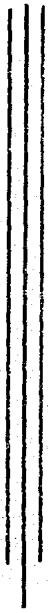


2 ej 44



UNIVERSIDAD NACIONAL
AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS



SOFTWARE DISPONIBLE
PARA
SERIES DE TIEMPO

TESIS PROFESIONAL
QUE PARA OBTENER EL TITULO DE
A C T U A R I A
P R E S E N T A
LUZ MARIA VARGAS MONTES

Director de Tesis:
GUSTAVO J VALENCIA RAMIREZ



TESIS CON
FALSA DE ORIGEN

MEXICO, D. F.

1989



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

SOFTWARE DISPONIBLE PARA SERIES DE TIEMPO

I	<u>INTRODUCCION Y OBJETIVOS</u>	Pag.
1.1	Introducción al tema de Series de Tiempo.	1
1.2	Que se busca con este trabajo.	13
II	<u>SERIES DE TIEMPO</u>	
2.1	Análisis de regresión.	15
2.2	Suavizamiento Exponencial.	19
2.3	Modelos ARIMA.	
2.3.1	Conceptos Fundamentales.	25
2.3.2	Modelos para Series de Tiempo Estacionarias.	28
2.3.3	Modelos para Series de Tiempo no Estacionarias.	38
2.3.4	Modelos Estacionales de Series de Tiempo.	41
III	<u>PARAMETROS NECESARIOS PARA LA IDENTIFICACION DE UN MODELO</u>	
3.1	Problemas en la Selección de un Modelo de Predicción.	44
3.2	Identificación de un Modelo.	48
IV	<u>CRITERIOS A CONSIDERAR PARA UN "BUEN" PAQUETE ESTADISTICO</u>	64
V	<u>SOFTWARE DISPONIBLE</u>	
5.1	Paquetes estadísticos con Series de Tiempo.	74
5.2	Análisis de cada paquete.	78
5.3	Ventajas y Facilidades.	189
VI	<u>CONCLUSIONES</u>	
6.1	Conclusiones acerca del Software.	194
6.2	El cuidado con los resultados.	197
	ANEXO A	
	Tabla de datos (serie G).	
	ANEXO B	
	Archivos ASCII, DIF, LOTUS y DBASE.	
	ANEXO C	
	Método Electra I.	
	ANEXO D	
	Bibliografía.	

I INTRODUCCION

1.1 INTRODUCCION AL TEMA DE SERIES DE TIEMPO.

Hoy en día, cuando se desea conocer sobre algún aspecto que se encuentre rodeado de incertidumbre, dentro de ciertas áreas, como por ejemplo la económica, no siempre se puede confiar en el instinto y la intuición, que muchas veces, no son lo suficientemente sensibles a las condiciones rápidamente cambiantes del ambiente en el que se está.

Ante esto, todas aquellas personas que deben tomar las decisiones necesitan conocer tanto herramientas para evaluar los problemas, como técnicas para formular modos de actuación alternativos. El análisis estadístico es un tipo de herramienta que actúa como soporte en la toma de decisiones, con un gran potencial influido por el modo en que los problemas están estructurados y son tomadas las decisiones. Mediante la estructuración de un problema, identificando supuestos y limitaciones, reuniendo y analizando datos numéricos relevantes, pueden aparecer más opciones, cuya variabilidad puede ser rigurosamente examinada.

De aquí, la gran importancia de poder tener los resultados anticipados en la toma de decisiones, sabiendo "que es lo mejor" dada una serie de suposiciones.

Definiendo entonces una predicción, como una estimación de cierta variable, de la cual no se conoce el valor para algún período en el tiempo.

Cuando se desea generar una predicción pueden analizarse los datos anteriores, es decir, valores que han representado el comportamiento del problema. Basándose en ellos puede generarse la predicción deseada.

Dentro de éstos datos disponibles se busca un posible patrón que distinga su comportamiento, para que con esta base y suponiendo que dicho patrón encontrado se conserva,

pueda hacerse la predicción. Es necesario hacer notar que este método producirá buenas predicciones si el patrón se conserva, de lo contrario se corre el riesgo de tener graves errores en la predicción, siendo esto un fenómeno que se pretende anticipar.

Se puede definir entonces a una Serie de Tiempo como una secuencia cronológica de observaciones de una variable en particular. Dichas observaciones, serán tomadas en puntos igualmente espaciados en el tiempo (anuales, trimestrales, mensuales, etc.).

Entre otras muchas aplicaciones podrían nombrarse el cálculo de una tasa de inflación, el movimiento de alguna acción dentro de la bolsa, el número de habitantes de cierta población de individuos, etc.

Con el fin de predecir a la variable de interés por medio del análisis de una Serie de Tiempo, se toman de este análisis algunos elementos importantes que a continuación se enuncian:

LA TENDENCIA, es el aumento o disminución que caracteriza a la serie durante un período de tiempo.

