmero de pronôsticos se refiere a cuantos perîodos adelante se quiere pronosticar. El mâximo nûmero es 104, el mînimo es 1.

El origen del pronôstico es el punto en el cual se desea empezar a pronosticar. Usualmente es la ûltima observaciôn del conjunto de datos, el valor (N) es el nûmero de observaciones que se tienen. El origen mînimo es N / 2 (la mitad de los datos).

El porcentaje de confianza es 95, es decir, que se puede esperar que el 5% de los datos caigan fuera del intervalo de confianza especificado.

El pronôstico actual estâ dado por una grâfica especîfica de datos originales, pronôstico y error y la tabla de datos, pronôstico y error.

En resumen, la experiencia es lo escencial para utilizar este paquete y para generar pronôsticos, FORECAST PLUS es mucho muy sencillo de utilizar, los pasos a seguir se pueden definir como sigue:

a. Usar el paquete de exploraciôn para examinar los datos.

b. Escoger una têcnica de pronôstico (BOX -JENKINS).

c. Probar la têcnica. Revisar los parâmetros y volver a probar la têcnica. Si los resultados se consideran apropiados, usarlos para el pronôstico, si no, probar con nuevos parâmetros ô cambiar la têcnica.

3.2 STATGRAPHICS.

El paquete STATGRAPHICS (Statistical Graphics System) comprende una gran variedad de temas como : funciones de distribución,análisis de varianza, análisis de regresión , diseño de experimentos, análisis numérico, análisis de series de tiempo, pronósticos, muestreo, etc. Por la variedad de temas es un poco más complicado de usar en caso de que el usuario no tenga experiencia. Es de gran utilidad leer el manual y a su vez intentar utilizar el paquete. STATGRAPHICS tiene una guía de usuarios que se divide en 10 partes:

1) Tutorial de Introducción.
2) Descripción de como usar el paquete STATGRAPHICS (requerimientos del sistema)
3) Descripción detallada del manejo de datos.
4) a 8) Referencias del manual para el análisis de datos y procedimientos de gráficas.
9) Descripción de como usar STATGRAPHICS con el sistema APLPLUS/PC
   Para usar STATGRAPHICS de este modo se necesita tener el sistema, versión 4.0 ó después, y estar familiarizado con el lenguaje de programación APL.

10) Descripción de Apéndices incluyendo glosario.

Este paquete está contenido en 4 discos:

START-UP DISK
PROGRAM DISK 1
PROGRAM DISK 2
SAMPLE DATA SETS/on line help disk (muestras para ejemplos)

El último disco contiene datos de muestra que se usan para correr ejemplos, este disco se maneja en la unidad de disco correspondiente al disco de datos, y no a la correspondiente al paquete. También contiene pantallas con texto de ayuda.

GENERALIDADES

Para poder trabajar con Statgraphics es necesario tener disponible un cierto hardware y un cierto software:

- una computadora personal IBM,XT, AT ó compatible.

- 384 kilobytes de acceso a memoria random

- un teclado (ó tablero)

- 2 unidades de disco de doble lado y doble densidad ó una unidad de doble lado y doble densidad y un disco duro.

- un adaptador gráfico y monitor monocromático ó monitor de color.

- sistema operativo MS-DOS, versión 2.0 en adelante.

Para iniciar la sesión de trabajo hay que definir el ambiente en el que se va a trabajar:

- número de monitores (1 ó 2), el valor default es 1, es monocromático ó de color.

- unidad de disco en donde se encuentra el paquete y unidad de disco en donde estará el disco de datos.

Se despliega una tabla mostrando todos los adaptadores gráficos que se aceptan, la resolución y el número de colores disponibles con el tipo de monitor recomendado para el adaptador.

Se tiene la opción de guardar esta información en el "automatic logon", así la siguiente sesión en STATGRAPHICS no se tendrá que dar otra vez la misma información.

El menú principal contiene 22 secciones del nivel A a la V agrupadas bajo los encabezados principales de este menú, éste da acceso a cualquier análisis de datos y a procedimientos gráficos.

La primera vez que se despliega este menú, el cursor está posicionado en la primera selección del menú, (MANEJO DE DATOS).

Existen dos formas de mover el cursor:

- Usar las teclas de flechas para el control del cursor.

- Teclear la letra correspondiente a la selección deseada.

SELECCION DE TRAYECTORIA RAPIDA

Una vez falimiliarizados con STATGRAPHICS se puede trabajar con las distintas opciones que este paquete tiene disponibles, seleccionàndolas del menù principal con las teclas de control de cursor (flechas) ó utilizando la "TRAYECORIA RAPIDA" la cual es de gran utilidad. Utilizando la trayecoria ràpida se puede accesar CUALQUIER procedimiento desde CUALQUIER menù tecleando el número y letra correspondiente y presionando "return". Por ejemplo, se puede seleccionar A1 para entrar a Manejo de Datos a la opción de manipulación de variables definidas. seleccionada. Cuando se usa la trayectoria ràpida, el sistema se dirige directamente al procedimiento deseado por lo que el asterisco dentro del submenù ayuda a verificar si el procedimiento es el correcto mientras que el paquete carga los programas necesarios.

MANEJO DE DATOS (Data Management)

Una vez seleccionada esta opción aparece en la pantalla el menù que corresponde a manejo de datos:

MANEJO DE DATOS

1. Manipular Variables Definidas.
2. Editor de Datos "full-screen".
3. Leer variables definidas de un archivo SG.
4. Grabar en archivos SG.
5. Importar datos de un archivo ASCII.
6. Exportar datos de un rachivo ASCII.
7. Importar datos de un archivo DIF.
8. Exportar datos de un archivo DIF.
9. Importar una hoja de càculo de lotus 123.
10. Exportar una hoja de càlculo de lotus 123.
11. Recodificar valores extraños.

De la misma manera que en 3.1 , se explicaràn unicamente las opciones necesarias para crear un archivo, modificarlo, etc.

STATGRAPHICS permite crear y almacenar variables en formatos especiales en disco llamados archivos statgraphics.

El nombre de la variable puede contener hasta 10 caracteres y puede consistir de mayúsculas y/ó minúsculas y/ó números, deberá de empezar con una letra. Las letras mayúsculas son consideradas distintas a las minúsculas, así que un nombre escrito en mayúsculas será distinto a ese mismo nombre escrito en minúsculas. Dos variables pueden llamarse igual si no pertenecen a un mismo archivo mientras que no sean usadas al mismo tiempo. Es obvio que es recomendable que no se repitan nombres de variables para evitar confusiones.

A) MANIPULACION DE VARIABLES DEFINIDAS

Este procedimiento permite manipular los datos. Es recomendable utilizar esta opción cuando se necesite:

- Crear rapidamente nuevas variables, cuando las variables son relativamente pequeñas

- Ejecutar cálculos rápidos (raíz cuadrada por ejemplo)

- Modificar datos en tipo y forma.

- Renombrar variables.

Dentro de este procedimiento se pueden ejecutar diez operaciones ó transformaciones; ninguna de éstas quedan salvadas en un archivo. Si se maneja una variable definida de un archivo, ésta se convierte en una variable RAM al momento de ser transformada, los valores previos son borrados (de memoria), es decir que los valores previamente guardados en un archivo no se modifican, y se dice que los valores nuevos asignados son una variable RAM (random access memory).

B) EDITOR DE DATOS "FULL SCREEN"

Este editor está diseñado para que el usuario almacene sus datos. Se puede trabajar con variables RAM definidas en el espacio de trabajo ó con variables previamente almacenadas en un archivo. Este editor permite:

- Inicializar la entrada de un conjunto de datos grande.

- Realizar cambios permanentes en las variables.

- Insertar y/ó borrar valores.

- Localizar y cambiar algún valor en particular.

- Desplegar e imprimir datos.

Los datos pueden ser desplegados en el editor de tres maneras:

- Numêricos, caracteres escalares, vectores, matrices de caracteres que contengan 10 ò màs columnas, cada una ocupa una columna en el editor.

- Matrices numêricas ocupando una columna en el editor por cada columna que contenga la matriz. El nombre de la variable se despliega arriba de cada columna (en el editor), y asî se identifican columnas de una misma matriz.

- Matrices de caracteres con màs de 10 columnas, ocupan una columna del editor por cada 10 columnas de la matriz. El nombre de la variable es desplegado arriba de cada columna en el editor y asî identificar columnas sucesivas de la misma matriz.

Cuando se estan proporcionando los datos y algunos de estos contiene un valor perdido (missing value) hay que dejar el espacio en blanco en la posiciòn que corresponde a este valor. Si este valor ocurre en la ûltima posiciòn de la columna, el sistema no reconoce el espacio en blanco; para evitar èste problema hay que dar un valor adicional en la celdilla abajo del valor perdido y luego borrar los espacios extra.

En STATGRAPHICS, el còdigo de un "missing value" para valores numêricos està dado por -32763. Cuando se salvan los datos en un archivo utilizando las teclas F2 ò F3, cualquier espacio en blanco en variables numêricas serà reemplazado por el valor de còdigo antes mencionado.

C) LEER VARIABLES DE UN ARCHIVO SG

El procedimiento para leer variables permite leer datos de un archivo y usar las variables para el anàlisis de datos y en procedimientos gràficos. Cuando se selecciona este procedimiento el sistema despliega los nombres de las variables disponibles; hay que recordar que los archivos que contienen a las variables, son archivos con un formato que utiliza statgraphics en particular.

D) GRABAR A UN ARCHIVO STATGRAPHICS

Este procedimiento permite crear, copiar, renombrar, borrar archivos statgraphics asî como grabar variables en los archivos.

Se necesita, como primer paso, crear un archivo antes de poder almacenar variables en èl. Se puede alamcenar tantas variables como se quiera en un solo archivo, pero es màs fàcil manejar los datos si se crea un archivo diferente para cada grupo de variables. Una vez creado el archivo, se pueden añadir variables en êl ò actualizar variables cuyos valores van cambiando durante la sesiòn de trabajo; estos cambios son guardados en un archivo cuando se salvan por medio de la opciòn (2) del menû ò cuando se utiliza este procedimiento.

Cuando se selecciona esta opción se despliega una lista de todos los archivos con formato statgraphics. Se pueden realizar 5 operaciones,con los archivos, dentro de esta opción(copiar,borrar, copiar un archivo nuevo,renombrar, grabar).

También se le pueden hacer transformaciones a las variables, el sistema despliega la lista de variables disponibles en el archivo, y las 6 operaciones a realizar son:
- COMMENTAR (COMMENT): cambiar el comentario asignado a la variable. Escoger la variable deseada y presionar C, se podrá ahora teclear cualquier otro comentario, puede contener hasta 30 caracteres.

- DESPLEGAR (DISPLAY): desplegar en pantalla el contenido de la variable seleccionada.

- BORRAR (DELETE): Borrar variables de archivo(s).

- NUEVO (NEW): Grabar una nueva variable en un archivo. Se tendrá que proporcionar el nombre de la variable y el comentario.

- RENOMBRAR (RENAME): Cambiar de nombre a una variable.

- ACTUALIZAR (UPDATE): Actualizar una variable de un archivo que ha cambiado durante la sesión; presionar U. La variable será actualizada en el archivo y aparecerán el día y la hora en la lista de nombres de variables lo cual confirma cuando se actualizó.

E) RECODIFICAR VALORES EXTRAÑOS (MISSING VALUES)

STATGRAPHICS reconoce el código de valores extraños en los datos de las variables. En muchas situaciones los procedimientos excluyen todos los casos que contengan estos valores en cualquiera de las variables. Cuando se proporcionan datos dentro de variables RAM ó dentro de archivos, los valores extraños deberán ser blancos; si son datos de caracter (alfanuméricos) los blancos se respetan, si son datos numéricos el blanco se reemplaza por el valor de código (-32768). Cuando se selecciona el procedimiento de recodificación de valores extraños se despliegan los archivos disponibles. Se pueden recodificar valores numéricos tanto de archivos previamente almacenados como de variables RAM en el espacio de trabajo. Las operaciones disponibles son:

- DESPLEGAR (D): Se pueden desplegar las variables antes ó después de recodificar para verificar el contenido antes de corregir y después de haberlo hecho.

- RECODIFICAR (R): Si la variable es salvada en un archivo, el valor perdido será recodificado y salvado en el archivo.

- RECODIFICAR TODO (A): Con variables guardadas en un archivo todas las variables deberán contener el mismo valor de código para poder utilizar esta opción.

ANALISIS DE SERIES DE TIEMPO

El procedimiento de Anàlisis de Series de Tiempo en STATGRAPHICS provee de 15 opciones , las cuales se enlistan en un menû:

MENU DE ANALISIS DE SERIES DE TIEMPO

1. Grafica de secuencia de tiempo horizontal.
2. Grafica de secuencia de tiempo vertical.
3. Grafica de subseries estacionales
4. Funciòn de autocorrelaciòn
5. Funciòn de Autocorrelaciòn parcial.
6. Funciòn de correlaciòn cruzada
7. Diferenciaciòn ordinaria y estacional.
8. Remover media ò tendencia.
9. Transformaciòn de Box-Cox
10. Periodograma.
11. Periodograma integral.
12. "data tapering"
13. Grafica vs. frecuencias de fourier.
14. Box-Jenkins
15. Matriz de correlaciòn.

1. GRAFICA DE SECUENCIA DE TIEMPO, HORIZONTAL

Este procedimiento crea una lìnea contìnua de una ò dos series de tiempo. Se pueden especificar los valores de las series y los puntos en el tiempo ò graficar unicamente valores contra ìndices (enteros, desde 1 hasta el nùmero de observaciones).

Los valores extraños (missing values) son excluidos de la gràfica y la informaciòn de entrada para que se despliegue la gràfica es:

- Nombre de la variable que contiene los datos.
- Nombre de la variable que contiene los puntos en el tiempo.
  (Este vector deberà ser de la misma longitud que el de los datos originales).
- Nombre de la variable que contiene la segunda serie de tiempo.
  (este vector deberà ser de la misma longitud de los 2 anteriores).

- Escala de ejes (derecha ó izquierda).
  Esta opción aparece cuando se están graficando 2 variables de series de tiempo.

2. GRAFICA DE SUBSERIES ESTACIONALES.

La gráfica de subseries estacionales despliega los valores de una series de tiempo discreta que presenta períodos. Cada grupo de datos representa todos los valores de un período para todos los ciclos.

Las líneas horizontales representan valores promedio de las observaciones para todos los ciclos del período correspondiente. Las líneas verticales son graficadas desde los promedios de los valores actuales para cada observación. Los datos de entrada para este procedimiento son:

- El nombre de la variable que contiene los datos.

- Longitud del período estacional, valor default es 12.

- Estación de el primer punto de datos, i.e, el período de inicio de los datos. El valor default es 1, si los datos, por ejemplo, tienen un ciclo anual y empiezan en junio, el valor será 6.

3. FUNCION DE AUTOCORRELACION.

La función de autocorrelación calcula los coeficientes de autocorrelación de una serie de tiempo. El sistema grafica barras verticales para retrazos (lags) desde 1 hasta el número que se desee. Se pueden utilizar los coeficientes de autocorrelación para probar estacionalidad u otra periodicidad, ó como un paso preliminar para determinar un modelo, es decir, identificar modelos "tentativos". La información de entrada a proporcionar es:

- Nombre de la variable que contiene los datos.

- número de "lags" que se desean.

Una vez desplegada la gráfica, si se presiona la tecla <RETURN> se despliega la tabla de coeficientes de autocorrelación estimados así como el error estandar estimado; y existe la posibilidad de guardar estas estimaciones en el disco.

4. FUNCION DE AUTOCORRELACION PARCIAL.

Este procedimiento grafica la PACF usando barras verticales (igual al caso anterior), cuya altura es proporcional al valor del coeficiente. La información de entrada para desplegar la gráfica es la misma que en el caso anterior. También se despliega la gráfica de coeficientes estimados y error estándar pudiendo ser guardada.

5. DIFERENCIAS SIMPLES Y ESTACIONALES

Este procedimiento permite ejecutar diferencias en la serie. Las diferencias ordinarias restan el valor en el tiempo t del valor en el tiempo t+1, el resultado es un vector de datos con una observación menos que en la serie original.

Las diferencias estacionales eliminan los efectos estacionales restando el valor que ocurre una ESTACION antes del valor actual. El resultado es una estación menos que los datos originales. El sistema permite guardar estas diferencias en una variable. La información de entrada es:

- Nombre de la variable que contiene los datos.

- Orden de las diferencias. Si se teclea (1) las diferencias son ordinarias, si se necesitan diferencias estacionales proporcionar la longitud estacional.

- Nombre de la variable en la que se quiere salvar los resultados. Si se presiona la tecla <return> éstos no serán guardados.

El sistema crea una variable RAM con el nombre de la nueva variable y despliega un mensaje diciendo cuantas observaciones fueron guardadas y esta variable se puede usar en otros procedimientos. En caso de querer usar esta variable en otra sesión de statgraphics se tienen que usar los procedimientos descritos en "Manejo de Datos" para guardarla en disco.

6. PERIODOGRAMA INTEGRAL

El procedimiento del periodograma integral verifica en donde los datos, en una serie de tiempo, son aleatorios. El procedimiento grafica la suma acumulada del periodograma ordinario normalizada a (0,1) en la escala vertical; también incluye límites de Kolmogorov-Smirnov del 75% y 95% para una distribución uniforme de ordenadas. La información de entrada es:

- Nombre de la variable de datos.

- Decir si se quiere restar la media de los datos.

METODO DE BOX AND JENKINS

Este procedimiento permite modelar con la serie original ó diferenciada, es necesario utilizar las teclas de funciones (function keys) para accesar los procedimientos de estimación, diagnóstico y predicción. Los datos de entrada para ejecutar Box and Jenkins son:

- Nombre de la variable que contiene los datos.

Una vez dado este nombre aparece un panel en la pantalla, en el que se accesan otros datos de entrada correspondientes al modelo "tentativo" y la información necesaria para estimación y diagnóstico, éstos son los siguientes:

ORDEN DE DIFERENECIAS ORDINARIAS: Orden de las diferencias a realizarse en la estimación del modelo.

0 ---> Sin diferencias

1 ---> primeras diferencias

2 ---> segundas diferencias

CONSTANTE CONTENIDA EN EL MODELO: Si se desea incluiur la constante al modelo se responde YES, en caso contrario se responde NO.

ORDEN DE TERMINOS AR : número de parámetros autorregresivos a estimar (orden p de 0 a 6).

ORDEN DE TERMINOS MA: número de parámetros de medias móviles a estimar (orden q de 0 a 6).

ORDEN DE DIFERENCIAS ESTACIONALES: Número de diferencias estacionales a aplicarse en la estimación del modelo.

LONGITUD DE LA ESTACIONALIDAD: proporcionar el "span",es decir, el orden estacional. Dar cero si no es estacional el modelo (un span=12 se usa para datos mensuales).

ORDEN DE TERMINOS SAR: Número de parámetros autorregresivos estacionales a estimar en el modelo (orden P de 0 a 6).

ORDEN DE TERMINOS SMA: Número de parámetros de medias móviles a estimarse en el modelo (orden Q de 0 a 6).

NUMERO MAXIMO DE RETRASOS (LAGS) PARA LA GRAFICA DE ACF: Número máximo de retrasos que se desea en gráfica de la función de autocorrelación.

NUMERO DE RETRASOS PARA LA PRUEBA DE CHI CUADRADA: Número de términos para la prueba de chi cuadrada de bondad de ajuste de en las autocorrelaciones de los residuales.

NUMERO MAXIMO DE RETRASOS PARA LA GRAFICA DE PACF: Número máximo de retrasos que se desea en la gráfica de la función de autocorrelación parcial.

NUMERO DE PRONOSTICOS QUE SE DESEAN: Número de puntos a ser incluidos en el pronóstico sobre el modelo ya ajustado.

CRITERIO 1 PARA TERMINAR: El sistema terminará el proceso de estimación cuando la suma de cuadrados de los residuales entre iteraciones alcance el cambio mínimo especificado en este punto (mayor o igual que cero).

CRITERIO 2 PARA TERMINAR: El sistema terminará el proceso de estimación cuando el parámetro estimado entre iteraciones alcance el mínimo cambio especificado en este punto.

Como se dijo antes esta es a la información que se actualiza en el panel dependiendo del modelo que se quiera estimar. Por otro lado, también es posible realizar algunos de los procedimientos descritos en la sección anterior, de este mismo capítulo, como son las graficas de ACF y PACF, diferencias ordinarias y estacionales, etc. Estos procedimientos se accesan con las opciones que aparecen en la parte inferior de la pantalla y utilizando las teclas de funciones (F2,F3,...) y son las siguientes:

F2 (SERIES) : Crea las gráficas de los datos originales ó datos transformados (diferencias). En este último caso, el sistema requiere del orden de las diferencias ordinarias o estacionales:

0 ---> No diferenciar

1 ---> primeras diferencias

2 ---> segundas diferencias

F3 (ACF) : Grafica la funcioñ de autocorrelación muestral de los datos originales ó transformados.

F4 (PACF) : Grafica la función de autocorrelación parcial muestral de los datos originales ó transformados.

En estos dos procedimientos, se despliega la tabla de coeficientes de autocorrelación y el error estandar estimados con la posibilidad de ser salvados.

F5 (ESTIM) : Se entra en el proceso de estimación de los parámetros del modelo y de las estadísticas. Si la estimación resulta satisfactoria se despliega una tabla con los parámetros, errores estándar aproximados, niveles de significancia etc. Esta tabla incluye la varianza estimada de los residuales basada en la media del error estandar y en la prueba de chi cuadrada.

Las siguientes funciones se deben utilizar una vez estimado el modelo:

F6 (RACF) : Grafica la función de autocorrelación de los residuales del modelo.