EL CICLO, se refiere a aumentos y disminuciones recurrentes cada determinado tiempo, éstos cambios son cuantificados de acuerdo a la magnitud de crestas y valles.

LAS VARIACIONES CICLICAS, son patrones periódicos en la serie generalmente anuales. Estos patrones pueden deberse al clima o la temporada en la que se encuentra la serie, son por lo regular mensuales o trimestrales. En todo esto, es importante notar que una sola observación anual no será suficiente para justificar dichas variaciones.

LAS FLUCTUACIONES IRREGULARES, son movimientos (aumentos o disminuciones) dentro de una serie que no siguen un patrón regular o esperado; estas fluctuaciones son causadas por hechos no usuales, los cuales no pueden ser predichos (temblores, guerras, etc.), o bien porque en la serie contemplada no se presentó anteriormente un caso de este tipo, lo que trae como consecuencia un error en el análisis.

Este tipo de observaciones llamadas aberrantes deben ser sometidas a un examen cuidadoso para ver si la razón de esta peculiaridad puede ser determinada. Han sido propuestas algunas reglas para desechar este tipo de observaciones, como es, realizar el análisis sin ellas, sin embargo el quitar observaciones de forma automática no es siempre el mejor procedimiento, ya que, a veces este o éstos puntos muestran una combinación de ciertas circunstancias que pudieran ser de vital interés en el estudio. Como regla general los puntos aberrantes deberán dejarse fuera del análisis sólo si se conocen las causas que los produjeron (como podría ser un error en la recolección de los datos).

De otra manera lo mejor será darles un tratamiento más cuidadoso.

Los fenómenos, es decir, el ciclo, la tendencia, las fluctuaciones irregulares y las variaciones cíclicas anteriormente mencionados pueden ocurrir ya sea por separado, o bien en una combinación de ellos, y en algunos casos todos al mismo tiempo. Por esto, no existe una técnica única, que ayude a predecir el comportamiento de la Serie de Tiempo, a la que pueda llamarse "mejor" que otra. Simplemente, se puede decir que una técnica puede ser buena para analizar una serie con una única tendencia, y no muy buena para analizar una que cuente con alguna combinación de los elementos mencionados.

Por lo anterior, uno de los problemas más importantes en la predicción, es el de encontrar el método apropiado para predecir el patrón de la serie de datos con la que se cuenta, una vez escogido el método que se va a utilizar, el siguiente paso a seguir, es el de analizar la serie de datos de tal forma que los diferentes elementos que la componen puedan ser estimados.

Las diferentes estimaciones de dichos elementos, deberán combinarse para generar una predicción, es decir, primero deberán identificarse y predecirse los componentes que forman parte de la serie, como son ciclos, tendencias, variaciones, etc. para que en segundo lugar se combinen esas predicciones y se obtenga otra que resultará ser la que represente al patrón de la serie en cuestión.

ERRORES EN LA PREDICCIÓN.

Por desgracia todas las situaciones que se desean predecir contienen un cierto grado de incertidumbre, para lo cual, es necesario incluir un componente de error dentro de la descripción de una Serie de Tiempo. Esta componente deberá contemplar el error que se comete al generar la predicción debido a fluctuaciones inexplicables o impredecibles en los datos, y además el error que producirá el como se predicen las demás variables necesarias para estimar a la variable de interés (en algunas técnicas de predicción). En la práctica, es difícil predecir con gran exactitud los elementos de ciclo, tendencia, fluctuaciones irregulares y variaciones cíclicas, por lo que la componente de error en la predicción representa los efectos combinados de los elementos de error, y la exactitud con que la técnica de predicción pueda representar el patrón de la serie.

De lo anterior se deduce que una componente de error grande producirá predicciones erróneas, o bien, indicará que la técnica utilizada de predicción no es capaz de representar los patrones de tendencia, ciclo, fluctuaciones y variaciones cíclicas de una manera aceptable y por lo tanto resulta ser inapropiada.

TIPOS DE PREDICCIÓN.

Existen principalmente dos tipos de predicción:

LA PREDICCIÓN PUNTUAL, es un valor que representa la mejor predicción de la variable de interés, sin embargo, por ser sólo un número el que se toma como predicción, generalmente no es igual al verdadero valor de la variable, y es por esto que no siempre resulta ser lo más adecuado, aunque por supuesto puede llegar a ser una buena aproximación.

Por lo anterior ese punto suele sustituirse por un INTERVALO DE CONFIANZA, que es un rango de valores calculado en el que se cree, con un cierto porcentaje de confianza, que el valor de la variable está dentro de dicho rango. Estos intervalos en la práctica suelen ser muy útiles para la toma de decisiones.