F7 (RPACF) : Grafica la función de autocorrelación parcial de los residuales del modelo.

F8 (INTPER) : Produce el periodograma integral de los residuales del modelo.

F9 (FORCST) : Grafica el pronóstico usando el modelo ya ajustado con un 50% y 95% de límites de confianza. Es posible guardar este pronóstico así como los límites.

NOTA: Todos los procedimientos que grafican preguntan al usuario si se desea la gráfica de datos originales ó diferenciados.

3.3 OTROS PAQUETES

Es importante saber que existen otros paquetes que manejan series de tiempo y en particular el mètodo de Box-Jenkins. Los dos paquetes expuestos en las secciones 3.1 y 3.2 son para ser utilizados en computadoras personales (PC) y asì como èstos hay otros similares y algunos de ellos se explican en esta secciòn.

Tambièn existen paquetes para computadoras mini, macro, etc. que generalmente se manejan en base a llamados a las subrutinas del paquete (biblioteca) a travès de un programa fuente elaborado por el usuario. Esto podrìa resultar màs complejo para el usuario ya que se necesitan conocimientos de programaciòn. Por ejemplo si se desean obtener las diferencias, ordinarias dentro del programa se hace el llamado a la subrutina que realiza diferencias ordinarias. Estas bibliotecas no manejan menus, asì para obtener varios resultados en una misma corrida se tendràn que hacer diferentes llamados a subrutinas distintas dentro del programa. En esta secciòn se explican màs detenidamente estas bibliotecas asì como algunos paquetes para PC.

BIBLIOTECAS (1)

Generalmente estos paquetes manejan una gran variedad de temas, como por ejemplo:

---> Ceros y Polinomios

---> Ecuaciones Diferenciales

---> Minimizar y Maximizar una Funciòn

---> Ajuste de Curvas

---> Eigenvalores y Eigenvectores

etc...

A) INTERNATIONAL MATHEMATICAL AND SCIENTIFIC LIBRARY (IMSL)

La biblioteca IMSL, contiene un gran nùmero de funciones y subrutinas que cumplen con diferentes propòsitos. La biblioteca està disponible en precisiòn sencilla (imsl.olb) y doble precisiòn (imsldp.olb).

(1) Informaciòn proporcionada por el Instituto Mexicano del Petròleo.

La biblioteca se encuentra particionada en paquetes y cada uno de ellos cubre cada uno de los temas. Para diferenciar un paquete de otro y accesar el tema que se desea, cada uno comienza con una letra relacionada con el tema, por ejemplo:

Anàlisis de Varianza: Todas las subrutinas de este paquete comienzan con la letra A.

Estadìstica Bàsica: Todas las subrutinas de este paquete comienzan con la letra B.

El paquete que maneja series de tiempo es:

PREDICCION; ECONOMETRIA, SERIES DE TIERMPO, TRANSFORMADAS

ECONOMETRIA, SERIES DE TIEMPO, TRANSFORMADAS (Todas las subrutinas de este paquete comienzan con la letra F).

Algunas de las subrutinas son:

---> Estimaciòn preliminar de paràmetros autorregresivos en un proceso estocàstico ARIMA.

---> Media, varianza, autocovarianza, autocorrelaciones y autocorrelaciones parciales para una serie de tiempo estacional.

---> Predicciòn de series de tiempo y lìmites de probabilidad usando un modelo ARIMA (Box-Jenkins).

y existen como estas otras màs.

B) NUMERICAL ALGORITHMS GROUP. (NAG)

La librerìa NAG, es una colecciòn de algoritmos para la soluciòn de problemas numèricos en computadora. Se encuentra dividida en capìtulos los cuales comprenden anàlisis numèrico ò estadìstico. Algunos de los temas que se manejan son:

---> Raìces de una ò màs ecuaciones trascendentales.

---> Sumatoria de series.

---> Ecuaciones diferenciales ordinarias.

---> Ecuaciones diferenciales parciales.

---> ANALISIS DE SERIES DE TIEMPO.

Cada capítulo tiene un nombre, de uno ó tres caracteres y un título, es decir

S aproximación de funciones espectrales
D02 Ecuaciones diferenciales ordinarias

El tema de interès, series de tiempo corresponde a la subrutina

G13 Anàlisis de series de tiempo

Algunas de las subrutinas para series univariadas son:

---> Diferencias estacionales y no estacionales.

---> Función de autocorrelación muestral.

---> Autocorrelación parcial muestral.

---> Estimación preliminar de un modelo ARIMA.

---> Actualización de un conjunto de datos para predicción.

etc...

En estas dos bibliotecas (IMSL y NAG), también existen subrutinas para series bivariadas que no se especifican , ya que no son de interès en esta tesis. El inconveniente que se presenta en el uso de estas bibliotecas es la necesidad de los argumentos en las subrutinas, ya que es necesario saber cuales necesita cada subrutina, cuales son de entrada y cuales de salida y que representa cada uno de ellos. Cada subrutina necesita de esta información para realizar los càlculos. Comparando con los paquetes para computadoras personales es una forma muy distinta de trabajar.

PAQUETES PARA PC.

Actualmente existe un grupo de programas de anàlisis estadístico llamado "THE STATISTIC SERIES"(*) para ser utilizados en APPLE II ó IBM PC.

HUMAN SYSTEM DYNAMICS se propuso proveer programas de anàlisis estadístico que satisfacieran las necesidades de investigación y de docencia. Dentro de STATISTIC SERIES existen

(*) Pertenece a Human System Dynamics"

una gran variedad de paquetes.

1.-PC REGRESSION

Este paquete provee múltiples rutinas de regresión para que se pueden utilizar, por ejemplo :

---> Microbiología: siguiendo la pista del vigor de las bacterias.

---> Indicadores financieros en busca de un mejor mercado.

---> Psicología: predicción de la conducta de un grupo.

Se pueden manipular los datos usando un amplio número de opciones para transformar los datos.

PC REGRESSION incluye "lead and lag" para el análisis de series de tiempo, medias móviles y suavizamiento exponencial.

PC REGRESSION ACEPTA:

2 a 37 variables por regresión.

creación del archivo de datos.

archivos del paquete PC STATISTICIAN

archivos del paquete PC ANOVA

archivos .dif de Lotus 1-2-3

archivos ASCII.

PC REGRESSION CALCULA:

Estadística descriptiva.

Correlación múltiple cuadrada.

Prueba t para coeficientes.

Matriz de correlación.

Varianza- matriz de covarianza.

Correlación mútiple.

Correlación parcial

Error estándar

PREDICTED SCORES

RESIDUAL SCORES

El costo de este paquete es de 200 dls.

2.-PC REGRESS II

Este paquete provee rutinas (profesionales) de regresión múltiple pudiendo ser utilizadas , por ejemplo por ingrenieros petroleros para predecir la buena localización del crudo.

Este paquete permite crear archivos, borrar, editar, crear subarchivos, buscar y seleccionar records y transformar datos.

Incluye transformaciones para ayudar en el análisis de series de tiempo

REGRESS II ACEPTA:

1 a 20 variables por archivo

Archivos con formatos de impresión VISICALC

CALCULA:

Prueba t para coeficientes

Matriz de correlación

Matriz de covarianza

Estadístico de Durbin-Watson

Standard Error

Correlación Serial.

Estos dos paquetes son parte del grupo de "statistic series", posiblemente no contengan el tema de análisis de series de tiempo ni tampoco el Método de Box and Jenkins pero sí proporcionan herramientas para realizar el análisis.

Por último cabe mencionar que también existe IMSL para PC's, y el paquete estadístico SPSS que también maneja series de tiempo.

CAPITULO 4

APLICACION Y ANALISIS COMPARATIVO

Hasta ahora se han explicado las bases teóricas en las que se apoya el análisis de Ingresos por Turismo así como los dos paquetes que se van a utilizar.

Este capítulo es importante ya que contiene el objetivo principal de esta tesis. Por otro lado, el fenómeno de ingresos por turismo muestra muy bien la forma de manejar los conceptos explicados en los capítulos anteriores, ya que se ilustra la existencia de tendencia, ciclos, patrones, etc.

En la ACF y PACF se observan picos en Marzo-Abril, Junio, Julio, Diciembre que se pueden interpretar como temporadas altas de vacaciones. Se puede observar también que año con año se repite un comportamiento (existe un patrón en los datos).

Para realizar este análisis se utilizan los paquetes FORECAST PLUS y STATGRAPHICS (capítulo 3), ésto tiene como objetivo comparar la eficiencia de cada uno. Se pretende que esta comparación tenga un logro didáctico, ya que dependiendo de la que ofrezca cada uno y de las necesidades a cubrir se pueda decidir cual de los dos es más conveniente usar.

Por último, se ha venido mencionando lo importante que es el juicio personal al realizar un análisis como éste, y es en este capítulo en donde se pone en práctica.

Al finalizar este capítulo se obtendrán 2 resultados:

---> El pronóstico de valores futuros realizado por los dos

paquetes y

---> Análisis comparativo de los paquetes.

Cabe señalar que no se pretende llegar a mostrar que un paquete sirve y el otro no, es decir, que sólo uno es bueno, si no que se pretende hacer notar que ventajas y desventajas existen en el uso de estos paquetes y que ofrece cada uno para realizar el análisis.

Las gráficas que se presentan en este capítulo corresponden a los dos paquetes de manera indistinta, para facilitar el análisis.

Los datos históricos disponibles son ingresos por turismo mensuales (dólares), a partir de enero de 1978 hasta 1988, la lista completa de éstos aparece en el capítulo 2.

4.1 ANALISIS INICIAL
(Identificación del modelo)

A) DATOS ORIGINALES.

Obsèrvese la figura 1a y figura 1b (gràfica del forecast plus)

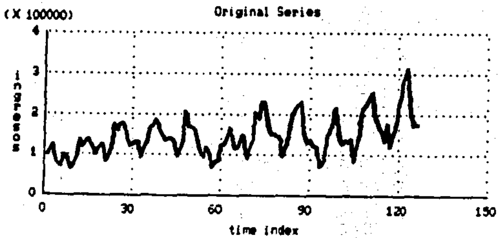


figura 1a

OBSERVACIONES:

1. Ligera tendencia creciente.

2. No se observa mucha variación de los datos con respecto a la media

3. Se observa la repetición de un comportamiento año con año. Esto es màs notorio a partir de 1984, existe el mismo comportamiento de septiembre a septiembre.

4. Los picos màs altos que se observan corresponden a los meses de FEBRERO, MARZO o ABRIL y DICIEMBRE.

5. Los picos màs bajos corresponden al mes de septiembe.

6. La repetición de picos refleja estacionalidad (estacionalidad multiplicativa), s=12 (12 puntos por cada ciclo).

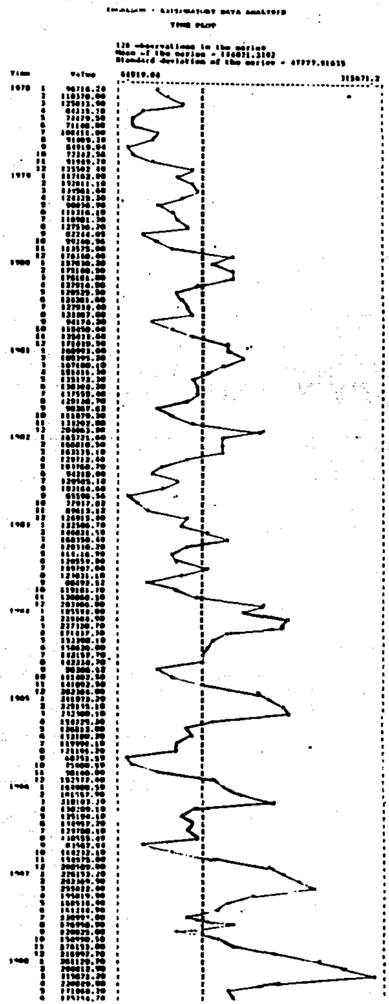

TIME PLOT
Time
Value
1970
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988

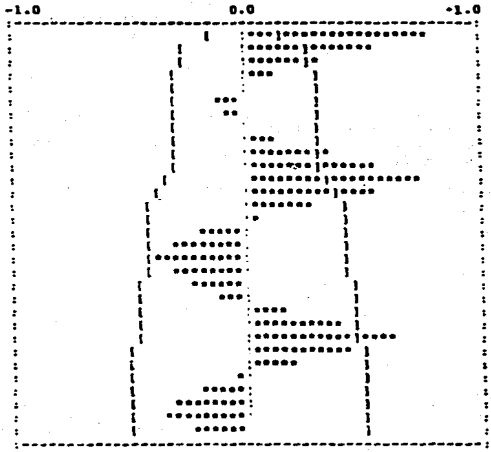
figura 1b

INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

AUTOCORRELATION FUNCTION

126 Observations in the Working Series
Mean of the Working Series = 144021.3
Standard Deviation of the Working Series = 47777.91

Lag	Value	T-Value
1	0.78	8.78
2	0.54	4.07
3	0.31	2.05
4	0.13	0.86
5	-0.01	-0.03
6	-0.12	-0.80
7	-0.07	-0.42
8	0.02	0.11
9	0.14	0.87
10	0.36	2.27
11	0.58	3.54
12	0.77	4.28
13	0.55	2.72
14	0.28	1.31
15	0.03	0.15
16	-0.18	-0.83
17	-0.32	-1.46
18	-0.42	-1.90
19	-0.34	-1.48
20	-0.23	-0.97
21	-0.10	-0.43
22	0.15	0.62
23	0.40	1.69
24	0.62	2.59
25	0.44	1.72
26	0.19	0.72
27	-0.03	-0.10
28	-0.22	-0.83
29	-0.31	-1.18
30	-0.36	-1.34
31	-0.25	-0.92



[ ] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 31 Degrees of Freedom = 516.6

A.1) FUNCION DE AUTOCORRELACION MUESTRAL

La gráfica de estas función se encuentran en la figura 2,

i) Los valores significativos son:

1,2,3,10,11,12,13,17,18,19,23,24,25

de éstos los que más sobrepasan del intervalo de confianza son:

1,2,11, 12, 13,24

El valor del estadístico t para los valores restantes a estos últimos es:

VALOR	ESTADISTICO t
17	-1.46
19	-1.48
23	1.69
25	1.72

como /t/ <= 2 tal vez se pueden considerar como cero. De los valores que se consideran significativos 1 y 2 muestran indicios de parte autorregresivos y 11,12 13, 23,24,25 muestran indicios de estacionalidad.

ii) Estadístico de Box Pierce con 31 grados de libertad =516.6

figura 2

A.2) FUNCION DE AUTOCORRELACION PARCIAL MUESTRAL

Figura 3:

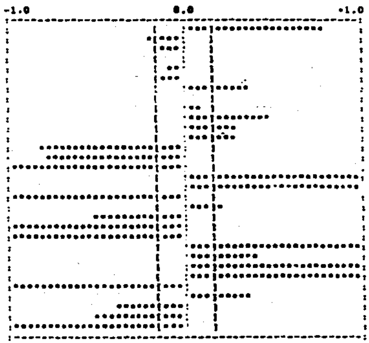
Esta gráfica NO muestra estacionaridad. NO decae a cero rapidamente. Esto indica que los datos no son estacionarios.

INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

PARTIAL AUTOCORRELATION FUNCTION

126 Observations in the Working Series
Mean of the Working Series = 144021.3
Standard Deviation of the Working Series = 47777.91

Lag	Value	T-Value
1	0.78	8.78
2	-0.18	-2.03
3	-0.14	-1.55
4	-0.02	-0.17
5	-0.08	-0.85
6	-0.11	-1.26
7	0.34	3.83
8	0.00	0.03
9	0.10	1.10
10	0.49	5.47
11	0.27	3.03
12	0.28	3.13
13	-0.86	-9.60
14	-0.81	-9.08
15	-1.00	-11.22
16	1.00	11.22
17	1.00	11.22
18	-1.00	-11.22
19	0.18	2.06
20	-0.54	-6.06
21	-0.99	-11.08
22	-1.00	-11.22
23	1.00	11.22
24	0.41	4.56
25	1.00	11.18
26	1.00	11.22
27	-1.00	-11.22
28	0.37	4.20
29	-0.41	-4.57
30	-0.53	-5.91
31	-1.00	-11.22



[ ] = Estimated Two-Standard Error Limits

figura 3

A.3) BOX PLOT

La têcnica de Box Plot refleja una estacionalidad multiplicativa, èsto se observa en la figura 4.

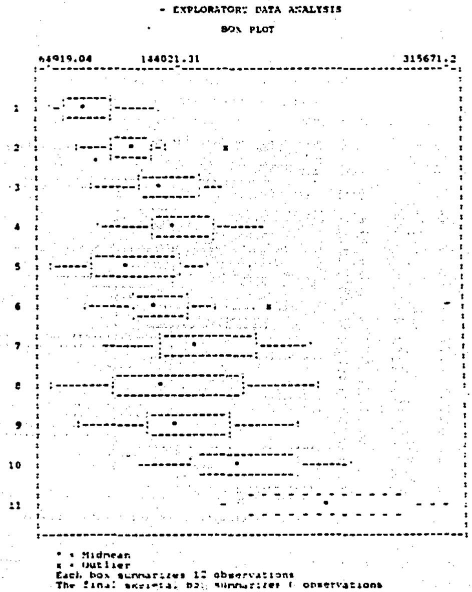
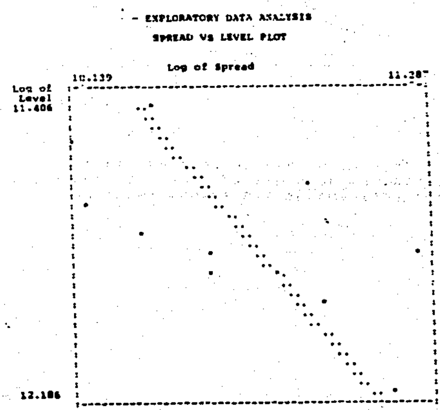


figura 4

Todo lo anterior sugiere realizar alguna transformación a los datos, ya que hasta el momento se puede concluir que la serie NO es estacionaria.

A.4) SPREAD VS. LEVEL PLOT

En la gráfica generada por esta técnica (Fig 5), aparece una serie de puntos formando una línea inclinada, ésto indica que la varianza no es estable en los datos. La transformación sugerida (tal vez no la mejor para este caso en particular) es tomar el logaritmo natural de los datos. Este será el paso a seguir despueśs de observar el periodograma integral de la sertie original. Así se podra observar si esta transformación es la ideal.



* Summarizes Subsets of 12 Consecutive Observations
Least Squares Estimate of Slope = 1.0576
Approximate Transformation to Stabilize Spread = Log

figura 5

A.5) PERIODOGRAMA INTEGRAL

Observando la figura 6, se observa que los datos no son estacionarios, el tener valores muy altos indica la presencia de tendencia, la existencia de saltos sistemáticos (escalones) indica estacionalidad.

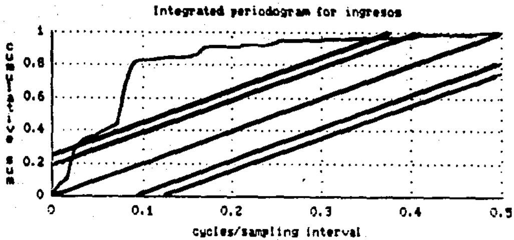


figura 6

De este primer reconocimiento se puede concluir que se tienen datos con tendencia, estacionalidad y no estacionarios. Por lo que tendrán que realizar diferencias y/ó transformaciones.

B) LOGARITMO NATURAL DE LOS DATOS

La gráfica generada (time plot) del logaritmo natural se puede observar en la figura 7.

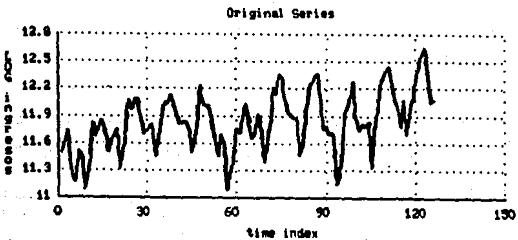


figura 7

OBSERVACIONES:

1.- Comparando esta gráfica con la de los datos originales, se observa que no existe un cambio significativo en el comportamiento de los datos.

2.- Se sigue observando un patrón en los datos, es decir, se observa estacionalidad.

3.- La varianza no es estable.

4.- Se elimina un poco la tendencia.

8.1) FUNCION DE AUTOCORRELACION MUESTRAL

Figura 8.

i. Los valores significativos son:

1,2,11,12,13,18,24

estos valores, practicamente son los mismos que la ACF de la serie original.

ii. La gráfica decae a cero, no muy rapidamente.

iii. Estadístico de Box-Pierce con 31 grados de libertad = 467.1

Disminuye con respecto al anterior pero todavía es bastante grande ya que:

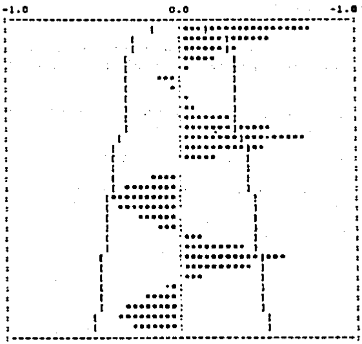
$$Q > X^2_{31} \Rightarrow \text{no se puede considerar como ruido blanco.}$$

Log of INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

AUTOCORRELATION FUNCTION

125 Observations in the Working Series
Mean of the Working Series = 11.8253
Standard Deviation of the Working Series = .325341

Lag	Value	T-Value
1	0.76	6.55
2	0.52	3.93
3	0.30	2.07
4	0.18	1.20
5	0.04	0.23
6	-0.11	-0.71
7	-0.05	-0.35
8	0.03	0.22
9	0.10	0.65
10	0.30	1.96
11	0.53	3.33
12	0.73	4.25
13	0.49	2.52
14	0.21	1.04
15	-0.02	-0.09
16	-0.18	-0.85
17	-0.32	-1.54
18	-0.44	-2.09
19	-0.36	-1.65
20	-0.23	-1.03
21	-0.14	-0.62
22	0.10	0.45
23	0.35	1.56
24	0.58	2.52
25	0.38	1.58
26	0.13	0.51
27	-0.07	-0.28
28	-0.20	-0.80
29	-0.30	-1.21
30	-0.37	-1.49
31	-0.27	-1.04



[ ] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 31 Degrees of Freedom = 467.1

figura 8

B.2) FUNCION DE AUTOCORRELACION PARCIAL MUESTRAL.

La figura 9 muestra el comportamiento de esta funciôn.

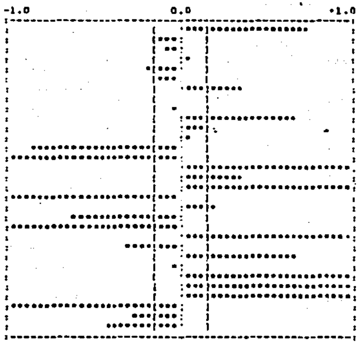
Esta grâfica NO muestra estacionaridad, no decae a cero rapidamente. No se vè un cambio significativo con respecto a la PACF de la serie original.

Log of INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

PARTIAL AUTOCORRELATION FUNCTION

126 Observations in the Working Series
Mean of the Working Series = 11.8251
Standard Deviation of the Working Series = .325341

Lag	Value	T-Value
1	0.76	8.55
2	-0.16	-1.76
3	-0.08	-0.89
4	0.06	0.67
5	-0.19	-2.13
6	-0.12	-1.38
7	0.38	4.23
8	-0.02	-0.17
9	-0.05	-0.52
10	0.66	7.43
11	0.17	1.96
12	0.06	0.64
13	-0.87	-9.79
14	-1.00	-11.22
15	1.00	11.22
16	0.36	4.01
17	1.00	11.22
18	-1.00	-11.22
19	0.21	2.34
20	-0.63	-7.02
21	-1.00	-11.22
22	1.00	11.22
23	-0.31	-3.50
24	0.67	7.57
25	-0.04	-0.45
26	1.00	11.22
27	1.00	11.22
28	1.00	11.22
29	-1.00	-11.22
30	-0.27	-3.01
31	-0.46	-5.11



[ ] = Estimated Two-Standard Error Limits

figura 9

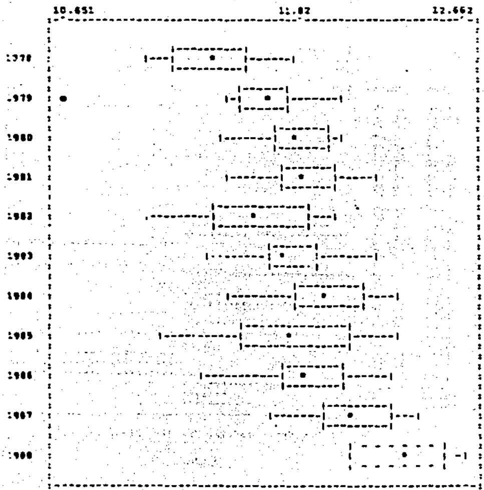
B.3) BOX-PLOT.

Obsèrvese la figura 10.

Presencia de estacionalidad. Se observa una media y varianza constante.

Log of INGRESOS - EXPLORATORY DATA ANALYSIS

BOX PLOT



* = Midmean
o = Extreme Outlier
Each box summarizes 12 observations
The final partial box summarizes [illegible] observations

FIGURA 10

B.4) SPREAD VS. LEVEL PLOT

Obsérvese la figura 11.

La línea vertical muestra que los datos son estacionarios. Esta transformación mejora el modelo.

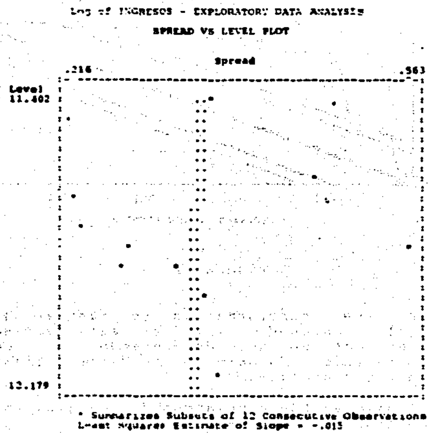


figura 11

B.5) PERIODOGRAMA INTEGRAL

Figura 12.

Comparando con el periodograma de los datos originales (figura 6) no existe mucha diferencia, se sigue observando tendencia y los datos no son estacionarios, así como saltos escalonados indicando estacionalidad en los datos.

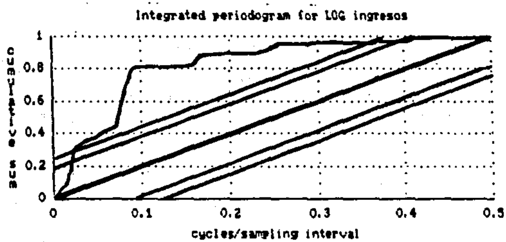


figura 12

Al analizar esta transformación se observa que no es la transformación màs adecuada, no es mala pudiendo combinarse posteriormente con diferencias, dependiendo de como se vaya presentando en análisis. No mejoran los datos y todavía no se tiene una idea de un modelo tentativo.

El siguiente paso es tomar primeras diferencias ordinarias para quitar esa ligera tendencia que se observa.

C) PRIMERAS DIFERENCIAS ORDINARIAS

La gráfica de las primeras diferencias se observa en la figura 13

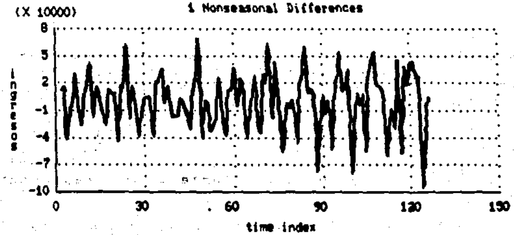


figura 13

OBSERVACIONES:

1. No existe tendencia
2. Media y varianza constante
3. Se observa estacionalidad. Se repite el mismo patrón cada año.

C.1) FUNCION DE AUTOCORRELACION MUESTRAL.

Observar la figura 14 .

Los valores significativos son:

3,6,12,18,24

Observando la funciôn , va mejorando en comparaciôn con las anteriores; los valores de 3,6,18 no son sumamente grandes por lo que pueden no ser significativos para el modelo. Los valores de 12 y 24 indican la presencia de un comportamiento estacional.

i. La distancia entre los dos valores significativos es de 12 perîodos , lo cual sugiere un span=12. Esta funciôn sugiere un modelo SAR ô SMA, observando que se trunca en Q=4. Esto se deberà tomar en cuenta si la PACF refleja un comportamiento que represente a uno de estos modelos.

ii. El estadîstico de Box-Pierce con 31 grados de libertad = 260.8 es considerablemente menor al del logartimo de los datos con los mismos grados de libertad.

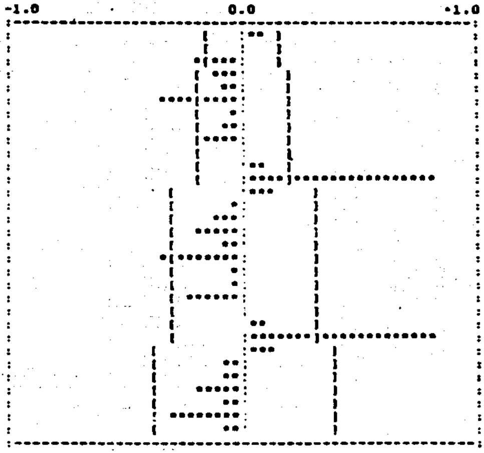
INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

AUTOCORRELATION FUNCTION

125 Observations in the Working Series
Mean of the Working Series = 632.3076
Standard Deviation of the Working Series = 31138.38

Degree of Regular Differencing = 1

Lag	Value	T-Value
1	0.07	0.78
2	-0.02	-0.22
3	-0.19	-2.11
4	-0.13	-1.37
5	-0.06	-0.67
6	-0.36	-3.79
7	-0.04	-0.38
8	-0.07	-0.70
9	-0.19	-1.84
10	0.00	0.00
11	0.09	0.85
12	0.84	7.74
13	0.12	0.82
14	-0.06	-0.36
15	-0.11	-0.74
16	-0.20	-1.27
17	-0.07	-0.45
18	-0.35	-2.26
19	-0.03	-0.18
20	-0.02	-0.15
21	-0.22	-1.37
22	-0.01	-0.08
23	0.08	0.49
24	0.83	4.99
25	0.13	0.68
26	-0.08	-0.40
27	-0.10	-0.49
28	-0.21	-1.04
29	-0.09	-0.45
30	-0.32	-1.59
31	-0.07	-0.35



[ ] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 31 Degrees of Freedom = 260.8

figura 14

C.2) FUNCION DE AUTOCORRELACION PARCIAL MUESTRAL

Al observar la figura 15.

Se vè una mejorìa en comparaciòn con las dos anteriores. El valor del perìodo 28 es grande y està muy lejos, por lo que en la etapa de estimaciòn podrìa eliminarse sin haberse tomado en cuenta.

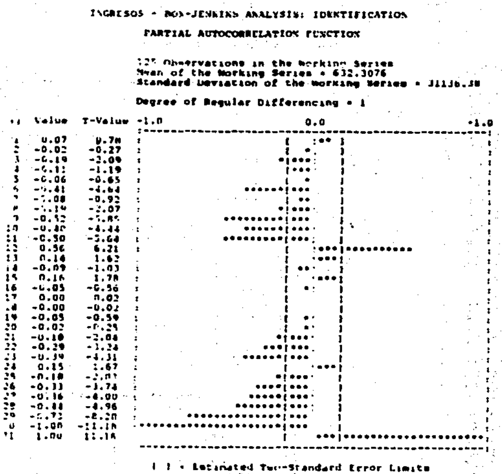


figura 15

C.3) BOX PLOT

Se observa (fig 16) que los datos fluctúan un poco más alrededor de la media; se observa estacionalidad.

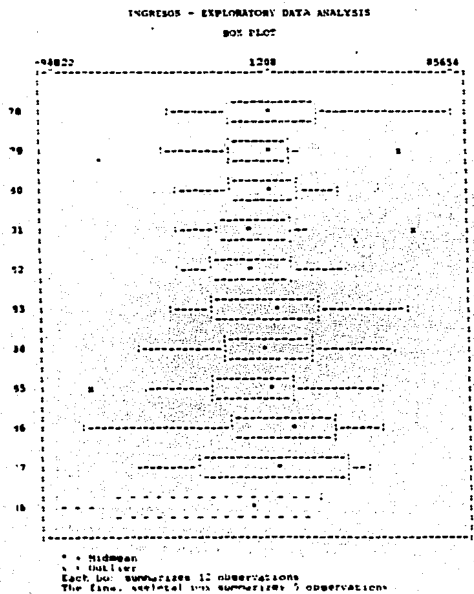


figura 16

C.4) SPREAD VS. LEVEL PLOT

En la gráfica generada (fig 17) se observa una línea ligeramente inclinada. Los datos se pueden considerar estacionarios.

Aunque esta tècnica no sugiere una transformaciòn, es necesario seguir transformando los datos para llegar a "modelos tentativos".

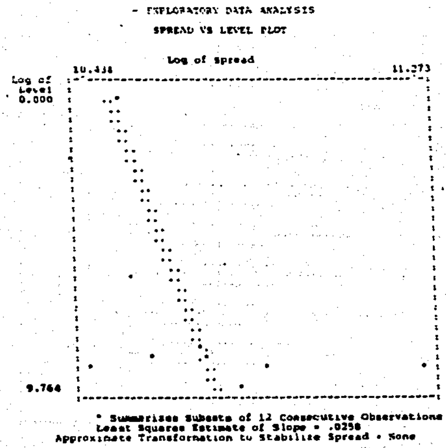


figura 17

C.5) PERIODOGRAMA INTEGRAL

Figura 18; Se observan los datos alrededor de la media,los picos que se observan son signo de estacionalidad. La figura muestra que se eliminó la tendencia.

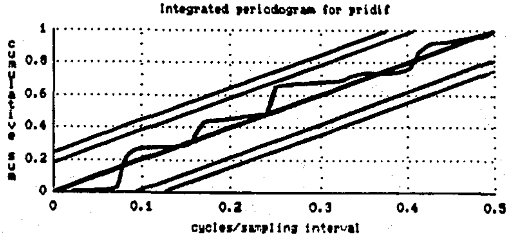


figura 18

En este paso se eliminó la tendencia y se estacionarizó la serie.

D) PRIMERAS DIFERENCIAS ESTACIONALES

Tomando un "span" s=12 la gràfica de datos se presenta en la figura 19.

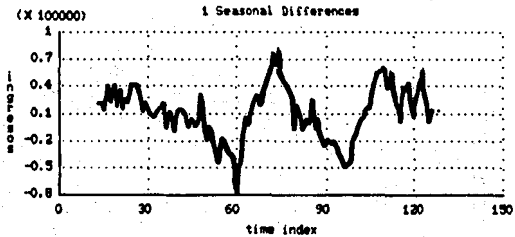


figura 19

OBSERVACIONES:

1. La media es constante

2. La varianza es constante; existe variación pero no es significativa.

3. Se elimina la estacionalidad

4. Practicamente la serie es estacionaria ya que tiene media y varianza constante.

D.1) FUNCION DE AUTOCORRELACION MUESTRAL

Figura 20,

i. La función de autocorrelación decae a cero muy lentamente.

ii. El estadístico de Box-Pierce con 28 grados de libertad, Q = 611.3, es demasiado grande.

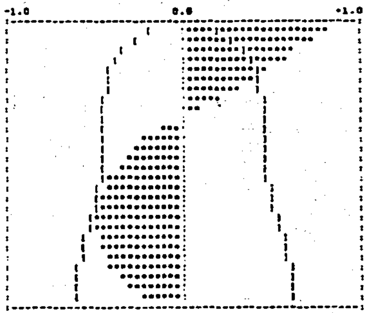
INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

AUTOCORRELATION FUNCTION

114 Observations in the Working Series
Mean of the Working Series = 11351.72
Standard Deviation of the Working Series = 30189.55

Degree of Seasonal Differencing = 1

Lag	Value	T-Value
1	0.84	8.96
2	0.75	5.19
3	0.66	3.72
4	0.59	2.99
5	0.50	2.35
6	0.40	1.81
7	0.31	1.35
8	0.21	0.89
9	0.10	0.42
10	0.01	0.03
11	-0.10	-0.43
12	-0.25	-1.07
13	-0.29	-1.22
14	-0.33	-1.40
15	-0.38	-1.54
16	-0.43	-1.73
17	-0.46	-1.80
18	-0.49	-1.88
19	-0.50	-1.86
20	-0.54	-1.95
21	-0.51	-1.79
22	-0.48	-1.66
23	-0.48	-1.59
24	-0.44	-1.43
25	-0.42	-1.34
26	-0.37	-1.16
27	-0.31	-0.97
28	-0.25	-0.76



[ ] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 28 Degrees of Freedom = 636.1

figura 20

D.2) FUNCION DE AUTOCORRELACION PARCIAL MUESTRAL

Figura 21.

Los valores significativos son

1,12,23

El valor 1 indica la presencia de parte autorregresiva ó medias móviles. Los valores 12y 23 indican presencia estacional, sugieren un modelo estacional autorregresivo y/ó medias móviles con s=12, es decir, SAR(1), SMA(1) ó SARMA(1,1).

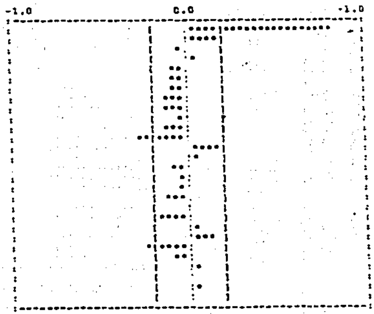
INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

PARTIAL AUTOCORRELATION FUNCTION

114 Observations in the Working Series
Mean of the Working Series = 11351.72
Standard Deviation of the Working Series = 30185.55

Degree of Seasonal Differencing = 1

Lag	Value	T-Value
1	0.84	8.96
2	0.17	1.84
3	-0.05	-0.49
4	0.03	0.36
5	-0.07	-0.70
6	-0.10	-1.05
7	-0.07	-0.73
8	-0.11	-1.18
9	-0.12	-1.30
10	-0.05	-0.56
11	-0.14	-1.45
12	-0.30	-3.18
13	0.19	2.07
14	0.02	0.25
15	-0.08	-0.89
16	-0.04	-0.38
17	-0.04	-0.41
18	-0.11	-1.21
19	0.00	0.02
20	-0.21	-2.23
21	0.05	0.57
22	0.10	1.12
23	-0.24	-2.60
24	-0.09	-0.91
25	0.05	0.51
26	-0.01	-0.09
27	0.03	0.29
28	0.01	0.09



[ ] = Estimated Two-Standard Error Limits

figura 21

D.3) BOX-PLOT

La gráfica de box-plot corresponde a la figura 22. Claramente se observa que la serie tiene una media constante y una varianza constante.

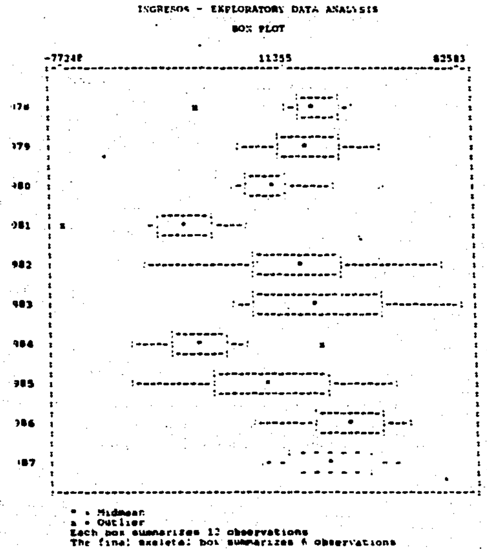


figura 22

D.4) SPREAD VS. LEVEL PLOT

En la figura 23 se observa una banda ligeramente inclinada. La varianza es dependiente del nivel; al igual que en la figura 17, no se sugiere ninguna transformación, pero se concluye que el utilizar unicamente diferencias estacionales no es el mejor camino.

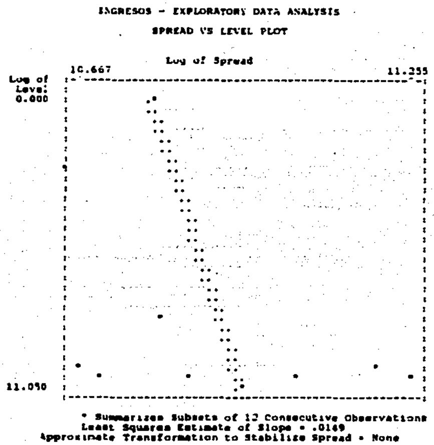


figura 23

D.5) PERIODOGRAMA INTEGRAL

La figura 24 muestra un modelo con tendencia. No se refleja estacionalidad ya que se eliminó al realizar las diferencias estacionales. Se puede conluir que el modelo ya no es estacional pero tiene una tendencia creciente.

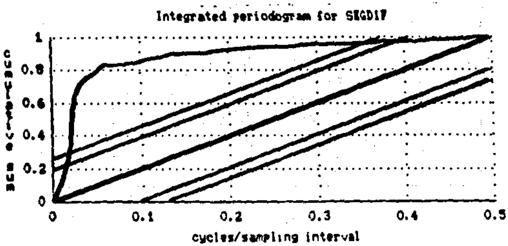


figura 24

Hasta el momento se ha visto que al hacer transformaciones y/ó diferencias de manera separada se elminan ciertas cosas pero se acentúan otras y no se ha logrado hacer estacionario el modelo. El siguiente paso es combinar transformaciones con diferencias.

Hasta aquí , el tomar diferencias ordinarias (primeras diferencias) ha dado como resultado un mejor comportamiento de los datos. Se empezará trabajando con estas diferencias.

Se observó también que el tomar primeras diferencias ordinarias elimina la tendecia creciente que presentan los datos originales y el tomar diferencias estacionales elimina la estacionalidad, por lo que es obvio pensar que tomar ambas mejore el modelo y tal vez sea el definitivo para continuar con estimación. La transformación de logartimo se descarta ya que no presentó la mejor opción.

E) DIFERENCIAS ORDINARIAS Y DIFERENCIAS ESTACIONALES.

La figura 25 muestra el comportamiento de los datos con primeras diferencias ordinarias y primeras diferencias estacionales.

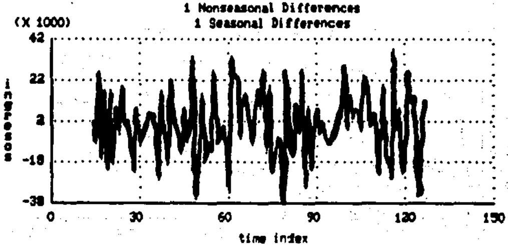


figura 25

OBSERVACIONES:

1. Se observa media y varianza constante.
2. No hay tendencia.
3. No hay estacionalidad.

E.1) FUNCION DE AUTOCORRELACION MUESTRAL.