MÉTODOS DE PREDICCIÓN.

Se sabe que no existe una sola técnica de predicción ideal, además de que existe una gran diversidad de métodos para la predicción de eventos futuros, englobados principalmente en dos importantes tipos:

- Métodos cualitativos
- Métodos cuantitativos

Métodos Cualitativos de Predicción.

Estos métodos utilizan las opiniones de expertos en el tema a tratar, quienes predicen el comportamiento de la serie. Una de las posibles razones para escoger estos métodos podría ser el no contar con todas las observaciones pasadas de la serie en estudio, o bien, el que para algunos decisores sea difícil comprender las técnicas de predicción matemáticas, y por lo mismo no confíen en ellas para tomar sus decisiones.

Algunos decisores debido ya sea a la información de que disponen, o a su experiencia, son capaces de dar una predicción sobre determinado evento sin utilizar ninguna técnica propiamente dicha.

Estas técnicas son utilizadas además, para predecir cambios en los patrones de la secuencia de los datos. Esto se hace ya que al utilizar los valores pasados de la variable de interés, para generar predicciones acerca de dicha variable, se está basando en el hecho de que el patrón de esta variable permanecerá igual, por lo que el cambio de dicho patrón no podrá predecirse utilizando los datos anteriores.

A continuación se describen las técnicas de predicción cualitativa más comunes:

- *Curva Subjetiva Ajustada.*

Debe ser generada por expertos en el tema, basada en conocimientos previos de dicho tema en estudio. Uno de los grandes problemas de esta técnica, es el de decidir que tipo de curva será la apropiada (una curva-S, logarítmica, exponencial, etc.). Por lo que quien hace la predicción deberá ser una persona con experiencia y juicio.

- *Método Delphi.*

Consiste en formar un grupo de expertos quienes conocen el tema a tratar y son capaces de producir predicciones en dicho tema. Son dirigidos por una sola persona o por un pequeño grupo de individuos. Este método se desarrolla entrevistando a cada uno de los expertos por separado, siendo sus opiniones revisadas y resumidas por el coordinador (o grupo coordinador), quien dará a conocer el resumen de las opiniones a todos los expertos, sin dar a conocer el autor de cada opinión. Después de esto, de igual manera se vuelve a entrevistar a los expertos, y se procede a resumir la información, todo esto con el fin de llegar después de varias rondas a conclusiones específicas que serán utilizadas para la toma de decisiones.

- *Comparaciones Tecnológicas (Independientes en el tiempo)*

El método predice cambios en cierta área, monitoreando cambios que toman lugar en otra área similar. Este tipo de predicción tiene dos grandes problemas; El primero es que la persona que predice debe identificar que la tendencia de un área primaria sea similar al área de interés, y segundo que debe utilizar su experiencia para determinar la relación entre la tendencia primaria y los eventos a predecir.

Métodos Cuantitativos de Predicción.

Pueden utilizarse también, métodos matemáticos y estadísticos para efectuar la predicción.

En estas técnicas se analizan los datos anteriores, con el propósito de predecir valores futuros de la variable de interés.

Estos métodos es posible utilizarlos cuando los datos se tienen disponibles y tienen un patrón que se espera persista.

Las técnicas más comunes son:

- *Modelos Estructurales y Económétricos.*

Este método consiste en buscar un conjunto de funciones matemáticas, que representen una relación en el desarrollo del problema, en donde es necesario incluir un término de error, el cual es interpretado como una variable aleatoria sujeta a una distribución de probabilidad con media cero, convirtiéndolo así en un modelo estadístico.

Es importante notar, que en algunas circunstancias las variables exógenas al modelo, son conocidas al tiempo que la predicción es preparada, aunque muchas otras permanecen desconocidas hasta periodos futuros.

Se trata de que las técnicas sólo se basen en información disponible al hacer la predicción.

- *Modelos Determinísticos.*

Se trata aquí a la variable de interés como una función determinística del tiempo

$$Z_t = f(t)$$

y generalmente

$$Z_t = \alpha_0 + \alpha_1 t^1 + \dots + \alpha_k t^k$$

(Polinomio de grado k)

en la serie de tiempo Z_t .

Si tomamos a k suficientemente grande, en particular $k = (T - 1)$ (ésto es el número de observaciones disponibles menos uno), los coeficientes podrán escogerse de tal forma que el polinomio pase por cada observación, teniéndose T constantes ($\alpha_0, \alpha_1, \dots, \alpha_{T-1}$) desconocidas.

Pero posiblemente no se crea que los valores grandes de T sean realmente importantes en la evolución de Z_t ; además, si se usa un polinomio para predecir más de un periodo, pueden existir errores en la predicción y este modelo no nos muestra que tan grandes pueden ser.