Observando la figura 26, los valores significativos son:

1,12

i. La funciòn disminuye rapidamente a cero, mostrando estacionalidad.

ii. Sugiere un modelo AR(1) x SAR(1), el tèrmino autorregresivo ordinario lo sugiere el valor significativo 1.

iii.Si se toma que esta funciòn se trunca despuès de 12, se puede pensar en un modelo AR(1) x SMA(1). Recuèrdese que la ACF de un SMA se trunca en sQ, en este caso Q=1

iv. Estadìstico de Box-Pierce con 28 grados de libertad = 43.6

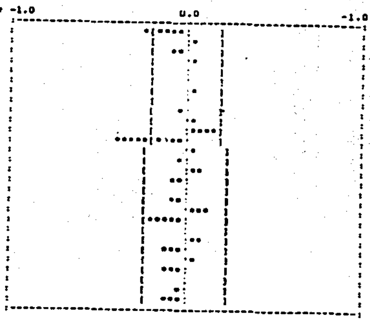
INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

AUTOCORRELATION FUNCTION

112 Observations in the Working Series
Mean of the Working Series =-52.5376
Standard Deviation of the Working Series = 16811.89

Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1

Lag	Value	T-Value
1	-0.22	-2.35
2	0.05	0.51
3	-0.08	-0.80
4	0.05	0.48
5	0.01	0.09
6	-0.01	-0.06
7	0.03	0.26
8	0.02	0.18
9	-0.06	-0.59
10	0.05	0.51
11	0.14	1.40
12	-0.41	-4.03
13	0.03	0.26
14	-0.04	-0.33
15	0.06	0.55
16	-0.10	-0.85
17	0.00	0.03
18	-0.07	-0.62
19	0.10	0.87
20	-0.22	-1.86
21	0.01	0.06
22	0.09	0.74
23	-0.10	-0.84
24	0.05	0.41
25	-0.10	-0.84
26	-0.00	-0.02
27	-0.04	-0.32
28	-0.12	-1.00



[ ] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 28 Degrees of Freedom = 43.6
Probability = .031

 figura 26

E.2) FUNCION DE AUTOCORRELACION PARCIAL MUESTRAL

Observar la figura 27;

Los valores significativos son:

1,12,24,28

El valor 28, no tiene importancia para el modelo y se puede asegurar que el la etapa de estimación y diagnóstico este valor se eliminará.

Esta función disminuye a cero con valores significativos en múltiplos de 12:

el valor de 12 es -0.39

el valor de 24 es -0.23

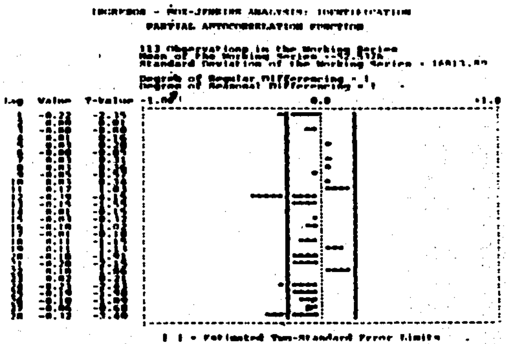


figura 27

si se grafica el valor de 36 será más pequeño, como se observa en la figura 27a.

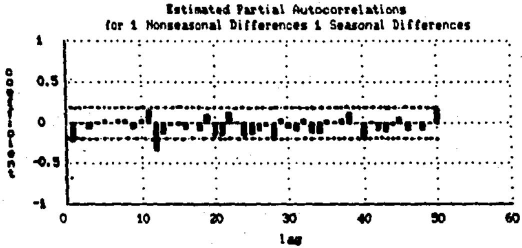


figura 27a

Este comportamiento sugiere un modelo multiplicativo AR(1) x SMA(1); recuérdese que la PACF de un modelo SMA disminuye a cero con valores significativos en múltiplos de "s".

E.3) BOX-PLOT

La figura 28 muestra un modelo estacionario, no refleja estacionalidad y el comportamiento es alrededor de la media.

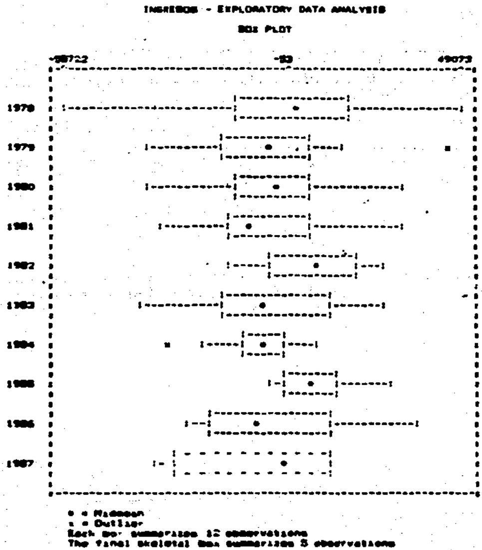


figura 28

E.4) SPREAD VS. LEVEL PLOT.

Observese que la figura 29, muestra una linea totalmente vertical, sin ninguna variación, ésto hace suponer que hasta el momento éste es el mejor modelo.

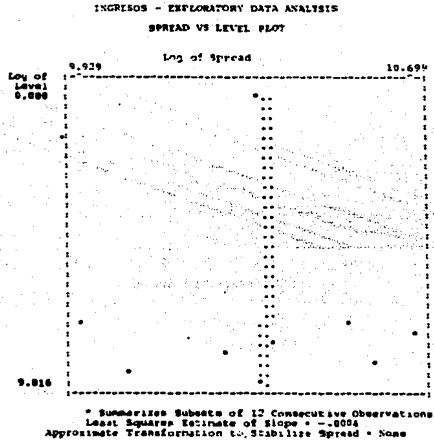


figura 29

E.5) PERIODOGRAMA INTEGRAL.

Figura 30: No muestra tendencia, el comportamiento de los datos es mejor con respecto a las gràficas del periodograma integral de los pasos anteriores.

Ligeros picos que pueden indicar presencia de estacionalidad.

Refleja un comportamiento un poco alejado de la media.

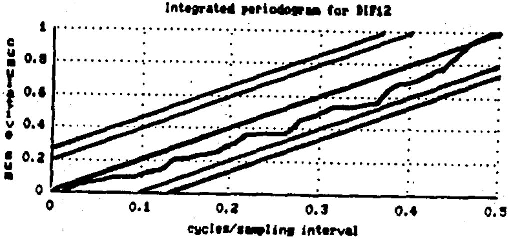


figura 30

Hasta el momento se puede concluir que se tienen dos modelos tentativos:

AR(1) x SAR(1)
AR(1) x SMA(1)

Con este paso la serie ya no contiene tendencia (observese comportamiento del periodograma integral). Todavia queda pendiente si ya se ha eliminado por completo la estacionalidad, que al parecer tambièn quedò eliminada. Para verificar èsto, se analizaràn los datos con primeras diferencias ordinarias y segundas estacionales.

F) PRIMERAS DIFERENCIAS ORDINARIAS,

SEGUNDAS DIFERENCIAS ESTACIONALES.

El comportamiento de los datos después de obtener primeras diferencias ordinarias y segundas estacionales se observa en la figura 31.

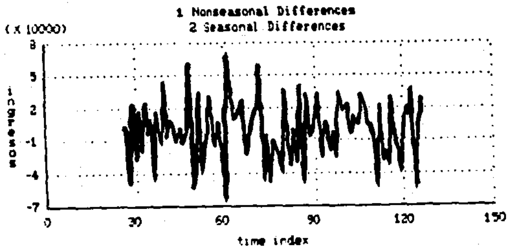


figura 31

OBSERVACIONES:

1. Refleja una serie estacionaria.

2. No hay tendencia.

3. No hay estacionalidad.

Comparando con la gráfica anterior (las ord. y las estac.), en ésta se observa que está sobrediferenciada, ya que presenta ciertos picos que en la otra no aparecen.

F.1) FUNCION DE AUTOCORRELACION MUESTRAL

FIGURA 32.

Los valores significativos son:

1,12

i. Se observa un valor grande en el período 20, lo cual no sucede en la ACF de la figura 26.

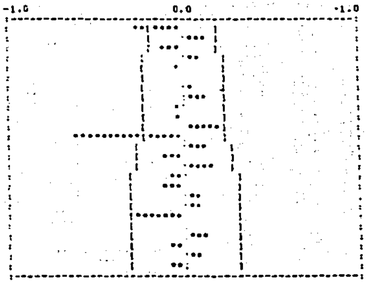
ii. Estadístico de Box-Pierce con 25 grados de libertad=79.6

INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

AUTOCORRELATION FUNCTION

101 Observations in the Working Series
Mean of the Working Series =-308.91
Standard Deviation of the Working Series = 31376.54

Lag	Value	T-Value
1	-0.29	-2.93
2	0.12	1.08
3	-0.13	-1.17
4	0.09	0.80
5	-0.06	-0.53
6	-0.01	-0.06
7	0.05	0.41
8	0.11	0.95
9	-0.03	-0.21
10	-0.04	-0.34
11	0.20	1.73
12	-0.62	-5.47
13	0.14	0.93
14	-0.12	-0.80
15	0.18	1.21
16	-0.08	-0.50
17	-0.11	-0.74
18	0.07	0.46
19	0.07	0.49
20	-0.28	-1.82
21	-0.00	-0.02
22	0.11	0.69
23	-0.08	-0.52
24	0.10	0.62
25	-0.08	-0.50



[ ] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 25 Degrees of Freedom = 79.6
Probability = 0

figura 32

F.2) FUNCION DE AUTOCORRELACION PARCIAL MUESTRAL

FIGURA 33.

Los valores significativos son:

1,12,13,19,22,24,25

Existen valores grandes que, hasta el momento, no habían aparecido y que complican el análisis del modelo.

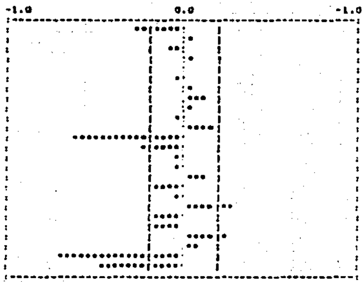
Por otro lado algunos de estos valores no tienen explicación para el modelo.

INGRESOS - BOX-JENKINS ANALYSIS: IDENTIFICATION

PARTIAL AUTOCORRELATION FUNCTION

101 Observations in the Working Series
Mean of the Working Series =-308.91
Standard Deviation of the Working Series = 11378.34

Lag	Value	T-Value
1	-0.29	-2.93
2	0.03	0.35
3	-0.09	-0.93
4	0.03	0.27
5	-0.02	-0.18
6	-0.05	-0.46
7	0.05	0.50
8	0.14	1.39
9	0.04	0.36
10	-0.04	-0.42
11	0.21	2.14
12	-0.63	-6.34
13	-0.26	-2.59
14	-0.05	-0.48
15	-0.04	-0.36
16	0.10	1.05
17	-0.22	-2.21
18	-0.04	-0.36
19	0.28	2.82
20	-0.17	-1.74
21	-0.21	-2.16
22	0.25	2.54
23	0.08	0.77
24	-0.73	-7.38
25	-0.50	-5.02



[ ] = Estimated Two-Standard Error Limits

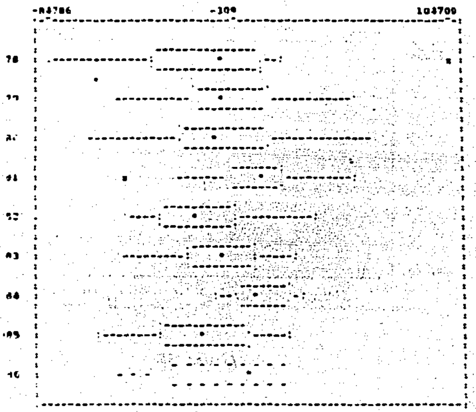
figura 33

F.3) BOX PLOT

La figura 34, muestra una serie estacionaria, no se observa variación estacional.

INGRESOS - EXPLORATORY DATA ANALYSIS

BOX PLOT



* = Midmean
x = Outlier
Each box summarizes 12 observations
The final skeletal box summarizes [illegible] observations

figura 34

F.4) SPREAD VS. LEVEL PLOT

Figura 35, se observa una línea vertical, los datos son estacionarios.

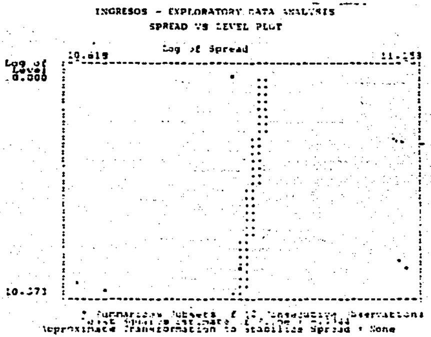


figura 35

F.5) PERIODOGRAMA INTEGRAL.

El periodograma que muestra la figura 36 no varía significativamente en comparación con el de la figura 30. Unicamente éste, presenta las mismas variaciones un poco más pronunciadas mientras que el anterior el comportamiento es más suave.

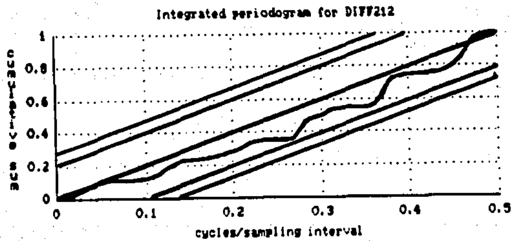


figura 36

Como conclusión se puede decir que el modelo está sobre diferenciado y no es el mejor para el análisis del fenómeno, ya que hay que tener presente que el camino a seguir debe ser lo más sencillo posible.

El resultado de esta primera etapa es que se debe utilizar el modelo E (primeras diferencias ordinarias, primeras diferencias estacionales).

Los modelos "tentativos" resultantes son:

AR(1) x SAR(1)
AR(1) x SMA(1)

4.2 ESTIMACION Y DIAGNOSTICO

Para iniciar esta etapa se empezarâ tomando el modelo más sencillo y observando el comportamiento de la ACF y PACF de los residuales, así como los resultados estadísticos que proporcionan los paquetes, paralelamente se realizará el diagnóstico de cada uno de los modelos. Con el fin de verificar los valores presentados en este capítulo, en el apêndice se encuentran los datos tal cual los presentan los paquetes.

Retomando los resultados de la etapa anterior, los modelos tentativos son:

AR(1) x SAR(1)

AR(1) x SMA(1)

En el transcurso de la realización de este capítulo, se tuvo la oportunidad de trabajar con la versión 2.1 del STATGRAPHICS (1) (siguiente versión a la explicada en el capítulo 3) la cual presenta cambios y mejoría.

Al realizar las estimaciones con esta nueva versión y compararlas con las de la versión anterior se notó que éstas variaban, y al compararlas con FORECAST PLUS las estimaciones nuevas eran muy semejantes. Para ejemplificar este comentario obsêrvese el caso AR(1).

A) AR(1)

Este es el modelo autorregresivo más sencillo que se puede tener.

Los resultados de estimación son:

FORECAST PLUS

PARAMETRO	VALOR	"t"
$\emptyset_1$	-0.2236	0.00

(1) a partir del modelo (8), las estimaciones se realizaron con esta nueva versión.

$SCR = 3.008303E10$

$S^2 = 2.406642E8$

$B\text{-}P = 33.4$

Nôtese que los valores estimados de SCR, S^2 son mucho muy grandes y el valor de "t" es 0.00. Esto se debe a que el paquete no soporta datos tan grandes. Para resolver este problema se dividieron los datos originales entre 100. Los resultados de estimación son:

PARAMETRO	VALOR	"t"
ϕ	-0.2236	-2.56

$SCR = 3008303$

$S^2 = 24066.42$

$B\text{-}P = 33.4$, $X^2 = 40.1133$

DIAGNOSTICO:

---> El valor de "t" indica que el parámetro autorregresivo sí es significativo, ya que / t / > 2.

---> La ACF de los residuales (figura 38) indica que falta por considerar parte estacional en el modelo.

---> 33.4 < 40.1133, las autocorrelaciones no significativas pueden considerarse como cero.

INGRESOS - BOX-JENKINS ANALYSIS
PARAMETER ESTIMATES AND STATISTICS

112 Observations Used for Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

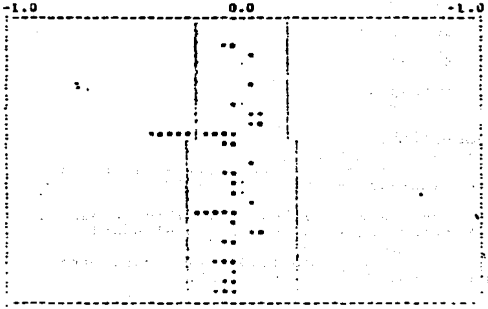
Sum of Squared Errors = 2698203
Residual Variance = 24060.42

Parameter	Value	T-Statistic
AR(1)	-0.2226	-2.36

INGRESOS - BOX-JENKINS ANALYSIS
RESIDUAL AUTOCORRELATION FUNCTION

112 Observations in the Residual Series
Mean of the Residual Series = -1.062998
Standard Deviation of the Residual Series = 164.6229

Lag	Value	T-Value
1	0.10	0.31
2	-0.03	-0.26
3	-0.07	-0.69
4	0.03	0.33
5	0.02	0.21
6	-0.03	-0.33
7	0.05	0.48
8	0.01	0.13
9	-0.03	-0.34
10	0.07	0.74
11	0.07	0.71
12	-0.41	-4.24
13	-0.07	-0.64
14	[illegible]	[illegible]
15	[illegible]	[illegible]
16	[illegible]	-0.50
17	-0.04	[illegible]
18	-0.05	-0.49
19	0.04	0.38
20	[illegible]	-1.35
21	[illegible]	[illegible]
22	[illegible]	[illegible]
23	[illegible]	[illegible]
24	[illegible]	[illegible]
25	[illegible]	[illegible]
26	[illegible]	[illegible]
27	[illegible]	[illegible]
28	[illegible]	[illegible]



[illegible] Estimated Two-Standard Error Limits

[illegible] Chi-Square Statistic with 27 Degrees of Freedom = [illegible]
Probability = [illegible]

figura 38

STATGRAPHICS

ESTIMACION:

PARAMETRO	VALOR	=t=
$\emptyset_1$	- 0.31403	-3.46231
MEDIA	-76.59663	-.05856
CTE.	-69.03640	

S^2 = 3.32122E8

X^2 = 27.0836

Los valores estimados por STATGRAPHICS versión 2.1 son:

PARAMETRO	VALOR	=t=
$\emptyset_1$	-0.22247	-2.3985
MEDIA	-67.3057	-0.05302
CTE.	-82.2794	

SCR = 3.00816E10
S^2 = 2.71006E8
X^2 = 21.8291

DIAGNOSTICO:

----> El parámetro AR sí es significativo.

----> La constante no es significativa en el modelo.

----> La media no es significativa en el modelo.

----> La ACF (figura 39) y la PACF (figura 40) de los residuales presentan valor significativo en 12. Por el comportamiento que se observa se puede decir que :

1.- La ACF decae a cero y la PACF se trunca después de 12, ésto sugiere un modelo SAR.

2.- La ACF se trunca después de 12 y la PACF decae a cero, ésto sugiere un modelo SMA.

----> Se observa que se eliminó el valor grande en el "lag" 1.

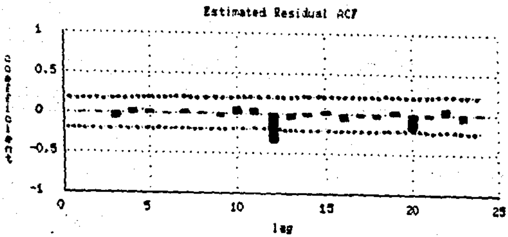


figura 39

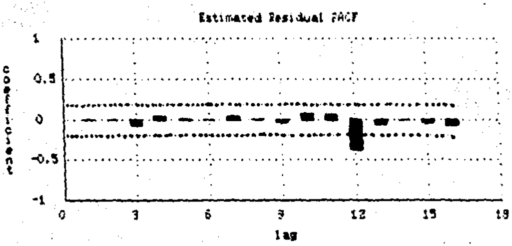


figura 40

PERIODOGRAMA INTEGRAL

En el periodograma integral (figura 41) se observan picos que se repiten, refleja estacionalidad.

El comportamiento de los residuales no sale de los límites de Kolmogorov-Smirnov, lo cual indica que el ajuste no es tan malo.

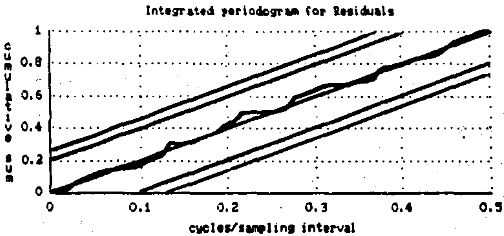


figura 41

B) SAR(1)

Este es el modelo autorregresivo estacional más sencillo.

<u>**FORECAST PLUS**</u>

ESTIMACION:

PARAMETRO	VALOR	=t=
Φ_1	-0.4293	-4.65

SCR= 2448564

S^2= 19588.51

B-P= 27.3 CON 24 g.l

DIAGNOSTICO:

---> El valor de =t= indica que 0 sí es significativo.

---> La ACF de los residuales (figura 42), ya no refleja un comportamiento estacional ya que éste fuê considerado en el modelo. El valor que se podría considerar alto es el del período 1, el valor de =t= correspondiente a ese período es -1.76 (relativamente significativo). Este valor alto sugiere considerar en el modelo parte autorregresiva.

---> $X^2_t > X^2_c$, 27.3 < 36.4150; las autocorrelaciones de los residuales pueden considerarse como cero.

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

1[illegible] Observations Used For Parameter Estimates
Degree of Regular Differencing = [illegible]
Degree of Seasonal Differencing = [illegible]
Seasonal Period = 12

Sum of Squared Errors = [illegible]8264
Residual Variance = 19[illegible]8.[illegible]

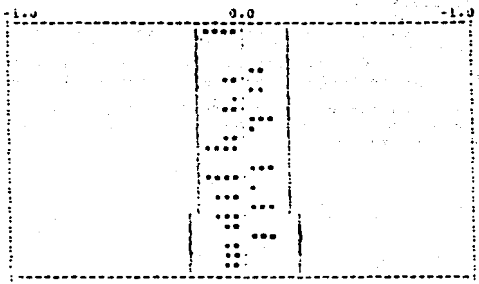
Parameter	Value	T-Statistic
SAR(1)	-0.4293	-4.65

INGRESOS - BOX-JENKINS ANALYSIS

RESIDUAL AUTOCORRELATION FUNCTION

11[illegible] Observations in the Residual Series
Mean of the Residual Series = -3.087894
Standard Deviation of the Residual Series = [illegible]56.4481

Lag	Value	T-Value
1	-0.17	-1.76
[illegible]	[illegible]	[illegible]
9	-0.09	-0.88
10	[illegible]	[illegible]
[illegible]	-0.07	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]5	-0.07	[illegible]



( ) = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with [illegible] Degrees of Freedom = [illegible]
Probability = [illegible]

figura 42

STATGRAPHICS

ESTIMACION:

PARAMETRO	VALOR	=t=
Φ_1	- 0.43664	-4.53242
MEDIA	-60.19718	-0.05749
CTE	-86.4815	

SCR = 2.66988E10

S^2 = 2.4053E8 CON 111 g.l

X^2 = 20.0084

Probabilidad de que los valores significativos sean ruido blanco= 0.394074

DIAGNOSTICO:

---> El parámetro Φ_1 sí es importante en el modelo ya que /t/=/-4.53242/ > 2.

---> La media y la constante no son significativas en el modelo.

---> La SCR posiblemente sea cercana a la estimada por FORECAST PLUS ya que la proporcionada por STATGRAPHICS no es el último valor.

---> La ACF de los residuales (figura 43) y la PACF (figura 44), muestran un valor grande en 1, sugieren considerar parte autorregresiva.

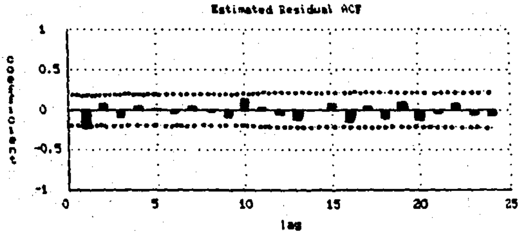


figura 43

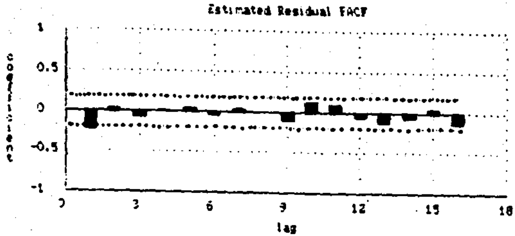


figura 44

PERIODOGRAMA INTEGRAL

El periodograma integral (figura 45) muestra que el comportamiento de los residuales no sale de los límites de Kolmogorov-Smirnov.

Ya no se observa el comportamiento estacional que presentó el periodograma anterior.

Falta ajuste en el modelo para lograr que este comportamiento se asemeje a una recta, es decir que los residuales se ajusten a la media.

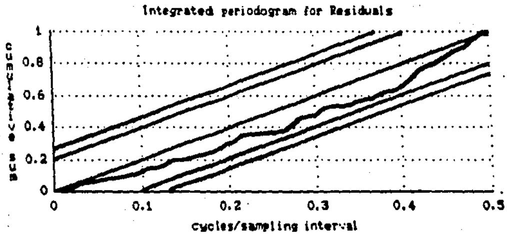


figura 45

C) SMA(1)

FORECAST PLUS

ESTIMACION:

PARAMETRO	VALOR	=t=
Θ_1	0.4527	4.70

SCR = 2654220

S^2 = 21233.76

B-P = 35.4 CON 27 g.l, X_t^2 = 40.1133

DIAGNOSTICO:

---> El parametro sí es significativo en el modelo.

---> La ACF de los residuales (figura 46) presenta un valor que sale de los límite en el período 1. Existen también valores muy cercanos a los límites en 10, 16,18,28 , éstos no tienen explicación en el modelo.

---> $X_t^2 > X_c^2$, las autocorrelaciones no significativas pueden considerarse cero.

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

113 Observations Used For Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = 2654220
Residual Variance = 21233.76

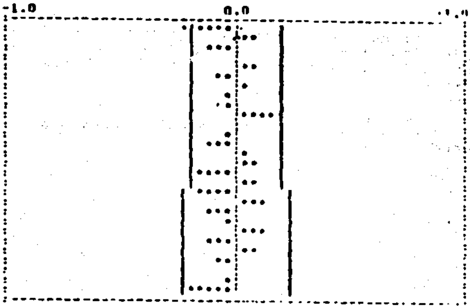
Parameter	Value	T-Statistic
SMA(1)	0.4527	4.70

INGRESOS - BOX-JENKINS ANALYSIS

RESIDUAL AUTOCORRELATION FUNCTION

112 Observations in the Residual Series
Mean of the Residual Series = -.955750
Standard Deviation of the Residual Series = 151.9100

Lag	Value	T-Value
1	-0.25	-2.63
2	0.09	0.89
3	-0.10	-1.03
4	-0.01	-0.06
5	0.06	0.64
6	-0.08	-0.81
7	0.04	0.40
8	-0.04	-0.35
9	-0.09	-0.85
10	0.13	1.36
11	0.01	0.14
12	-0.04	-0.41
13	-0.10	-0.98
14	0.04	0.36
15	0.06	0.57
16	-0.17	-1.73
17	0.06	0.57
18	-0.15	-1.45
19	0.11	0.97
20	-0.13	-1.22
21	-0.06	-0.49
22	0.11	0.95
23	-0.10	-0.88
24	0.07	0.63
25	-0.07	-0.64
26	-0.02	-0.13
27	-0.03	-0.28
28	-0.21	-1.81



[ ] - Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 27 Degrees of Freedom = 35.4
Probability = .129

figura 46

STATGRAPHICS

ESTIMACION:

PARAMETRO	VALOR	=t=
Θ_1	0.4526	4.4286
MEDIA	-80.589	-0.09362
CTE.	-80.589	

SCR = 2.65215E10

S^2 = 2.38932E8 con 111 g.l

X^2 = 21.3679

DIAGNOSTICO:

---> El parámetro Θ_1 sí es significativo en el modelo.

---> La media y la constante no son significativas en el modelo.

---> La ACF (fig.47) muestra un valor grande en 1 (sugiere comportamiento autorregresivo). Algunos valores cercanos a los límites de confianza.

---> La PACF (figura 48) muestra un valor grande en 1 (también sugiere comportamiento autorregresivo).

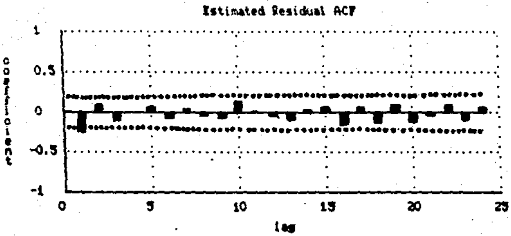


figura 47

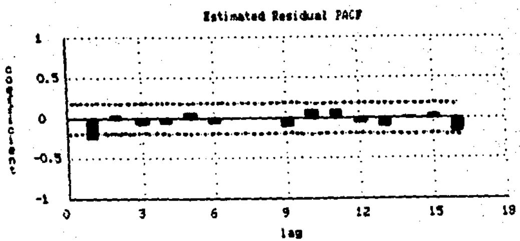


figura 48

PERIODOGRAMA INTEGRAL

La figura 49 muestra que ya no hay comportamiento estacional pero falta ajuste. No sale de los límites pero se aleja un poco de la media.

Lo que podemos concluir es que falta considerar la parte autorregresiva.

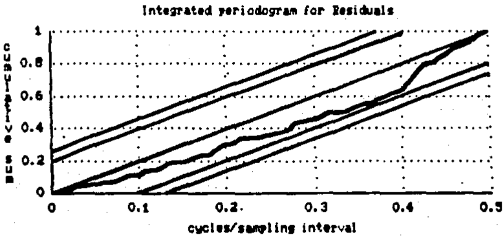


figura 49

D) AR(1) x SAR(1)

FORECAST PLUS

ESTIMACION:

PARAMETROS	VALOR	=t=
ϕ_1	-0.1680	-1.89
Φ_1	-0.4334	-4.65

SCR = 2378140
S^2 = 19186.61
B-P = 23 , X^2_t = 35.1725

DIAGNOSTICO:

---> El parámetro ϕ_1 no es significativo en el modelo ya el valor de /t/ es menor que 2.

---> El parámetro Φ_1 si es significativo para el modelo ya que el valor de /t/ > 2.

---> $X^2_t > X^2_c$, 23 < 35.1725, los residuales pueden considerarse como ruido blanco.

---> La ACF (figura 50) no muestra valores altos.

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

113 Observations Used For Parameter Estimation
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = 2?79140
Residual Variance = 19106.61

Parameter	Value	T-Statistic
AR(1)	-0.1680	-1.89
SAR(1)	-0.4134	-4.65

Correlation Matrix

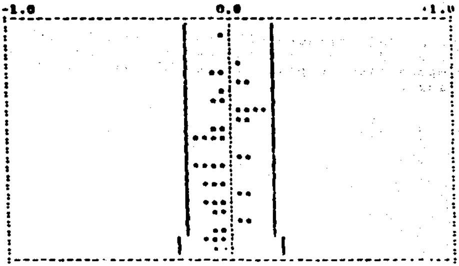
	AR(1)	SAR(1)
AR(1)	1.00	
SAR(1)	0.00	1.00

INGRESOS - BOX-JENKINS ANALYSIS

RESIDUAL AUTOCORRELATION FUNCTION

112 Observations in the Residual Series
Mean of the Residual Series = -4.0?1630
Standard Deviation of the Residual Series = 154.9673

Lag	Value	T-Value
1	-0.00	-0.04
2	-0.01	[illegible]
3	-0.02	[illegible]
4	0.01	[illegible]
5	0.05	[illegible]
6	-0.06	[illegible]
7	0.06	[illegible]
8	-0.05	[illegible]
9	-0.08	[illegible]
10	0.13	[illegible]
11	0.07	[illegible]
12	-0.09	[illegible]
13	-0.17	[illegible]
14	-0.00	[illegible]
15	0.08	[illegible]
16	-0.10	[illegible]
17	-0.01	[illegible]
18	-0.11	[illegible]
19	0.07	[illegible]
20	-0.14	[illegible]
21	-0.09	[illegible]
22	0.09	[illegible]
23	-0.08	[illegible]
24	-0.13	[illegible]
25	-0.08	-0.74



[ ] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 23 Degrees of Freedom = 21
Probability = .162

figura 50

STATGRAPHICS

ESTIMACION:

PARAMETROS	VALOR	=t=
ϕ_1	-0.22562	-2.42537
Θ_1	-.47422	-4.85555
MEDIA	-65.4323	-0.0784
CTE.	-118.2248	

SCR =2.53248E10
S^2 = 2.30225E8 110 g.1
X^2 = 12.4333

DIAGNOSTICO:

---> ϕ_1 sí es significativo para el modelo.

---> Θ_1 sí es significativo para el modelo.

---> La ACF (figura 51), muestra un valor grande en el período 13 (que no explica nada en el modelo), los demás valores son pequeños.

---> LA PACF (figura 52) muestra un valor relativamente grande en el período 14, los demás son pequeños. Hay que hacer notar que este valor tampoco tiene explicación en el modelo y no sale del intervalo de confianza.

---> $X^2_c < X^2_t$.

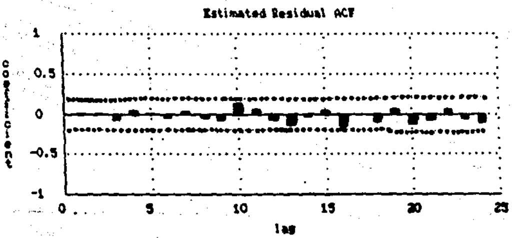


figura 51

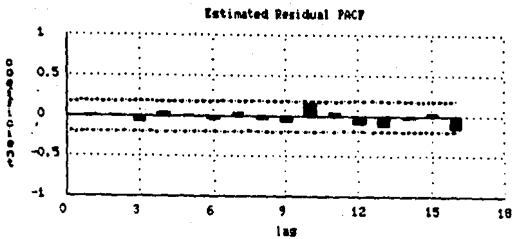


figura 52

PERIODOGRAMA INTEGRAL

En el periodograma integral (figura 53) se observa que los residuales se asemejan màs a una recta en comparaciòn a los programas presentados anteriormente.

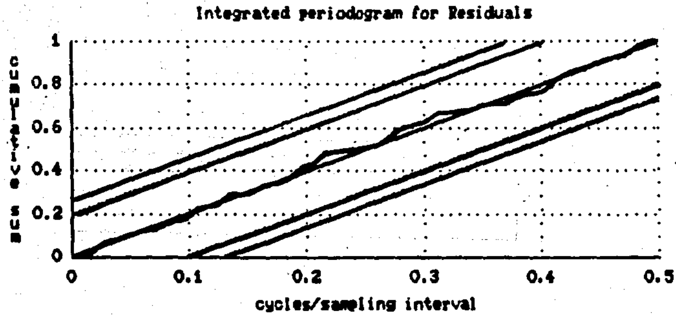

Integrated periodogram for Residuals
cumulative sum
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
0.4
0.5
cycles/sampling interval

figura 53

E) AR(1) x SMA(1)

FORECAST PLUS

ESTIMACION:

PARAMETROS	VALOR	=t=
ϕ_1	-0.2530	-2.91
Θ_1	0.4727	4.87

SCR = 2485194
S^2 = 20041.88
B-P = 24 con 26 g.l. X^2_t =38.8851

DIAGNOSTICO:

---> Los dos parámetros son significativos para el modelo.

---> La función de autocorrelación de los residuales (figura 54) no presenta valores significativamente grandes. El valor de =t= más grande es de -1.90 correspondiente al período 28 que podría considerarse como cero.

---> $X^2_c < X^2_t$, 24 < 38.8851, los residuales pueden considerarse como ruido blanco.

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

112 Observations Used For Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = [illegible]
Residual Variance = 20041.38

Parameter	Value	T-Statistic
AR(1)	-0.2530	-2.31
SMA(1)	0.3727	4.07

Correlation Matrix

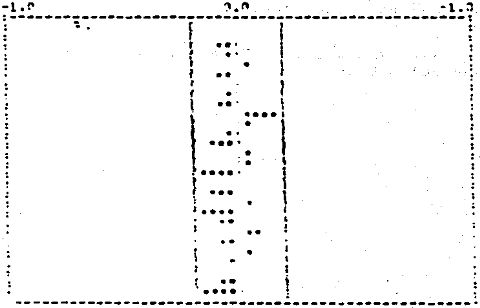
	AR(1)	SMA(1)
AR(1) :	1.00	
SMA(1) :	0.00	1.00

INGRESOS - BOX-JENKINS ANALYSIS

RESIDUAL AUTOCORRELATION FUNCTION

112 Observations in the Residual Series
Mean of the Residual Series = [illegible]
Standard Deviation of the Residual Series = 149.6194

Lag	Value	T-Value
[illegible]	0.01	0.39
[illegible]	0.01	0.10
[illegible]	[illegible]	[illegible]



[illegible] Estimated Two-Standard Error Limits

[illegible] Chi-Square Statistic with 26 Degrees of Freedom [illegible]
Probability [illegible]

figura 54

STATGRAPHICS

ESTIMACION:

PARAMETROS	VALOR	"t"
ϕ_1	-0.25094	-2.71305
Θ_1	0.47022	4.59116
MEDIA	-79.07205	4.59116
CTE.	-98.87855	

$SCR = 2.48524E10$
$S^2_e = 2.25931E8$ con 110 g.l
$X^2 = 11.5848$

DIAGNOSTICO:

---> Los dos parámetros son significativos en el modelo.

---> La ACF y la PACF (figuras 55 y 56) muestran un comportamiento de ruido blanco.

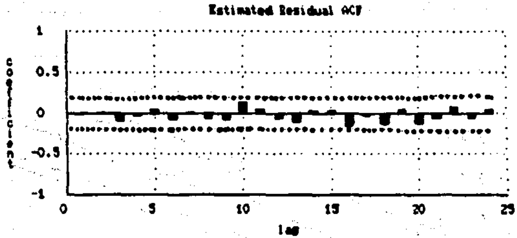


figura 55

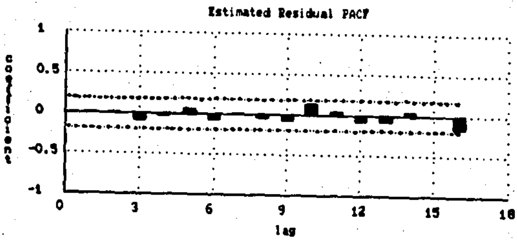


figura 56

PERIODOGRAMA INTEGRAL

El periodograma muestra que los residuales se mantienen dentro de los límites y que el comportamiento es el de una recta con ligeras variaciones aleatorias (figura 57). Comparando con los periodogramas anteriores, éste muestra un muy buen ajuste, junto con el del modelo AR x SAR.

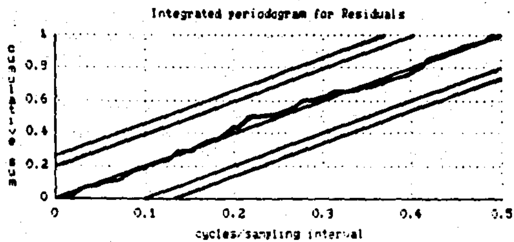


figura 57

Obsèrvense ahora los siguientes cuadros comparativos. En ellos se encuentran resumidos todos los valores estimados en esta secciòn y a partir de èstos se obtiene el resultado de cual es el mejor modelo.

CUADROS COMPARATIVOS

El modelo que tiene SCR, estadístico de BOX PIERCE y S más pequeñas es el D, AR x SAR. Esto indicaría que éste es el modelo apropiado, pero hay que observar que el parámetro autorregresivo no contiene información en el modelo. Por lo tanto al descartar este modelo se concluye que el modelo que mejor representa al fenómeno es el E, AR x SMA, ya que una vez eliminado el D, es el modelo que presenta SCR,S y BOX-PIERCE mínimos.

Los ingresos por turismo en México se comportan como:

$$Z_t = -0.2530\ Z_{t-1} + e_t - 0.4727\ e_{t-q}$$

(resultado de utilizar el paquete Forecast Plus).

C U A D R O   C O M P A R A T I V O

(FORECAST PLUS)

MOD / PAR	A	B	C	D	E
$\varnothing_1$	-0.22230			-0.1000 *	-0.2530
Φ_1		-0.4283		-0.4334	
Θ_1			0.4527		0.4727
SCR	3000303	2440584	2854220	237140	2405184
χ^2	33.4	27.3	35.4	23	24
s^2	24006.42	18500.5	21233.70	18200.01	20841.00

* parametro no significativo

C U A D R O   C O M P A R A T I V O

(STATGRAPHICS)

MOD / PAR	A	B	C	D	E
$\varnothing_1$	-0.22247			-0.22582	-0.25040
Φ_1		-0.43064		-0.47422	
Θ_1			0.45200		0.47022
SCR	3.00016E10	2.66006E10	2.05215E10	2.53540E10	2.40524E10
χ^2	21.0281	20.000	21.3679	12.4939	11.5040
s^2	2.71006E8	2.40530E8	2.30032E8	2.30225E8	2.25031E8

El modelo que presenta la SCR, S^2, X^2 mínimas es el E ; por lo que se concluye que el modelo que mejor representa el comportamiento de Ingresos por Turismo en México es :

AR(1) x SMA(1)

$$Z_t = -0.25049\, Z_{t-1} + e_t - 0.47022\, e_{t-12}$$

(Resultado de utilizar el paquete Statgraphics).

4.3 PRONOSTICO

Una vez que se ha encontrado el modelo que refleja el comportamiento de Ingresos por Turismo se procede a realizar el pronóstico de valores futuros.

<u>FORECAST PLUS:</u>

El pronóstico generado por este paquete lo muestra la figura 58. Se pronosticaron 12 valores y son:

1988

JULIO	AGOSTO	SEPT.	OCT.	NOV
153270.25	177975.82	129411.65	156749.24	184965.21
DIC				
228766.38				

1989

ENERO	FEBRERO	MARZO	ABRIL	MAYO
257653.30	284471.22	305897.24	226502.59	185334.56
JUNIO				
186413.77				

(Todos los datos son en miles de dólares)

Recuérdese que los valores utilizados en el FORECAST PLUS fueron divididos por 100, los valores presentados aquí ya fueron multiplicados por 100.