Si truncamos el polinomio a menos de (T - 1) términos, no pasará por todas las observaciones y se tendrá que introducir un término de error. El problema entonces será determinar el grado del polinomio, y las

propiedades del término de error. Además hay que notar que los valores grandes de T dominarán la expresión.

- *Predicción Casual.*

Esta técnica de predicción, se refiere a la identificación de otras variables relacionadas con la variable que se desea predecir, una vez identificadas, se desarrolla un modelo estadístico que describe dicha relación, derivando de esto la predicción de la variable de interés. Entonces se define a la variable de interés, precisamente como la variable dependiente, y a las demás variables como las variables independientes. El trabajo consiste en estimar la relación funcional entre unas y otras variables, utilizando así, el valor de las variables independientes para estimar el valor futuro de la variable de interés (dependiente).

Sin embargo, esto tiene varias desventajas. Primero, la dificultad para desarrollar el modelo, segundo, que también se requieren (en algunos casos) los datos pasados de todas las variables incluidas en el modelo y no sólo de la variable dependiente; y tercero, que la habilidad de predecir el valor de la variable de interés, dependerá de la habilidad para predecir el valor de las variables independientes.

- *Formulas de predicción Ad Hoc.*

Otro tipo de técnicas de predicción que dependen sólo de los datos pasados es:

$$\hat{Z}_t(l) = f_l(Z_1, \dots, Z_{t-1}, Z_t)$$

donde $Z_t(l)$ denota la predicción hecha al tiempo t y f una función de los datos pasados dependiendo de la predicción l .

Se podría suponer que

$$\hat{Z}_t(l) = Z_t$$

o bien que el cambio sea igual al anterior

$$\hat{Z}_t(l) = Z_t + (Z_t - Z_{t-1})$$

pero en muchos casos esto no ocurre.

Por otro lado si se tuviesen suficientes observaciones podría obtenerse un promedio de ellas.

Por ejemplo, de las últimas diez observaciones:

$$\hat{Z}_t(l) = 1/10 (Z_t + \dots + Z_{t-9}).$$

Sin embargo, podríamos querer dar un mayor peso a observaciones más recientes y uno menor a las pasadas,

para lo cual habría que modificar con "pesos" a las observaciones.

Existen muchas técnicas que consideran algunos de los aspectos anteriores, una de las más usuales es la de "Suavizamiento Exponencial" donde se considera :

$$(1 - \beta), (1 - \beta)\beta, (1 - \beta)\beta^2$$

donde

$$0 < \beta < 1$$

se dice entonces que

$$\hat{Z}_t(1) = (1-\beta) Z_t + (1-\beta)\beta Z_{t-1} + (1-\beta)\beta^2 Z_{t-2} + \dots$$

declina exponencialmente y se inclina hacia el promedio de la serie, con un peso que varía exponencialmente dependiendo de β .

Si se escoge β pequeña, al principio se tendrán valores grandes pero declinará rápidamente, y si por el contrario se escoge a β grande, al principio se tendrán valores pequeños que decaerán lentamente.

Cuando se desea predecir un periodo grande, se requiere tener un número considerable de datos previos, para lo cual se llegó a

$$\hat{Z}_t(1) = (1 - \beta)Z_t + \beta Z_{t-1}(1).$$

- Análisis de Series de Tiempo.

La característica principal del análisis de Series de Tiempo, es que la secuencia de observaciones de una variable dada, es vista como una realización de una distribución conjunta de variables aleatorias.

La secuencia de observaciones Z_1, Z_2, \dots, Z_n es una función de distribución conjunta $P_{1\dots N}(Z_1, Z_2, \dots, Z_n)$, donde los índices $1\dots N$ sea cual sea la función de densidad, pueda hacerse inferencia sobre observaciones obteniendo resultados para observaciones aun no realizadas. Este tipo de modelos se ven como un proceso estocástico, y se dice que la secuencia de observaciones va de acuerdo a leyes de probabilidad.

Por ejemplo, tómesese un camino aleatorio en el que cada cambio en el tiempo es independiente, y con una distribución probabilística con media cero.

Así

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

$$Z_t = Z_{t-1} + U_t.$$

U_t es una variable aleatoria con media cero e independiente para cada periodo

$$Z_1 = Z_0 + U_1$$

$$Z_2 = Z_0 + U_1 + U_2$$

.....

$$Z_t = Z_0 + U_1 + \dots + U_t$$

Suponga que se tiene Z_1, Z_2, \dots, Z_T y se desea Z_{T+1} . Note que Z_{T+1} es una variable aleatoria debido a que Z_T es fija y Z_{T+1} es igual a una observación, Z_T más una variable aleatoria U_{T+1} , además U_{T+1} puede describirse por una distribución de probabilidad de Z_{T+1} dada por los datos anteriores, Z_1, Z_2, \dots, Z_T y la media dada por $E(X|Y)$.