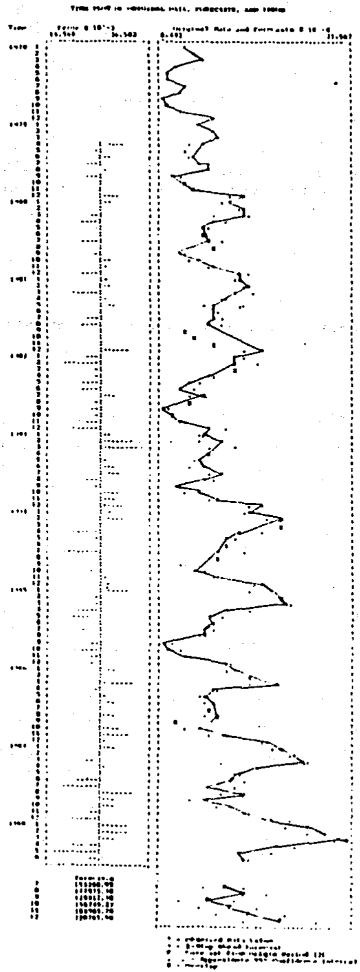


FIGURA 08

STATGRAPHICS:

El pronóstico generado por este paquete lo muestra la figura 59. Se pronosticaron 12 valores y son:

1988

JULIO	AGOSTO	SEPT.	OCT.	NOV.
156016	181956	127474	158310	182550

DIC.
229938

1989

ENERO	FEBRERO	MARZO	ABRIL	MAYO
258642	281496	303564	222358	185263

JUNIO
187275

Claramente se puede notar que la diferencia entre paquetes no es muy grande.

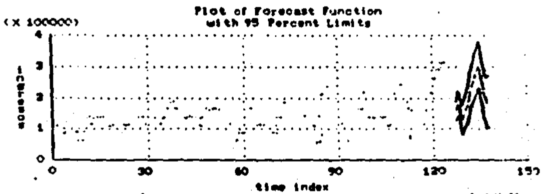


FIGURA 59

CONCLUSIONES

Resumiendo, los resultados obtenidos son:

---> El fenómeno de Ingresos por Turismo en México se comporta como un modelo:

$$AR(1) \quad x \quad SMA(1)$$

---> El modelo estimado por FORECAST PLUS es:

$$Z_t = -0.2530\, Z_{t-1} + e_t - 0.4727\, e_{t-12} \quad \text{(Modelo 1)}$$

---> El modelo estimado por STATGRAPHICS es:

$$Z = -0.25049\, Z_{t-1} + e - 0.47022\, e_{t-12} \quad \text{(Modelo 2)}$$

Ahora se puede decir que el valor t+1 de Ingresos por Turismo en México depende del valor anterior (t) y del valor del error aleatorio de 12 periodos atrás.

Los parámetros estimados por los dos paquetes no son exactamente iguales, ya que son estimaciones, pero la diferencia debe ser pequeña:

	0	
FORECAST PLUS	-0.2530	-0.4727
STATGRAPHICS	-0.25049	-0.47022
DIFERENCIA	0.00251	0.00248

Como se puede observar la diferencia es muy pequeña y se puede considerar que los paquetes dan como resultado el mismo modelo.

Las tablas (a) y (b) , que se encuentran al final de esta sección, muestran los valores pronosticados por FORECAST PLUS y STATGRAPHICS respectivamente , los valores reales(1) así como el error existente entre éstos. Obsérvese que el error es pequeño si se compara con la magnitud de los datos originales.

El anàlisis realizado en el desarrollo de esta tesis lleva a conlcluir que:

a. El fenômeno de ingresos por Turismo SI puede ser analizado adecuadamente por este mètodo.

b. Se logrô predecir 12 meses con poco error, sin que los valores tendieran a la media.

c. La diferencia que existe entre las predicciones y los valores reales se debe, en parte, a los fenômenos naturales (temblores, maremotos, etc) o situaciones que se fueron presentando y que no se pueden contemplar en las estimaciones ya que son situaciones desconocidas y que estàn fuera del alcance del ser humano.

Las ventajas y desventajas que se concluyen del uso de los dos paquetes son las siguientes:

FORECAST PLUS

VENTAJAS

a. Las gráficas que despliega en cuanto a tamaño son recomendables para un uso didàctico, ya que proporciona mucha informaciôn y son muy claras.

b. La utilizaciôn de menús facilita el uso del paquete.

c. No es necesario definir el ambiente de trabajo.

d. Contiene otros mêtodos de prediccîôn (aunque esto no es de importancia para esta tesis).

(1) Datos proporcionados por la hemeroteca del Banco de México

g. Todos los valores estimados, autocorrelaciones, desviación estândar, etc. pueden ser guardadas en disco.

<u>DESVENTAJAS:</u>

a. Es necesario definir el ambiente de trabajo. Es decir, es necesario que el usuario conozca el Hardware que está utilizando.

b. En las grâficas no aparecen los valores de las autocorrelaciones ni el estadístico =t= de cada una de ellas (aunque aparecen en tablas aparte.

c. En un mismo tamaño de grâfica la escala puede variar dependiendo de los datos, es decir que la escala de los datos originales y la de logaritmo de los datos puede ser distinta.

d. La versiôn de STRATGRAPHICS utilizada en un principio, realizaba menos iteraciones en la estimaciôn de parâmetros; pero como se pudo observar posteriormente con un mayor nômero de iteraciones (realizadas por la nueva versiôn) se logrô una mejor estimación de los parâmetros.

e. El paquete redondea los datos. Nôtese que los datos pronosticados por el paquete son enteros. Sôlo en ciertas ocasiones no redondea.

A travês del desarrollo de esta tesis y del anâlisis realizado, se puede concluir que es indispensable realizar este tipo de anâlisis con fenômenos en los que se encuentre necesario obtener valores futuros para la toma de desiciones y que estos valores no dependan de decisones polîticas, ô de fenômenos naturales, que uno, como analista no puede contemplar en el anâlisis ni en las estimaciones.

Los Ingresos por Turismo en Mêxico presentan un comportamiento relativamente sencillo en cuanto a la aplicaciôn del Mêtodo de Box y Jenkins. Ya que como se observô presentan ACF y PACF en las que se pudo ejemplificar la teorîa explicada en el Capîtulo 1.

También se puede decir que durante el anâlisis (Capîtulo 4) se hizo notar que es necesario, que SIEMPRE se utilice el juicio personal, ya que como se explicô anteriormente, existen factores que no serân contemplados por los paquetes y los resultados no pueden ser exactos. Otra justificaciôn a utilizar el razonamiento y conocimiento del fenômeno en particular es el que se está trabajando con muestras, y en este caso, es una muestra de poblaciones humanas que depende de encuestas y estimaciones explicadas en el Capîtulo 2.

e. Existe la opción de efectuar las transformaciones que se requieran en el momento que se desee.

f. La gráfica del pronóstico es mucho muy clara ya que presenta todos los datos más el pronóstico, y no únicamente el comportamiento de los datos pronosticados.

g. Dentro de esta misma gráfica presenta los valores futuros que se requieren.

h. Aunque se tuvo que dividir los datos entre 100, las estimaciones de los parámetros fueron bastante acertadas. Esto se logró corroborar gracias a la nueva versión del STATGRAPHICS.

DESVENTAJAS:

a. En el TIME PLOT de los datos, éstos no aparecen unidos por líneas, sino que el usuario tiene que hacerlo manualmente.

b. No existe facilidad de cambiar tamaño de gráficas a conveniencia del usuario.

c. No proporciona el periodograma integral.

d. No soporta datos muy grandes en el momento de realizar ciertas estimaciones.

STATGRAPHICS

VENTAJAS:

a. El tamaño de las gráficas que se despliega es muy útil para distintos objetivos, tanto para uso didáctico como para impresiones de artículos,tesis, etc.

b. En todas las gráficas los puntos aparecen unidos por líneas. Cabe mencionar que en la nueva versión, utilizada al final de esta tesis, existe la opción de obtener gráficas sólo con puntos, sólo con líneas ó con líneas y puntos.

c. El uso de los menús es muy sencillo.

d. Proporciona periodograma integral acumulativo.

e. Proporciona ACF y PACF de los residuales.

f. La versión 2.1, realmente mejoró en comparación con la anterior, sobre todo para un mejor entendimiento entre el usuario y el paquete.

g. Todos los valores estimados, autocorrelaciones, desviación estándar, etc. pueden ser guardadas en disco.

DESVENTAJAS:

a. Es necesario definir el ambiente de trabajo. Es decir, es necesario que el usuario conozca el Hardware que está utilizando.

b. En las gráficas no aparecen los valores de las autocorrelaciones ni el estadístico =t= de cada una de ellas (aunque aparecen en tablas aparte.

c. En un mismo tamaño de gráfica la escala puede variar dependiendo de los datos, es decir que la escala de los datos originales y la de logaritmo de los datos puede ser distinta.

d. La versión de STRATGRAPHICS utilizada en un principio, realizaba menos iteraciones en la estimación de parámetros; pero como se pudo observar posteriormente con un mayor número de iteraciones (realizadas por la nueva versión) se logró una mejor estimación de los parámetros.

e. El paquete redondea los datos. Nótese que los datos pronosticados por el paquete son enteros. Sólo en ciertas ocasiones no redondea.

A través del desarrollo de esta tesis y del análisis realizado, se puede concluir que es indispensable realizar este tipo de análisis con fenómenos en los que se encuentre necesario obtener valores futuros para la toma de desiciones y que estos valores no dependan de decisones políticas, ó de fenómenos naturales, que uno, como analista no puede contemplar en el análisis ni en las estimaciones.

Los ingresos por Turismo en México presentan un comportamiento relativamente sencillo en cuanto a la aplicación del Método de Box y Jenkins. Ya que como se observó presentan ACF y PACF en las que se pudo ejemplificar la teoría explicada en el Capítulo 1.

También se puede decir que durante el análisis (Capítulo 4) se hizo notar que es necesario, que SIEMPRE se utilice el juicio personal, ya que como se explicó anteriormente, existen factores que no serán contemplados por los paquetes y los resultados no pueden ser exactos. Otra justificación a utilizar el razonamiento y conocimiento del fenómeno en particular es el que se está trabajando con muestras, y en este caso, es una muestra de poblaciones humanas que depende de encuestas y estimaciones explicadas en el Capítulo 2.

TABLAS COMPARATIVAS DE VALORES

MES	FORECAST PLUS (+)	VALORES REALES (+)	ERROR
JULIO	153270.25	200163.1	46892.85 POR ABAJO DEL VALOR
AGOSTO	177975.82	192633.2	14657.38 POR ABAJO DEL VALOR
SEPT.	129411.65	128269.0	1142.65 POR ARRIBA DEL VALOR
OCT.	156749.24	143243.7	13505.54 POR ARRIBA DEL VALOR
NOV.	184965.21	180952.0	4013.21 POR ARRIBA DEL VALOR
DIC.	228766.38	264703.3	35936.92 POR ABAJO DEL VALOR
ENERO	257653.30	274502.5	16849.20 POR ABAJO DEL VALOR
FEBRERO	284471.22	263068.5	21403.72 POR ARRIBA DEL VALOR
MARZO	305897.24	333788.0	27890.76 POR ABAJO DEL VALOR
ABRIL	226502.59		
MAYO	185334.56		
JUNIO	188413.77		

TABLA (a)

(+) miles de dolares

Por último según los resultados obtenidos por los paquetes, la diferencia de pronósticos es muy pequeña por lo que los dos paquetes proporcionan resultados similares, como se muestra en las tablas (a) y (b), y sobre todo un muy buen resultado.

MES	STATGRAPHICS (+)	VALORES REALES (+)	ERROR
JULIO	156016	200163.1	44147.1 POR ABAJO DEL VALOR
AGOSTO	181956	192633.2	10677.2 POR ABAJO DEL VALOR
SEPT.	127474	128269.0	795 POR ABAJO DEL VALOR
OCT.	158310	143243.7	15062.3 POR ARRIBA DEL VALOR
NOV.	182550	180952.0	1598.0 POR ARRIBA DEL VALOR
DIC.	229938	264703.3	34815.0 POR ABAJO DEL VALOR
ENERO	258842	274502.5	15660 POR ABAJO DEL VALOR
FEBRERO	281496	263068.5	18427.0 POR ARRIBA DEL VALOR
MARZO	303564	333788.0	30224.0 POR ABAJO DEL VALOR
ABRIL	222350		
MAYO	185263		
JUNIO	187275		

TABLA (b)

(+) miles de dólares

A P E N D I C E

DATOS EJEMPLOS

Contents of file A:EJEMPLOS

Variable Name	Type	Rank	Length	Date	Time	Comment
FDIRECTORY				1/ 1/80	01:34	TESIS EJEMPLOS
VAR1	N	1	100	7/ 6/89	01:42	MA(1)
VAR2	N	1	100	7/ 6/89	01:42	ARMA(1,1)
VAR3	N	1	100	7/ 6/89	01:42	MA(2)
VAR4	N	1	100	7/ 6/89	01:42	ARMA(1,1)
VAR5	N	1	100	7/ 6/89	01:43	MA(1)
VAR6	N	1	100	7/ 6/89	01:43	AR(2)
VAR7	N	1	100	7/ 6/89	01:43	AR(2)
VAR8	N	1	100	7/ 6/89	01:43	AR(1)
VAR9	N	1	100	7/ 6/89	01:43	SAR(1)
VAR10	N	1	100	7/ 6/89	01:43	SARMA(2,1)
VAR11	N	1	100	7/ 6/89	01:43	MA(1) X SMA(1)
VAR12	N	1	50	7/ 6/89	01:43	AR(1) X SAR(1)

Thr Jul 6 1989 02:11:36 AM Page 1

Variable: A:EJEMPLOS.VAR2 (length = 100)

(1)	-0.3145	(19)	1.6142	(37)	2.6094	(55)	0.9905	(73)	1.0568
(2)	-1.3094	(20)	1.7508	(38)	1.4675	(56)	0.848	(74)	1.4022
(3)	-2.18	(21)	2.777	(39)	1.3391	(57)	1.7337	(75)	1.977
(4)	-2.1	(22)	1.6551	(40)	1.0181	(58)	1.7627	(76)	1.4165
(5)	-1.1996	(23)	2.4655	(41)	1.2062	(59)	3.1468	(77)	0.1493
(6)	-0.1027	(24)	2.1826	(42)	0.848	(60)	3.1797	(78)	-0.8581
(7)	-0.674	(25)	0.7016	(43)	0.4304	(61)	2.8976	(79)	-0.7793
(8)	0.3776	(26)	-2.2036	(44)	-2.1267	(62)	1.9907	(80)	1.3402
(9)	-0.7316	(27)	-0.0399	(45)	-1.8704	(63)	0.148	(81)	-1.0668
(10)	-2.179	(28)	0.9817	(46)	-2.3394	(64)	0.247	(82)	-1.2793
(11)	-0.7793	(29)	0.2323	(47)	-1.118	(65)	-1.1444	(83)	0.73
(12)	0.5637	(30)	2.1265	(48)	-0.1693	(66)	-0.0449	(84)	1.722
(13)	-0.794	(31)	2.7656	(49)	-1.3935	(67)	0.903	(85)	2.551
(14)	-0.4069	(32)	2.8481	(50)	-0.1398	(68)	0.0277	(86)	0.5063
(15)	-0.5974	(33)	1.9889	(51)	-3.3E-3	(69)	0.112	(87)	-0.45
(16)	-0.6611	(34)	1.6272	(52)	-0.7548	(70)	2.6283	(88)	-2.0298
(17)	-1.3964	(35)	0.9302	(53)	0.2822	(71)	0.5065	(89)	-2.2993
(18)	1.006	(36)	3.1598	(54)	1.2695	(72)	-0.3336	(90)	-2.1452

(91)	-1.7529
(92)	-0.6796
(93)	-0.7475
(94)	0.9668
(95)	1.2362
(96)	1.089
(97)	2.0514
(98)	1.2759
(99)	0.6754
(100)	-2.1749

Variable: A:EJEMPLOS.VAR1   (length = 100)

(1)	-0.3145	(19)	-1.5673	(37)	-2.4607	(55)	-0.8425	(73)	1.3647
(2)	-0.7433	(20)	0.363	(38)	0.061	(56)	0.3258	(74)	-0.7792
(3)	-0.1689	(21)	0.9458	(39)	0.5808	(57)	0.9186	(75)	0.6992
(4)	0.4873	(22)	-1.8137	(40)	-0.343	(58)	-0.6384	(76)	-0.9734
(5)	0.5833	(23)	2.0821	(41)	0.5318	(59)	1.6651	(77)	-0.5762
(6)	0.3456	(24)	-1.2135	(42)	-0.5507	(60)	-1.0529	(78)	-0.116
(7)	-1.2954	(25)	-0.8274	(43)	-8E-4	(61)	0.2062	(79)	0.6263
(8)	1.7198	(26)	-1.7856	(44)	-2.2972	(62)	-0.6369	(80)	1.8408
(9)	-2.1915	(27)	4.1941	(45)	2.2498	(63)	-1.0242	(81)	-4.0659
(10)	-0.2126	(28)	-1.4506	(46)	-1.3997	(64)	1.447	(82)	2.2911
(11)	2.1431	(29)	-0.8274	(47)	1.7636	(65)	-1.8662	(83)	1.3643
(12)	-1.2578	(30)	2.5759	(48)	-0.3314	(66)	2.3699	(84)	-0.4987
(13)	-7.5E-3	(31)	-1.0668	(49)	-1.6661	(67)	-0.3814	(85)	0.5263
(14)	0.4707	(32)	0.2106	(50)	2.3554	(68)	-1.2376	(86)	-2.5137
(15)	-0.5728	(33)	-0.7976	(51)	-2.2806	(69)	0.9475	(87)	0.973
(16)	0.1788	(34)	0.478	(52)	1.356	(70)	2.1915	(88)	-1.3629
(17)	-0.7928	(35)	-0.5402	(53)	0.4458	(71)	-4.0307	(89)	0.9022
(18)	2.9683	(36)	2.8378	(54)	0.2896	(72)	1.5877	(90)	-0.1036

(91)	0.2084
(92)	0.6858
(93)	-0.9155
(94)	1.9823
(95)	-1.2273
(96)	0.1443
(97)	1.0668
(98)	-1.5113
(99)	0.3637
(100)	-2.5827

Thr Jul 6 1989 02:13:01 AM Page 1

Variable: A:EJEMPLOS.VAR3 (length = 100)

(1)	-0.3145	(19)	-3.4557	(37)	-4.1993	(55)	-0.9309	(73)	0.1637
(2)	-0.9949	(20)	1.5737	(38)	1.8493	(56)	0.9344	(74)	-1.6062
(3)	0.2393	(21)	0.756	(39)	0.4983	(57)	0.7043	(75)	1.324
(4)	0.4692	(22)	-2.3151	(40)	-0.7208	(58)	-1.1654	(76)	-1.3607
(5)	0.1752	(23)	3.3367	(41)	0.7839	(59)	2.1778	(77)	0.1175
(6)	-0.0385	(24)	-2.5474	(42)	-0.874	(60)	-2.025	(78)	0.2006
(7)	-1.4601	(25)	0.0828	(43)	0.39	(61)	1.0594	(79)	0.6036
(8)	2.5727	(26)	-1.2889	(44)	-2.3143	(62)	-0.7098	(80)	1.3772
(9)	-3.3183	(27)	5.1663	(45)	3.6139	(63)	-0.5767	(81)	-5.1655
(10)	1.2327	(28)	-4.2607	(46)	-3.0367	(64)	2.0375	(82)	4.9555
(11)	2.1121	(29)	0.3113	(47)	2.6438	(65)	-2.8766	(83)	-0.3184
(12)	-2.6961	(30)	3.113	(48)	-1.5253	(66)	3.5495	(84)	-1.2888
(13)	0.8823	(31)	-2.6792	(49)	-1.3641	(67)	-1.9794	(85)	1.0121
(14)	0.4325	(32)	1.1064	(50)	3.3977	(68)	-0.8651	(86)	-2.7401
(15)	-0.8904	(33)	-0.8475	(51)	-3.8596	(69)	1.7591	(87)	2.6159
(16)	0.551	(34)	1.0431	(52)	2.8744	(70)	1.5297	(88)	-2.1407
(17)	-0.9462	(35)	-0.8535	(53)	-0.5214	(71)	-5.3066	(89)	1.6816
(18)	3.4286	(36)	3.1981	(54)	0.0619	(72)	4.3004	(90)	-0.8392

(91)	0.2234
(92)	0.5106
(93)	-1.3422
(94)	2.6019
(95)	-2.4704
(96)	1.0756
(97)	0.9888
(98)	-2.1271
(99)	1.4179
(100)	-2.8644

Variable: A:EJEMPLOS.VAR4 (length = 100)

(1)	-0.3145	(19)	-2.49	(37)	-3.2904	(55)	-1.0143	(73)	-0.1959
(2)	-0.6804	(20)	1.6784	(38)	1.5917	(56)	0.7863	(74)	-0.2881
(3)	-0.1271	(21)	0.2717	(39)	-0.2883	(57)	0.5858	(75)	0.924
(4)	0.2614	(22)	-1.5337	(40)	-0.0832	(58)	-0.6074	(76)	-1.1611
(5)	0.3672	(23)	2.6376	(41)	0.563	(59)	2.0364	(77)	-0.0682
(6)	0.2687	(24)	-2.0768	(42)	-0.681	(60)	-1.5176	(78)	-0.3128
(7)	-1.2407	(25)	0.2112	(43)	0.2954	(61)	1.092	(79)	0.5632
(8)	2.1028	(26)	-2.1392	(44)	-2.4805	(62)	-1.0195	(80)	1.5714
(9)	-2.9268	(27)	4.5297	(45)	2.7724	(63)	-0.5748	(81)	-4.2895
(10)	0.8492	(28)	-3.0444	(46)	-2.6851	(64)	1.3789	(82)	3.6658
(11)	1.3373	(29)	0.7964	(47)	2.7669	(65)	-2.406	(83)	-0.3973
(12)	-1.5884	(30)	2.0108	(48)	-1.4572	(66)	3.1328	(84)	0.1662
(13)	0.6797	(31)	-1.433	(49)	-0.8307	(67)	-1.5889	(85)	0.6666
(14)	0.043	(32)	1.1185	(50)	2.3563	(68)	-0.2704	(86)	-2.5104
(15)	-0.5234	(33)	-1.1406	(51)	-3.0837	(69)	0.8496	(87)	1.7434
(16)	0.3255	(34)	0.982	(52)	2.5137	(70)	1.8644	(88)	-2.3306
(17)	-0.9939	(35)	-0.9408	(53)	-0.7115	(71)	-4.2273	(89)	1.5818
(18)	3.1967	(36)	3.2185	(54)	0.8588	(72)	3.0806	(90)	-1.0124