Para este caso

$$\begin{aligned} E(Z_{T+1} | \dots, Z_{T-1}, Z_T) &= E(Z_T + U_{T+1} | \dots, Z_{T-1}, Z_T) \\ &= E(Z_T | \dots, Z_{T-1}, Z_T) + E(U_{T+1} | \dots, Z_{T-1}, Z_T) \\ &= Z_T + E(U_{T+1}) \\ &= Z_T \end{aligned}$$

De aquí que la posición esperada para el periodo $T+1$, es precisamente la del periodo T , es decir Z_T , note que se usa el hecho de que los demás datos anteriores de Z , no dicen nada acerca del siguiente valor de U , siendo esto debido a su independencia.

La varianza de Z_{T+1} será

$$\begin{aligned} V(Z_{T+1} | \dots, Z_{T-1}, Z_T) &= V(Z_T + U_{T+1} | \dots, Z_{T-1}, Z_T) \\ &= 0 + V(U_{T+1}) \\ &= \sigma_U^2 \end{aligned}$$

Nuevamente, utilizando que U_{T+1} es independiente de las demás, se concluye que σ_U^2 es la varianza de cualquier U .

El conocimiento de Z_{T+1} , toma entonces la forma del conocimiento acerca de la distribución de probabilidad de cada una de ellas.

Si por ejemplo U_T se distribuye como Normal, implicaría que la distribución de Z_{T+1} dadas las observaciones pasadas, es una campana centrada en Z_T , por lo que si se desea tomar a la predicción como un sólo punto (y no a un intervalo) lo ideal sería escoger a Z_T . En este punto la media tiene la

propiedad de tener el mínimo error cuadrático medio para el valor predicho.

Así

$$\begin{aligned} Z_T(2) &= E(Z_{T+2} | \dots Z_{T-1}, Z_T) \\ &= E(Z_T + U_{T+1} + U_{T+2} | \dots Z_{T-1}, Z_T) \\ &= Z_T + E(U_{T+1}) + E(U_{T+2}) \\ &= Z_T \end{aligned}$$

y

$$\begin{aligned} V(Z_{T+2} | \dots Z_{T+1}, Z_T) &= V(Z_T + U_{T+1} + U_{T+2} | \dots Z_{T-1}, Z_T) \\ &= 0 + \sigma^2 + \sigma^2 \\ &= 2\sigma^2 \end{aligned}$$

Es claro que la predicción con este método siempre será Z_T y la varianza $l\sigma^2$ donde l es el tiempo que se predice a la variable, ésto es, la desviación estándar será de $(l)^{1/2}\sigma$ y junto con ésto variarán los intervalos.

Z_{T+1} y Z_{T+2} son variables aleatorias no independientes y conjuntamente distribuidas, ya que, ambos términos contienen a U_{T+1} como elemento.

Su covarianza es

$$\begin{aligned} C(Z_{T+1}, Z_{T+2} | \dots Z_{T-1}, Z_T) &= \\ C[(Z_T + U_{T+1})(Z_T + U_{T+1} + U_{T+2}) | \dots Z_{T-1}, Z_T] &= \\ = V(U_{T+1}) + C(U_{T+1}, U_{T+2}) &= \\ = \sigma^2 & \end{aligned}$$

La cual se obtuvo considerando la independencia de U_{T+1} y U_{T+2} .

Si U_t se distribuye de forma Normal, entonces Z_{T+1} y Z_{T+2} son distribuidas conjuntamente como Normal, y si se tienen la esperanza y la covarianza puede especificarse perfectamente la distribución conjunta.

Este modelo de caminata aleatoria da un ejemplo de como un modelo de Series de Tiempo, expresado como un proceso estocástico, puede utilizarse para obtener información de la futura evolución de dicha serie.

Ahora después de discutir algunas de las técnicas de predicción, es necesario escoger alguna de ellas, para lo

cual, se debe tomar en cuenta los siguientes factores:

- La forma de predicción deseada.
- La longitud del período.
- El patrón de los datos.
- El costo de la predicción.
- La exactitud deseada.
- La disponibilidad de los datos.
- La facilidad de operación y entendimiento.

El primer factor a considerar en la búsqueda de un método de predicción, es la forma en la cual se desea la predicción. Dependiendo de las necesidades de cada persona, puede escogerse entre obtener como predicción un sólo punto, o bien, un intervalo de confianza. Esto podrá influir en el método de predicción a escoger, debido a que existen algunos métodos que teóricamente ofrecen intervalos de confianza, mientras que otros sólo un punto.