(91)	0.5892
(92)	0.3534
(93)	-0.9167
(94)	2.3065
(95)	-1.8933
(96)	1.1126
(97)	0.5711
(98)	-1.4283
(99)	0.9193
(100)	-3.0601

Variable: A:EJEMPLOS.VAR5 (length = 100)

(1)	-0.3145	(19)	2.036	(37)	1.6967	(55)	0.5312	(73)	1.2566
(2)	-1.2465	(20)	0.738	(38)	-0.55	(56)	0.0768	(74)	1.3178
(3)	-1.7608	(21)	1.8267	(39)	0.1896	(57)	1.2407	(75)	1.13
(4)	-1.0564	(22)	0.4042	(40)	0.2733	(58)	1.089	(76)	0.4899
(5)	0.128	(23)	0.9546	(41)	0.4761	(59)	2.0255	(77)	-0.963
(6)	0.9148	(24)	1.2159	(42)	0.2556	(60)	1.8995	(78)	-1.3475
(7)	-0.2869	(25)	-0.8255	(43)	-0.2369	(61)	0.8834	(79)	-0.5445
(8)	0.4538	(26)	-3.108	(44)	-2.4874	(62)	0.2347	(80)	1.9062
(9)	-0.4525	(27)	0.2792	(45)	-1.5779	(63)	-1.346	(81)	-1.0682
(10)	-2.3278	(28)	2.128	(46)	-0.8622	(64)	-0.4491	(82)	-1.8161
(11)	0.1107	(29)	-0.2854	(47)	-0.0459	(65)	-1.0678	(83)	1.7443
(12)	0.5453	(30)	1.6856	(48)	1.0426	(66)	0.0226	(84)	1.9882
(13)	-0.5775	(31)	2.3426	(49)	-1.0971	(67)	1.5326	(85)	1.7179
(14)	2.6E-3	(32)	1.231	(50)	0.1448	(68)	-0.3166	(86)	-0.7182
(15)	-0.1942	(33)	0.3557	(51)	-0.2804	(69)	-0.2958	(87)	-1.6125
(16)	-0.4347	(34)	0.1245	(52)	-0.6927	(70)	2.7128	(88)	-1.8746
(17)	-0.9975	(35)	-0.0581	(53)	0.9764	(71)	-0.1072	(89)	-1.6878
(18)	1.5359	(36)	2.359	(54)	1.4275	(72)	-1.7227	(90)	-0.7321

(91)	-0.4601
(92)	0.4845
(93)	0.0207
(94)	1.2664
(95)	1.3716
(96)	0.2597
(97)	1.39
(98)	0.4541
(99)	-0.4819
(100)	-2.6773

Variable: A:EJEMPLOS.VAR6 (length = 100)

(1)	-0.3145	(19)	-1.7579	(37)	-2.4366	(55)	-1.1595
(2)	-0.9949	(20)	2.1958	(38)	1.8523	(56)	1.1952
(3)	-0.4308	(21)	-0.2828	(39)	-1.2135	(57)	0.1306
(4)	-0.225	(22)	-0.0958	(40)	1.0637	(58)	0.3859
(5)	-0.4046	(23)	1.5193	(41)	-0.3769	(59)	1.6399
(6)	0.3424	(24)	-0.9295	(42)	0.2913	(60)	-0.4843
(7)	-0.91157	(25)	0.035	(43)	-0.3691	(61)	1.1629
(8)	1.7048	(26)	-2.6538	(44)	-2.1126	(62)	-0.9955
(9)	-2.528	(27)	3.8359	(45)	1.5297	(63)	-0.3552
(10)	0.5875	(28)	-2.4936	(46)	-2.4713	(64)	0.513
(11)	0.2688	(29)	1.7069	(47)	2.6475	(65)	-1.8459
(12)	-0.4	(30)	0.6079	(48)	-1.7272	(66)	2.4064
(13)	1.1E-3	(31)	0.6144	(49)	0.1842	(67)	-1.2374
(14)	0.1559	(32)	0.4738	(50)	0.7841	(68)	0.4466
(15)	-0.4768	(33)	-0.3823	(51)	-1.7201	(69)	-0.1896
(16)	0.1893	(34)	0.6254	(52)	1.5225	(70)	2.6553
(17)	-1.1041	(35)	-0.7509	(53)	-0.5464	(71)	-3.7001
(18)	2.9525	(36)	3.1741	(54)	1.4909	(72)	2.6836

(73)	-1.0395	(91)	0.8811
(74)	1.4297	(92)	-0.2185
(75)	-0.151	(93)	-0.14
(76)	0.1348	(94)	1.6647
(77)	-0.8807	(95)	-0.9547
(78)	-0.1763	(96)	1.1078
(79)	-0.0294	(97)	0.3728
(80)	1.8559	(98)	-0.5306
(81)	-3.6865	(99)	0.3338
(82)	2.8206	(100)	-2.9365
(83)	-0.8753		
(84)	1.834		
(85)	-0.1533		
(86)	-1.1571		
(87)	0.3438		
(88)	-2.0565		
(89)	0.9098		
(90)	-1.3751		

Variable: A:EJEMPLOS.VAR7 (length = 100)

(1)	-0.3145	(19)	-1.4989	(37)	-2.7756	(55)	-0.6496	(73)	2.232
(2)	-0.9949	(20)	0.4473	(38)	-0.0598	(56)	1.1233	(74)	-3.0387
(3)	0.0197	(21)	1.9277	(39)	2.0984	(57)	0.5707	(75)	2.0063
(4)	0.2965	(22)	-2.5153	(40)	-1.6777	(58)	-0.9053	(76)	-0.0236
(5)	0.1066	(23)	2.374	(41)	0.5871	(59)	2.2271	(77)	-1.9545
(6)	0.367	(24)	-0.3888	(42)	0.3893	(60)	-0.8152	(78)	0.846
(7)	-1.1488	(25)	-1.9998	(43)	-0.7826	(61)	-0.1392	(79)	0.5368
(8)	1.7856	(26)	-0.6616	(44)	-1.9998	(62)	0.3994	(80)	0.9364
(9)	-2.0612	(27)	3.9299	(45)	2.4054	(63)	-1.4211	(81)	-3.6383
(10)	-0.6926	(28)	-2.4082	(46)	-1.8554	(64)	1.3962	(82)	2.5863
(11)	2.9178	(29)	-0.9878	(47)	0.8999	(65)	-1.7313	(83)	1.6682
(12)	-2.2749	(30)	4.366	(48)	0.7489	(66)	1.7436	(84)	-2.1416
(13)	-0.2233	(31)	-2.2623	(49)	-2.5207	(67)	0.2194	(85)	1.8345
(14)	1.7803	(32)	-0.0888	(50)	2.8173	(68)	-1.9989	(86)	1.7986
(15)	-1.6737	(33)	1.2075	(51)	-2.0219	(69)	1.7932	(87)	0.0183
(16)	0.1428	(34)	-0.6114	(52)	0.2588	(70)	2.2169	(88)	-0.5543
(17)	-5.2E-3	(35)	-0.5345	(53)	1.7172	(71)	-4.9185	(89)	0.0396
(18)	2.1706	(36)	3.3929	(54)	-0.6705	(72)	2.5371	(90)	-0.1169

(91)	-0.056
(92)	0.7002
(93)	-0.9739
(94)	1.9834
(95)	-0.9302
(96)	-0.2438
(97)	1.9816
(98)	-1.9675
(99)	0.3259
(100)	-1.7102

==============

Variable: A:EJEMPLOS.VAR8 (length = 100)

(1)	-0.31452	(19)	1.0673	(37)	2.4572	(55)	0.6308
(2)	-1.24657	(20)	1.4044	(38)	1.7213	(56)	0.706
(3)	-1.9621	(21)	2.5098	(39)	1.7622	(57)	1.6445
(4)	-1.8543	(22)	1.3031	(40)	1.3749	(58)	1.5408
(5)	-1.1277	(23)	2.5609	(41)	1.6039	(59)	3.078
(6)	-0.2718	(24)	2.0499	(42)	1.1356	(60)	2.8857
(7)	-1.0087	(25)	0.8134	(43)	0.7895	(61)	2.8534
(8)	0.2798	(26)	-1.796	(44)	-1.7606	(62)	2.0816
(9)	-1.09814	(27)	0.7998	(45)	-1.0725	(63)	0.4801
(10)	-2.1487	(28)	0.9785	(46)	-1.989	(64)	0.883
(11)	-0.592	(29)	0.2263	(47)	-0.7323	(65)	-0.7605
(12)	-0.8298	(30)	2.3119	(48)	-0.2302	(66)	0.5878
(13)	-0.9564	(31)	2.4873	(49)	-1.5658	(67)	1.0459
(14)	-0.5285	(32)	2.7107	(50)	-2.5E-3	(68)	0.0595
(15)	-0.8063	(33)	1.9476	(51)	-1.2825	(69)	0.3734
(16)	-0.773	(34)	1.8593	(52)	-0.6944	(70)	2.75099
(17)	-1.5236	(35)	1.1882	(53)	0.1556	(71)	0.1317
(18)	1.0412	(36)	3.549	(54)	0.983	(72)	0.0379

(73)	1.341	(91)	-1.5964
(74)	1.342	(92)	-0.6919
(75)	1.98831	(93)	-1.00097
(76)	1.3488	(94)	0.8236
(77)	0.3094	(95)	0.731
(78)	-0.4842	(96)	0.7868
(79)	-0.3465	(97)	1.8579
(80)	1.5963	(98)	0.9578
(81)	-1.29	(99)	0.7071
(82)	-0.7944	(100)	-2.0643
(83)	0.9187		
(84)	1.4797		
(85)	2.3059		
(86)	0.2287		
(87)	-0.1367		
(88)	-1.7281		
(89)	-1.7753		
(90)	-1.8381		

Variable: A:EJEMPLOS.VAR9 (length = 100)

(1)	-0.31452	(19)	-0.3985	(37)	-1.3914	(55)	-0.0466
(2)	-0.9949	(20)	1.42	(38)	-2.5599	(56)	-0.524
(3)	-0.9648	(21)	0.3286	(39)	1.435	(57)	1.3753
(4)	-0.2845	(22)	-1.7209	(40)	8.5E-3	(58)	-1.3678
(5)	0.3557	(23)	2.42	(41)	-0.3319	(59)	3.5799
(6)	0.63027	(24)	-0.2837	(42)	3.3211	(60)	2.2254
(7)	-0.7912	(25)	-1.2618	(43)	0.1362	(61)	-1.451
(8)	1.0868	(26)	-2.8942	(44)	-0.9068	(62)	-0.8393
(9)	-1.322	(27)	1.3123	(45)	0.3695	(63)	-1.2911
(10)	-1.2702	(28)	0.0542	(46)	-1.9913	(64)	0.7697
(11)	1.1269	(29)	-1.0449	(47)	2.1682	(65)	-1.1105
(12)	-0.3562	(30)	4.3359	(48)	2.2527	(66)	4.0087
(13)	-0.5442	(31)	0.3189	(49)	-2.4948	(67)	0.5382
(14)	-0.5592	(32)	1.8569	(50)	-0.7978	(68)	-1.1964
(15)	-1.1554	(33)	0.0419	(51)	-0.1324	(69)	1.426
(16)	-0.3556	(34)	-1.0754	(52)	0.3384	(70)	1.3579
(17)	-0.6106	(35)	1.6367	(53)	0.4455	(71)	0.7949
(18)	2.7563	(36)	2.3714	(54)	3.5155	(72)	1.7128

(73)	0.1498	(91)	0.2513
(74)	-0.4022	(92)	1.3183
(75)	-0.1182	(93)	-1.5883
(76)	0.374	(94)	2.6835
(77)	-1.6581	(95)	1.8244
(78)	2.4751	(96)	1.894
(79)	0.4714	(97)	2.2221
(80)	0.9164	(98)	-2.0787
(81)	-1.4262	(99)	-0.3906
(82)	1.3239	(100)	-3.6857
(83)	2.1903		
(84)	2.115		
(85)	1.242		
(86)	-1.9377		
(87)	-0.4143		
(88)	-1.3195		
(89)	-1.7193		
(90)	1.5622		

Thr Jul 6 1989 02:51:25 AM Page 1

Variable: A:EJEMPLOS.VAR10 (length = 100)

(1)	-0.3145	(19)	0.2343	(37)	-0.3574	(55)	-0.4068	(73)	1.9826
(2)	-0.9949	(20)	0.5505	(38)	-0.864	(56)	7E-3	(74)	-0.8313
(3)	-0.9648	(21)	1.3862	(39)	0.1521	(57)	1.9093	(75)	1.6959
(4)	-0.2845	(22)	-0.7047	(40)	-0.0946	(58)	-0.2632	(76)	-0.4965
(5)	0.3557	(23)	1.5184	(41)	1.2118	(59)	2.7266	(77)	-0.6188
(6)	0.6302	(24)	1.1E-3	(42)	-1.1963	(60)	-1.1266	(78)	-1.8297
(7)	-0.7912	(25)	-0.6378	(43)	-0.6393	(61)	0.4688	(79)	-0.1826
(8)	1.0868	(26)	-1.8498	(44)	-2.201	(62)	-0.9671	(80)	0.6884
(9)	-1.322	(27)	2.8156	(45)	-1.1303	(63)	-0.0741	(81)	-3.7242
(10)	-1.2702	(28)	0.5094	(46)	-1.3178	(64)	0.762	(82)	-0.5497
(11)	1.1269	(29)	-0.7698	(47)	0.4887	(65)	-2.1106	(83)	0.717
(12)	-0.3562	(30)	1.7526	(48)	0.1839	(66)	2.0841	(84)	0.517
(13)	-0.2925	(31)	1.1125	(49)	-1.0186	(67)	1.1601	(85)	0.3032
(14)	0.2366	(32)	0.0687	(50)	2.8557	(68)	0.6989	(86)	-0.1552
(15)	-0.3835	(33)	0.5722	(51)	-2.7834	(69)	0.3402	(87)	-0.9002
(16)	-0.1279	(34)	1.0634	(52)	0.0738	(70)	3.6337	(88)	-1.8722
(17)	-0.8952	(35)	-0.9754	(53)	0.6067	(71)	-3.0673	(89)	0.7529
(18)	2.2521	(36)	2.8121	(54)	0.6459	(72)	1.0621	(90)	-0.79

(91)	-0.6431
(92)	1.1138
(93)	0.2741
(94)	0.0739
(95)	2.5824
(96)	-0.253
(97)	0.6939
(98)	-1.1983
(99)	-0.6122
(100)	-2.1294

Thr Jul 6 1989 02:53:44 AM Page 1

Variable: A:EJEMPLOS.VAR11 (length = 100)

(1)	-0.3145	(19)	-1.0552	(37)	0.1328	(55)	0.0117	(73)	0.5066
(2)	-0.9949	(20)	-0.7662	(38)	-2.0321	(56)	-1.8888	(74)	-0.4432
(3)	-0.9648	(21)	-1.6682	(39)	1.8646	(57)	0.0948	(75)	-1.1451
(4)	-0.2845	(22)	-0.2346	(40)	0.2816	(58)	-1.0993	(76)	0.576
(5)	0.3557	(23)	1.7576	(41)	-0.4311	(59)	0.3545	(77)	-1.3224
(6)	0.6302	(24)	-0.2864	(42)	1.7641	(60)	-0.3403	(78)	0.5064
(7)	-0.7912	(25)	-0.2332	(43)	0.4961	(61)	-1.2897	(79)	0.4844
(8)	1.0868	(26)	-1.4285	(44)	0.349	(62)	1.0877	(80)	-0.6829
(9)	-1.322	(27)	4.0713	(45)	0.4662	(63)	-1.215	(81)	4.1082
(10)	-1.2702	(28)	-0.7083	(46)	0.545	(64)	0.7384	(82)	2.4495
(11)	1.1269	(29)	-0.5653	(47)	0.5376	(65)	1.1545	(83)	-1.9506
(12)	-0.3562	(30)	2.7502	(48)	1.8344	(66)	2.0909	(84)	-0.98
(13)	-0.335	(31)	-0.8996	(49)	0.0874	(67)	-0.6754	(85)	0.3799
(14)	-0.7405	(32)	0.0726	(50)	1.1861	(68)	0.5189	(86)	1.6661
(15)	-0.6992	(33)	1.2364	(51)	1.5888	(69)	1.0663	(87)	0.3183
(16)	-0.2669	(34)	-0.5105	(52)	0.3118	(70)	-0.4589	(88)	-0.6075
(17)	0.1929	(35)	1.2868	(53)	0.2145	(71)	5.5352	(89)	-1.1244
(18)	2.1171	(36)	0.6229	(54)	-0.6065	(72)	0.2268	(90)	-0.7167

(91)	-9.3E-3
(92)	1.5577
(93)	-1.8441
(94)	0.7714
(95)	1.1497
(96)	0.5841
(97)	0.6991
(98)	-0.7733
(99)	-0.2808
(100)	-1.4197

ESTIMACIONES FORECAST PLUS

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

113 Observations Used For Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = 3008303
Residual Variance = 24066.42

Parameter	Value	T-Statistic
AR(1)	-0.2236	-2.56

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

113 Observations Used For Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = 2448564
Residual Variance = 19588.51

Parameter	Value	T-Statistic
SAR(1)	-0.4293	-4.65

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

113 Observations Used For Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = 2654220
Residual Variance = 21233.76

Parameter	Value	T-Statistic
SMA(1)	0.4527	4.70

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

113 Observations Used For Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = 2379140
Residual Variance = 19186.61

Parameter	Value	T-Statistic
AR(1)	-0.1680	-1.89
SAR(1)	-0.4334	-4.65

Correlation Matrix

	AR(1)	SAR(1)
AR(1)	1.00	
SAR(1)	0.00	1.00

INGRESOS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

113 Observations Used For Parameter Estimates
Degree of Regular Differencing = 1
Degree of Seasonal Differencing = 1
Seasonal Period = 12