El segundo factor es la longitud entre cada observación. Las predicciones pueden generarse para puntos en el tiempo que pueden ser diarios, semanales, mensuales, trimestrales, anuales, etc. Esta longitud puede clasificarse en:

- Inmediato : Menor a un mes.
- Corto Plazo : De uno a tres meses.
- Mediano Plazo : Más de 3 meses pero menor a 2 años.
- Largo Plazo : Dos años o más.

Generalmente cuando las predicciones son a largo plazo, puede provocar que el generar predicciones con gran exactitud sea más difícil, llegando a ser más útiles las técnicas cualitativas de predicción, cuando el lapso de tiempo entre observaciones se incrementa.

El tercer factor es el que siguen los datos, si este patrón existe ya sea con alguna tendencia, fluctuación, variación cíclica, o alguna combinación de ellas determinará la técnica a utilizar.

El costo de la predicción es también un factor relevante. Entre algunos de los costos a considerar se tienen, primero el costo del desarrollo del método (en algunos métodos es necesario el seguimiento de determinados procesos), donde la complejidad y el costo de éstos métodos varía de técnica a técnica. Segundo, el costo de la recolección de los datos que dependerá de la técnica utilizada, ya que, para algunas técnicas se requiere tan sólo un pequeño grupo de datos, mientras que en otras se necesitan un mayor número de ellos. Finalmente el costo de la operación actual de la técnica, ya que pueden existir diferencias en complejidad que varían de técnica a técnica y de persona a persona.

Otro factor, de igual importancia, es la exactitud que se desee para la predicción. Existen casos en los que una predicción que tiene un error de a lo más 20% puede ser

aceptable, pero para otros casos, un error de a lo más 1% podría ser desastroso.

Es necesario saber si se tienen o no disponibles los datos anteriores a la fecha de predicción, ya que por ejemplo, las técnicas cuantitativas requieren de dichos datos. Así, este factor es determinante para decidirse por cierta técnica. Además de esto, deberá considerarse la exactitud y el tiempo necesario para que los datos sean seleccionados, debido a que la falta tanto de exactitud, como de datos en el conjunto de observaciones, producirá obviamente la falta de exactitud en las predicciones.

Finalmente, la facilidad y el entendimiento de la técnica a utilizar, será también un factor importante que determinará el decidirse por utilizar cierta técnica de predicción, ya que si la persona que efectúe la predicción o el decisor no confía en los resultados de la técnica, o bien no los entiende, traerá como consecuencia el que dicho proceso de predicción no sea tomado en cuenta.

El buscar una técnica de predicción para ser utilizada en determinada situación, trae como consecuencia el encontrar una técnica que balancee los factores discutidos. Es obvio que el "mejor" método para cierta situación no es siempre el más "exacto". En vez de esto, el método para predecir que deberá utilizarse será el que cubra las necesidades de predicción al menor costo y con las menores inconveniencias.

1.2 OBJETIVOS DE ESTE TRABAJO.

Una vez habiendo explorado superficialmente las principales técnicas de predicción, el siguiente paso a seguir será el de tomar para este estudio el Análisis de Series de Tiempo.

El objetivo principal de este trabajo no es el de presentar la base teórica con todos los fundamentos matemáticos inherentes que sustentan el desarrollo de este análisis, sino que tomando como hipótesis que esta herramienta de predicción existe y es utilizada en muchos y muy diversos aspectos de la vida diaria, se intentará simplificar la forma de efectuar dicho análisis por medio de la utilización de paquetes computacionales en microcomputadoras PC's y compatibles.

Esto es muy interesante, debido a la creciente facilidad de disposición de computadoras personales que realicen los cálculos matemáticos de una forma rápida y precisa. Afortunadamente el hardware y el software de las computadoras personales están llegando a satisfacer las demandas de un mayor número de personas con requerimientos de un mayor poder de proceso. La memoria y la capacidad de almacenamiento han aumentado también, mientras que el precio se ha mantenido o en algunos casos ha disminuido.

Actualmente, ya se puede disponer de programas que lleven a cabo análisis estadísticos que pueden competir con los que tienen las grandes computadoras. Además que algunos de los distribuidores de programas de grandes sistemas ofrecen versiones para computadoras personales (por ejemplo el reciente SPSS/PC de SPSS Inc.). Los paquetes de grandes sistemas aunque ofrecen unas posibilidades de modelado y de pronóstico muy poderosas son muy costosos y a veces muy complicados de usar.

Así, el soporte en la toma de decisiones con sus exigencias en cuanto a análisis, suministrará un suelo fértil para el desarrollo de paquetes estadísticos.

Todas aquellas personas que tengan que tomar decisiones, y deseen utilizar algún método estadístico tendrán que escoger entre los productos de software disponibles, el "mejor" paquete para su uso personal. Entre las preguntas que intentarán contestar están las siguientes:

- ¿Qué características tienen los paquetes estadísticos disponibles comercialmente para las PC's?
- ¿Qué criterios deberán ser utilizados para comparar y evaluar el software estadístico?

A pesar de que las respuestas a estas preguntas serán distintas en cada caso individual, el marco para el análisis y la decisión será muy similar. Este trabajo mostrará precisamente el desarrollo de este marco de análisis y decisión "similar".

Se ha centrado la atención en programas de software para computadoras personales, tomando en cuenta que una amplia comunidad de usuarios se ve afectada por la disponibilidad de estas máquinas, y mientras que los programas estadísticos pueden no ser necesarios para todos los usuarios, los tipos de problemas y de decisiones que quedan sujetos al rigor de los análisis estadísticos serán probablemente de importancia estratégica para muchas personas.

Unos de los posibles usuarios para paquetes estadísticos en PC's serían los profesionales de la estadística y los analistas que han usado los paquetes de grandes sistemas para su toma de decisiones, así como los investigadores y por supuesto los alumnos del área estadística en diversas instituciones educativas.

En resumen, se puede decir que en este trabajo se buscará analizar el mayor número de paquetes estadísticos, se verá si con ellos puede o no efectuarse el Análisis de Series de Tiempo y se tratará de evaluar a dichos paquetes para que dependiendo de las necesidades de cada individuo, pueda elegirse el que en cada caso será el "mejor" software disponible para Series de Tiempo.

Finalmente, se presentarán resultados sobre lo obtenido tanto en materia de software, como resultados estadísticos obtenidos de los mejores paquetes.

II SERIES DE TIEMPO.

Antes de adentrarse en lo concerniente a este trabajo y discutir acerca del software disponible para Series de Tiempo, es necesario revisar rápidamente la teoría fundamental de lo que son las Series de Tiempo. Esto se hace con el fin de tener una visión mas amplia de lo que se busca en un paquete estadístico. De tal manera, que nos garantice que el análisis que estamos efectuando es realmente el que se desea.

Se discutirán tres tipos fundamentales en el análisis de Series de Tiempo :

- Análisis de Regresión.
- Suavizamiento Exponencial.
- Modelos ARIMA.

2.1 ANALISIS DE REGRESION

En esta sección se ilustrará rápidamente el procedimiento de predicción utilizando el modelo de regresión múltiple para conocer los valores futuros de una Serie de Tiempo [2].

El objetivo es conocer para un periodo t en el tiempo el valor denotado por Y_t , llamado "variable dependiente", el cual se cree que está parcialmente determinado por una o más "variables independientes" (X_1, X_2, \dots, X_p). Esto mientras se cuente con los valores para dichas variables para el periodo t en el cual se desea tener la predicción, así podremos predecir Y_t en base a las variables independientes X_1, X_2, \dots, X_p .

Una posible relación entre las variables de interés es la expresada en el siguiente Modelo de Regresión Múltiple

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp} + \epsilon_t$$

donde $\beta_0, \beta_1, \dots, \beta_p$ son constantes desconocidas. Dichas constantes son llamadas parámetros y relacionan Y_t con $X_{t1}, X_{t2}, \dots, X_{tp}$. El término ϵ_t es un componente irregular llamado componente aleatorio de error, que describe la influencia combinada en Y_t de otros factores, además de las variables independientes, como son el error de medición o el error estocástico que toma en cuenta el efecto en Y_t de todas las demás variables independientes que no fueron incluidas explícitamente en el modelo. Una posible razón para que dichas variables no sean incluidas en el modelo, es que no pueda tenerse acceso a la información que determina el valor de dichas variables para el periodo en cuestión, o bien, que ni siquiera se sepa que estas variables puedan afectarlo.

Se asume que el valor esperado de ϵ_t es cero, es decir, para un periodo suficientemente grande el promedio de los errores aleatorios será de cero. Sin embargo, en particular para algún periodo t podría no ser cero, consecuentemente la Serie de Tiempo Y_t puede representarse por un nivel promedio el cual cambia con el tiempo de acuerdo a la expresión $\beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp}$ combinada con las fluctuaciones aleatorias representadas por ϵ_t .

Usando los datos disponibles anteriores puede mostrarse que las estimaciones para $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ de "mínimos cuadrados" son b_0, b_1, \dots, b_p respectivamente, es decir, si se calcula para cada periodo t el error cuadrado $(Y_t - \hat{Y}_t)^2$ el cual es el cuadrado de la diferencia entre la observación actual Y_t y la predicción \hat{Y}_t hecha para ese periodo con la ecuación:

$$\hat{Y}_t = b_0 + b_1 X_{t1} + b_2 X_{t2} + \dots + b_p X_{tp}$$

dichos errores $(Y_t - \hat{Y}_t)$ deberán ser "razonablemente pequeños" para cada uno de los t periodos.