Sum of Squared Errors = 2485194
Residual Variance = 20041.88

Parameter	Value	T-Statistic
AR(1)	-0.2530	-2.91
SMA(1)	0.4727	4.87

Correlation Matrix

	AR(1)	SMA(1)
AR(1)	1.00	
SMA(1)	0.00	1.00

PRIMERAS ESTIMACIONES STATGRAPHICS

```
ESTIMATION BEGINS .....
--------------------------------------------------------------------------------
                         SUMMARY OF FITTED MODEL
--------------------------------------------------------------------------------
         parameter      estimate  stnd.error     t-value   prob(>~t~)
         AR ( 1)        -.31403      .09070     -3.46231      .00076
         MEAN          -76.59663  1308.07612    -.05856      .95341
         CONSTANT      -69.03640
--------------------------------------------------------------------------------
MODEL FITTED TO DIFFERENCES OF ORDER 1
MODEL FITTED TO SEASONAL DIFFERENCES OF ORDER 1 WITH SEASONAL LENGTH = 12
ESTIMATED WHITE NOISE VARIANCE = 3.32122E8 WITH 110 DEGREES OF FREEDOM.
CHI-SQUARE TEST STATISTIC ON FIRST 20 RESIDUAL AUTOCORRELATIONS = 27.0836
    WITH PROBABILITY OF A LARGER VALUE GIVEN WHITE NOISE = 0.10272

NUMBER OF ITERATIONS PERFORMED: 1

Press ENTER to continue.



 1HELP   2LABEL  3SAVSC  4RECORD 5      6      7      8       9REVIEW 10QUIT
PRINT    TUE JAN 1  1980  01:09:00 AM  VERSION 1.1                   REC:OFF
```

```
ESTIMATION BEGINS .....
ITERATION 1:  RESIDUAL SUM OF SQUARES ..... 3.4302E10
ITERATION 2:  RESIDUAL SUM OF SQUARES ..... 3.40285E10
--------------------------------------------------------------------------------
                         SUMMARY OF FITTED MODEL
--------------------------------------------------------------------------------
         parameter      estimate  stnd.error     t-value   prob(>~t~)
         SAR( 12)       -.43212      .09944    -4.34545      .00003
         MEAN          -60.34979  1255.21130    -.04808      .96175
         CONSTANT      -52.53805
--------------------------------------------------------------------------------
MODEL FITTED TO DIFFERENCES OF ORDER 1
MODEL FITTED TO SEASONAL DIFFERENCES OF ORDER 1 WITH SEASONAL LENGTH = 12
ESTIMATED WHITE NOISE VARIANCE = 3.43718E8 WITH 99 DEGREES OF FREEDOM.
CHI-SQUARE TEST STATISTIC ON FIRST 20 RESIDUAL AUTOCORRELATIONS = 29.7255
     WITH PROBABILITY OF A LARGER VALUE GIVEN WHITE NOISE = 0.055398

NUMBER OF ITERATIONS PERFORMED: 3

Press ENTER to continue.


 1HELP   2LABEL   3SAVSC   4RECORD 5      6       7       8       9REVIEW 10QUIT
PRINT    TUE JAN 1  1980  01:21:00 AM  VERSION 1.1                       REC:OFF
```

```
ESTIMATION BEGINS .....
ITERATION 1:  RESIDUAL SUM OF SQUARES ..... 3.30701E10
ITERATION 2:  RESIDUAL SUM OF SQUARES ..... 3.27856E10
------------------------------------------------------------------------------
                          SUMMARY OF FITTED MODEL
------------------------------------------------------------------------------
          parameter      estimate  stnd.error     t-value   prob(>~t~)
          SMA( 12)         .52152      .10287     5.06976      .00000
          MEAN          -75.08990   901.85959    -.08326       .93381
          CONSTANT      -52.53805
------------------------------------------------------------------------------
MODEL FITTED TO DIFFERENCES OF ORDER 1
MODEL FITTED TO SEASONAL DIFFERENCES OF ORDER 1 WITH SEASONAL LENGTH = 12
ESTIMATED WHITE NOISE VARIANCE = 3.31158E8 WITH 99 DEGREES OF FREEDOM.
CHI-SQUARE TEST STATISTIC ON FIRST 20 RESIDUAL AUTOCORRELATIONS = 29.005
     WITH PROBABILITY OF A LARGER VALUE GIVEN WHITE NOISE = 0.0659064

NUMBER OF ITERATIONS PERFORMED: 3

Press ENTER to continue.


 1HELP   2LABEL  3SAVSC  4RECORD 5      6      7       8       9REVIEW 10QUIT
PRINT    TUE JAN 1  1980  12:35:00 AM  VERSION 1.1                      REC:OFF
```

```
ESTIMATION BEGINS .....
ITERATION 1:  RESIDUAL SUM OF SQUARES ..... 3.16331E10
ITERATION 2:  RESIDUAL SUM OF SQUARES ..... 3.13906E10
--------------------------------------------------------------------------------
                        SUMMARY OF FITTED MODEL
--------------------------------------------------------------------------------
         parameter      estimate  stnd.error     t-value    prob(>~t~)
         AR ( 1)        -.25870      .09014     -2.86987      .00503
         SAR( 12)       -.38925      .09715     -4.00666      .00012
         MEAN          -81.04521  1047.39068     -.07738      .93848
         CONSTANT      -66.12981
--------------------------------------------------------------------------------
MODEL FITTED TO DIFFERENCES OF ORDER 1
MODEL FITTED TO SEASONAL DIFFERENCES OF ORDER 1 WITH SEASONAL LENGTH = 12
ESTIMATED WHITE NOISE VARIANCE = 3.20308E8 WITH 98 DEGREES OF FREEDOM.
CHI-SQUARE TEST STATISTIC ON FIRST 20 RESIDUAL AUTOCORRELATIONS = 18.5635
     WITH PROBABILITY OF A LARGER VALUE GIVEN WHITE NOISE = 0.419151

NUMBER OF ITERATIONS PERFORMED: 3

Press ENTER to continue.

 1HELP   2LABEL  3SAVSC  4RECORD 5      6      7       8       9REVIEW 10QUIT
PRINT    TUE JAN 1  1980  12:12:00 AM  VERSION 1.1                    REC:OFF
```

```
ESTIMATION BEGINS .....
ITERATION 1:  RESIDUAL SUM OF SQUARES ..... 2.93536E10
ITERATION 2:  RESIDUAL SUM OF SQUARES ..... 2.91521E10
--------------------------------------------------------------------------------
                          SUMMARY OF FITTED MODEL
--------------------------------------------------------------------------------
          parameter       estimate  stnd.error      t-value    prob(>~t~)
          AR ( 1)          -.33390      .09525     -3.50558      .00069
          SMA( 12)          .51940      .10326      5.02984      .00000
          MEAN           -80.31197   624.07430      -.12869      .89787
          CONSTANT       -70.08036
--------------------------------------------------------------------------------
MODEL FITTED TO DIFFERENCES OF ORDER 1
MODEL FITTED TO SEASONAL DIFFERENCES OF ORDER 1 WITH SEASONAL LENGTH = 12
ESTIMATED WHITE NOISE VARIANCE = 2.97356E8 WITH 98 DEGREES OF FREEDOM.
CHI-SQUARE TEST STATISTIC ON FIRST 20 RESIDUAL AUTOCORRELATIONS = 11.9829
     WITH PROBABILITY OF A LARGER VALUE GIVEN WHITE NOISE = 0.848117

NUMBER OF ITERATIONS PERFORMED: 3

Press ENTER to continue.

 1HELP   2LABEL  3SAVSC  4RECORD 5       6       7       8       9REVIEW 10QUIT
PRINT    TUE JAN 1  1980  12:46:00 AM  VERSION 1.1                      REC:OFF
```

SEGUNDAS ESTIMACIONES STATGRAPHICS

Estimation begins.....
Initial: RSS = 3.00818E10 b = -0.221789 -52.5381
Final: RSS = 3.00816E10 ...stopped on criterion 2

Summary of Fitted Model for: ingresos

Parameter	Estimate	Stnd.error	T-value	P-value
AR (1)	-.22247	.09276	-2.39850	.01813
MEAN	-67.30574	1269.39863	-.05302	.95781
CONSTANT	-82.27949			

Model fitted to differences of order 1
Model fitted to seasonal differences of order 1 with seasonal length = 12
Estimated white noise variance = 2.71006E8 with 111 degrees of freedom.
Estimated white noise standard deviation (std err) = 16462.3
Chi-square test statistic on first 20 residual autocorrelations = 21.8291
with probability of a larger value given white noise = 0.292842
Backforecasting: no Number of iterations performed: 1

```
Initial:     RSS = 3.41693E10  b = 0.1 -52.5381
Iteration 1: RSS = 2.75291E10  b = -0.257924 -19.8993
Iteration 2: RSS = 2.6732E10  b = -0.401067 -59.6504
Iteration 3: RSS = 2.66992E10  b = -0.432895 -58.6063
Final:       RSS = 2.66988E10   ...stopped on criterion 2
```

Summary of Fitted Model for: ingresos

Parameter	Estimate	Stnd.error	T-value	P-value
SAR(12)	-.43664	.09634	-4.53242	.00001
MEAN	-60.19718	1047.13987	-.05749	.95426
CONSTANT	-86.48156			

```
Model fitted to differences of order 1
Model fitted to seasonal differences of order 1 with seasonal length = 12
Estimated white noise variance = 2.40533E8 with 111 degrees of freedom.
Estimated white noise standard deviation (std err) = 15509
Chi-square test statistic on first 20 residual autocorrelations = 20.0084
    with probability of a larger value given white noise = 0.394074
Backforecasting: no                      Number of iterations performed: 4
```

```
Estimation begins.....
Initial:     RSS = 2.96522E10  b = 0.1 -52.5381
Iteration 1: RSS = 2.67757E10  b = 0.356523 -47.1103
Iteration 2: RSS = 2.65214E10  b = 0.453927 -71.1712
Final:       RSS = 2.65215E10   ...stopped on criterion 2
```

Summary of Fitted Model for: ingresos

Parameter	Estimate	Stnd.error	T-value	P-value
SMA(12)	.45259	.10220	4.42856	.00002
MEAN	-71.79397	860.78387	-.08341	.93368
CONSTANT	-71.79397			

```
Model fitted to differences of order 1
Model fitted to seasonal differences of order 1 with seasonal length = 12
Estimated white noise variance = 2.38932E8 with 111 degrees of freedom.
Estimated white noise standard deviation (std err) = 15457.4
Chi-square test statistic on first 20 residual autocorrelations = 21.3678
     with probability of a larger value given white noise = 0.316839
Backforecasting: no                    Number of iterations performed: 3
```

```
Initial:     RSS = 3.24748E10  b = -0.221789 0.1 -52.5381
Iteration 1: RSS = 2.61018E10  b = -0.209498 -0.263451 -29.1401
Iteration 2: RSS = 2.53404E10  b = -0.215042 -0.407432 -55.3466
Final:       RSS = 2.53431E10   ...stopped on criterion 2
```

Summary of Fitted Model for: ingresos

Parameter	Estimate	Stnd.error	T-value	P-value
AR (1)	-.22289	.09303	-2.39579	.01827
SAR(12)	-.48451	.09767	-4.96095	.00000
MEAN	-64.03302	858.95118	-.07455	.94071
CONSTANT	-116.24539			

```
Model fitted to differences of order 1
Model fitted to seasonal differences of order 1 with seasonal length = 12
Estimated white noise variance = 2.30392E8 with 110 degrees of freedom.
Estimated white noise standard deviation (std err) = 15178.7
Chi-square test statistic on first 20 residual autocorrelations = 12.5208
    with probability of a larger value given white noise = 0.819218
Backforecasting: no                     Number of iterations performed: 3
```

```
Initial:     RSS = 2.81655E10  b = -0.221789 0.1 -52.5381
Iteration 1: RSS = 2.51624E10  b = -0.216227 0.369432 -53.9996
Iteration 2: RSS = 2.48585E10  b = -0.23728 0.478538 -80.2948
Final:       RSS = 2.48524E10  ...stopped on criterion 2
```

Summary of Fitted Model for: ingresos

Parameter	Estimate	Stnd.error	T-value	P-value
AR (1)	-.25049	.09233	-2.71305	.00774
SMA(12)	.47022	.10242	4.59116	.00001
MEAN	-79.07205	652.40079	-.12120	.90375
CONSTANT	-98.87855			

```
Model fitted to differences of order 1
Model fitted to seasonal differences of order 1 with seasonal length = 12
Estimated white noise variance = 2.25931E8 with 110 degrees of freedom.
Estimated white noise standard deviation (std err) = 15031
Chi-square test statistic on first 20 residual autocorrelations = 11.5848
     with probability of a larger value given white noise = 0.867917
Backforecasting: no                           Number of iterations performed: 3
```

PRONOSTICO STATGRAPHICS

Variable: A:PRONOST.matriz (length = 12 3)

(1,1) 156016	(1,2) 126221	(1,3) 185811
(2,1) 181956	(2,2) 144721	(2,3) 219190
(3,1) 127474	(3,2) 83065.8	(3,3) 171882
(4,1) 158310	(4,2) 107958	(4,3) 208662
(5,1) 182550	(5,2) 126835	(5,3) 238265
(6,1) 229938	(6,2) 169344	(6,3) 290531
(7,1) 258642	(7,2) 193532	(7,3) 323752
(8,1) 281496	(8,2) 212164	(8,3) 350829
(9,1) 303564	(9,2) 230252	(9,3) 376876
(10,1) 222358	(10,2) 145271	(10,3) 299445
(11,1) 185263	(11,2) 104578	(11,3) 265948
(12,1) 187275	(12,2) 103145	(12,3) 271405

PRONOSTICO: 12 Primeros valores de la matriz

DATOS REALES QUE SE PRONOSTICARON

CUADRO IV-11

VIAJEROS AL INTERIOR

AÑOS TRIMESTRES Y MESES	NUMERO DE TURISTAS 1/ (miles) Total	Vía Aérea 1/	Vía Terrestre 2/	GASTO MEDIO (dólares) Total	Vía Aérea	Vía Terrestre	GASTO MEDIO DIARIO (dólares) Total	Vía Aérea	Vía Terrestre	PERMANENCIA MEDIA (días) Total	Vía Aérea	Vía Terrestre
[illegible]	3 767	2 172	1 [illegible]	373.[illegible]	510.[illegible]	177.[illegible]	[illegible] 15	[illegible]	14.[illegible]	10.3	[illegible]	12.5
[illegible]	4 749	[illegible]	1 757	[illegible]	[illegible]	121.15	37.[illegible]	54 4[illegible]	12.[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	4 [illegible]	3 [illegible]	1 [illegible]	419 57	543 76	[illegible]	44 [illegible]	64 70	16.77	[illegible]	[illegible]	11.[illegible]
[illegible]	4 [illegible]	2 [illegible]	1 515	[illegible].72	[illegible] 74	187.[illegible]	44.[illegible]	[illegible]	18.[illegible]	9.1	8.5	[illegible]
[illegible]	4 [illegible]	2 [illegible]	1 675	[illegible] 41	[illegible]	[illegible]	[illegible]	51 [illegible]	17.[illegible]	[illegible]	9.7	[illegible]
I	1 719	[illegible]	373	[illegible]	576 37	[illegible]	47 42	[illegible] 15	18.4[illegible]	10 9	10.3	12.[illegible]
II	1 [illegible]	[illegible]	401	[illegible]	477 [illegible]	176.97	[illegible]	[illegible]	[illegible]	9.5	[illegible]	[illegible]
III	1 [illegible]1	[illegible]	422	334 [illegible]	430 36	176 [illegible]	33 14	44 3[illegible]	19 [illegible]	10 1	[illegible]	[illegible]
IV	1 [illegible]	[illegible]	470	376 42	[illegible].17	[illegible]	[illegible] 12	54 11	17 [illegible]	[illegible]	[illegible]	10.[illegible]
[illegible]	5 4[illegible]7	3 6[illegible]	1 772	[illegible]	510.[illegible]	218 [illegible]	43.31	[illegible]	19 27	9.7	[illegible]	11 4
I	1 4[illegible]	1 [illegible]	[illegible]	404 57	562.7[illegible]	211 15	[illegible]	[illegible] 74	17.[illegible]	[illegible]	[illegible]	12 [illegible]
II	1 304	672	432	482 44	[illegible]	220 01	[illegible] 24	[illegible] 14	18.[illegible]	10.[illegible]	8.7	12.7
III	1215	777	43[illegible]	[illegible]	[illegible]	220 43	[illegible]	[illegible]	21 [illegible]	[illegible]	[illegible]	10 1
IV	1 42[illegible]	910	516	387 4[illegible]	[illegible]	[illegible] 76	42 57	[illegible]	18.4[illegible]	8.1	[illegible]	10 7
[illegible]	5 [illegible]	3 [illegible]	2 [illegible]26	44[illegible]	54[illegible]	2[illegible].19	42 4[illegible]	61 57	18 77	10 5	[illegible]	13.3
I	1 [illegible]	1 227	471	511.13	615.10	240 87	46 71	[illegible] 18	15 76	10 [illegible]	9 3	15.3
II	1 3[illegible]	[illegible]	[illegible]	433.27	530.[illegible]	[illegible]	42.[illegible]	[illegible]	[illegible] 25	10 7	8.0	12.7
III	1 237	740	497	471 70	530.23	[illegible].70	41 [illegible]	[illegible].75	21.32	10.1	[illegible]	11 [illegible]
IV	1 451	840	6[illegible]	[illegible] 75	512.[illegible]	[illegible]	37.[illegible]	[illegible]	18 [illegible]	10.7	8.8	13.7
[illegible]												
I	1 842	1 145	497	530 87	812.[illegible]	317 41	45.[illegible]	64.37	18 20	11.7	8.5	17.4
[illegible]												
Julio	495	3[illegible]	103	287.[illegible]	474 [illegible]	231.73	38.15	48 43	21 42	10.2	9.8	10.8
Agosto	44[illegible]	[illegible]	1[illegible]2	3[illegible].10	[illegible]	212.[illegible]	40.[illegible]	38.20	22.30	9.7	[illegible]	9.5
Septiembre	310	155	115	387 [illegible]	[illegible]	215.[illegible]	40.[illegible]	51 76	22.[illegible]	9.5	8.4	8.6
Octubre	357	230	128	[illegible].30	[illegible].24	197.[illegible]	47.[illegible]	61 72	22.25	[illegible]	8.2	8.9
Noviembre	447	[illegible]7	131	3[illegible]	[illegible].72	[illegible] 75	4[illegible]	[illegible].15	1[illegible] 76	[illegible]	7.8	[illegible]
Diciembre	[illegible]	3[illegible]	227	373.40	474.[illegible]	215.[illegible]	37.[illegible]	54.[illegible]	18.04	10.[illegible]	8.7	11.0
[illegible]												
Enero	517	[illegible]	13[illegible]	[illegible]	[illegible].12	[illegible]	[illegible]	[illegible]	15 15	10.1	9.8	13.[illegible]
Febrero	[illegible]	[illegible]	13[illegible]	530.21	[illegible]	307.[illegible]	[illegible]	[illegible]	[illegible]	10.9	9 4	13.3
Marzo	[illegible]	[illegible]	197	487.[illegible]	812.34	[illegible].71	[illegible]	[illegible]	14 7[illegible]	11.7	9.4	[illegible]
Abril	[illegible]	[illegible]	14[illegible]	472.[illegible]	[illegible].17	[illegible]	[illegible]	[illegible].31	[illegible].27	10.7	[illegible]	[illegible]
Mayo	411	[illegible]	147	410.76	911.[illegible]	251.[illegible]	[illegible]	[illegible].79	19 3[illegible]	10.1	8.4	[illegible]
Junio	[illegible]	[illegible]	1[illegible]	[illegible].31	[illegible] 91	[illegible]	[illegible]	[illegible].51	21.17	[illegible]	[illegible]	11.3
Julio	[illegible]	301	[illegible]	404.3[illegible]	510.[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10 [illegible]	8 5	12 1
Agosto	[illegible]	271	[illegible]	[illegible]	[illegible]	[illegible]	42 4[illegible]	[illegible]	[illegible]	10.3	9 5	11 7
Septiembre	[illegible]	170	125	[illegible]	[illegible].07	[illegible]	43.14	[illegible]	23.77	[illegible]	9 1	[illegible]
Octubre	[illegible]	311	14[illegible]	[illegible]	[illegible]	212.[illegible]	43.[illegible]	[illegible]	[illegible]	9.2	8.1	11 [illegible]
Noviembre	[illegible]	[illegible]	1[illegible]	434.77	[illegible].47	[illegible]	41.[illegible]	[illegible]	18.75	10.[illegible]	[illegible]	13.5
Diciembre	[illegible]	[illegible]	[illegible]	[illegible]	491.3[illegible]	277.7[illegible]	[illegible]	[illegible]	[illegible]	11.[illegible]	[illegible]	13.2
[illegible]												
Enero	3[illegible]	372	133	543 57	[illegible]	375.[illegible]	4[illegible]	61 [illegible]	19 [illegible]	11 8	10.2	10 8
Febrero	4[illegible]	374	124	52[illegible].25	[illegible] 81	[illegible]	4[illegible]	[illegible] 25	18 5[illegible]	10 7	8.7	10.7
Marzo	[illegible]	4[illegible]	2[illegible]	5[illegible]	812 [illegible]	[illegible]	41.84	63 81	17.3[illegible]	12.5	9.6	10.7

[illegible]

1/ [illegible]

2/ [illegible]

[illegible]

[illegible]

DATOS REALES (miles de dólares)

JULIO	200163.1
AGOSTO	192633.2
SEPTIEMBRE	128269.0
OCTUBRE	143243.7
NOVIEMBRE	180952.0
DICIEMBRE	264703.3
1989	
ENERO	274502.8
FEBRERO	263068.5
MARZO	333788.0

{8} MOOD,
GRAYBILL
Introducción a la Teoria Estadistica.
Ed. Aguilar

{9} MENDENHALL, WILLIAM.
Introducción a la Probabilidad
y Estadistica.
Ed. Wadsworth International Iberoamerica.

{10} Forecast Plus for de IBM PC.
Programs and Manual.
Copyright Walonik Associates, 1984,1985.

{11} Manual Statgraphics.
Copyright 1985 STSC, INC.

{12} VARGAS, AGUAYO ALBERTO
"Serie de Documentos de Investigación"
Manual Metodológico
Documento 21 "La Encuesta de Turismo Receptivo"
Banco de México
Junio 1980.

{13} RAMIREZ, BLANCO DANIEL
Teoria General de Turismo
Editorial Diana
1987

{14} México: Perfil de una Nación
INSTITUTO NACIOAL DE ESTADISTICA GEOGRAFIA E INFORMATICA
(INEGI)
BANCO NACIONAL DE MEXICO, FOMENTO CULTURAL BANAMEX.
1987.

BIBLIOGRAFIA:

(1) MAKRIDAKIS, WHELLWRIGHT
Forecasting Methods and Applications.
Ed. Wiley

(2) BOX, GEORGE E.P.
JENKINS, GWILYM M.
Time Series Analisys,
Forecasting and Control
Ed. Holden Day, 1976
Capitulos: 6,7,8,9.

(3) ANDERSON O.D.
Time Series Analysis and Forecasting
The Box-Jenkins Approach
Ed. Butterworths

(4) JUDGE, GEORGE G.
HILL, R.
GRIFFITHS, WILLIAM E.
Introduction to the Theory
and Practice of Econometrics
Ed. Wiley
Capitulo: 25

(5) DAMODAR, GUJARATI.
Econometria BÁsica.
Ed. Mc Graw Hill
Capitulos: 1,5,6,7,8

(6) COCHRAN, WILLIAM G.
TÁcnicas de Muestreo.
ED. CECSA
Capitulos: 5, 8, 11.

(7) DES RAJ
Teoría del Muestreo.
Fondo de Cultura EconÓmica.
Capitulos: 2, 4, 5.

(8) MOOD,
GRAYBILL
Introducción a la Teoria Estadistica.
Ed. Aguilar

(9) MENDENHALL, WILLIAM.
Introducción a la Probabilidad
y Estadistica.
Ed. Wadsworth International Iberoamerica.

(10) Forecast Plus for de IBM PC.
Programs and Manual.
Copyright Walonik Associates, 1984,1985.

(11) Manual Statgraphics.
Copyright 1985 STSC, INC.

(12) VARGAS, AGUAYO ALBERTO
"Serie de Documentos de Investigación"
Manual Metodológico
Documento 21 "La Encuesta de Turismo Receptivo"
Banco de México
Junio 1980.

(13) RAMIREZ, BLANCO DANIEL
Teoria General de Turismo
Editorial Diana
1987

(14) México: Perfil de una Nación
INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA
(INEGI)
BANCO NACIONAL DE MEXICO, FOMENTO CULTURAL BANAMEX.
1987.