Así llamando a

$$\sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (\text{donde } n \text{ es el número de observaciones pasadas con las que se cuenta})$$

la suma de cuadrados de los errores que depende de las estimaciones por mínimos cuadrados de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$.

Se llaman a b_0, b_1, \dots, b_p las estimaciones por mínimos cuadrados de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ porque puede demostrarse por manipulaciones algebraicas que estas estimaciones dan un valor de

$$\sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

menor que cualquier otra estadística.

Estos cálculos pueden realizarse con la ayuda de paquetes computacionales de regresión, que hacen las estimaciones de mínimos cuadrados al dar como datos al paquete, la serie de observaciones anteriores.

Llamando entonces al nivel promedio de la serie μ_t que es

$$\mu_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp}$$

El valor μ_t representa el promedio de todos los valores de la variable dependiente Y_t , que pueden ser siempre calculados cuando los valores de las variables independientes son fijos. Si b_0, b_1, \dots, b_p son las estimaciones por mínimos cuadrados de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ entonces el punto estimado $\hat{\mu}_t$ es

$$\hat{Y}_t = b_0 + b_1 X_{t1} + b_2 X_{t2} + \dots + b_p X_{tp}$$

Entonces, si se asume que el promedio del valor de la componente de error es cero en un periodo grande, el valor supuesto para el siguiente ϵ_t es precisamente cero. De donde se sigue que

$$\begin{aligned} Y_t &= \mu_t + \epsilon_t \\ &= \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp} + \epsilon_t \end{aligned}$$

Notese que el número de periodos observados n deberá ser mayor que $(p+1)$, el número de parámetros en el modelo con el fin de calcular las estimaciones de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$. Este trabajo puede ser también ejecutado por algún paquete estadístico.

Mientras que las estimaciones por mínimos cuadrados $b_0, b_1, b_2, \dots, b_p$ son sólo los posibles valores de los parámetros desconocidos $\beta_0, \beta_1, \beta_2, \dots, \beta_p$. Se ve que

$\hat{Y}_t = b_0 + b_1 X_{t1} + b_2 X_{t2} + \dots + b_p X_{tp}$
 será diferente de

$$\mu_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp}$$

$$Y_t = \mu_t + \epsilon_t$$

Por ésto el paso a seguir es determinar que tanto difiere \hat{Y}_t de μ_t y Y_t . Esto es, se desea encontrar el intervalo de confianza para μ_t y Y_t .

Para ésto deben considerarse tres hipótesis sobre la componente aleatoria de error ϵ_t . La primera es que ϵ_t deberá seguir una función de distribución de probabilidad Normal, lo que implica que los valores de la variable dependiente Y_t que pueden calcularse cuando se conocen los valores de las variables dependientes son distribuidos en una curva Normal alrededor de μ_t . La segunda hipótesis es que la varianza de Y_t , la cual mide la dispersión de todos los posibles valores de la variable dependiente Y_t alrededor del nivel promedio μ_t deberá ser la misma para toda t . Esto es, que para toda combinación de valores de las variables independientes la varianza de Y_t será la misma. La hipótesis final es que los valores de la Serie de Tiempo Y_1, Y_2, Y_3, \dots en diferentes periodos son independientes uno de otro. A estas hipótesis se les llama hipótesis de inferencia. Es probable que en la práctica estas tres hipótesis no se cumplan exactamente para la construcción de un modelo debido a alguna situación particular. Por lo tanto, es necesario hacer notar que en la construcción del modelo se deben incluir las variables independientes que afecten "significativamente" a la variable dependiente. Además, se tienen como ayuda las gráficas de residuales en las que puede determinarse si las hipótesis están incluidas (en forma aproximada) en el modelo de regresión.

Usando la teoría estadística y las tres hipótesis, puede mostrarse que existe un error $E_{\mu_t}(100 - \alpha)$ tal que se tiene un $(100 - \alpha)$ por ciento de confianza que la predicción Y_t estará en un margen de $E_{\mu_t}(100 - \alpha)$ de Y_t . Entonces se dice que un intervalo con $(100 - \alpha)$ por ciento de confianza para Y_t es

$$(\hat{Y}_t - E_{\mu_t}(100 - \alpha), \hat{Y}_t + E_{\mu_t}(100 - \alpha))$$

Este intervalo dice que se tiene con un $(100 - \alpha)$ por ciento de confianza el valor actual de la Serie de Tiempo, es decir, Y_t no será menor a $(\hat{Y}_t - E_{\mu_t}(100 - \alpha))$ ni mayor a $(\hat{Y}_t + E_{\mu_t}(100 - \alpha))$.

El error $E_{\mu_t}(100 - \alpha)$ en el intervalo de confianza para μ_t mide la incertidumbre causada por el hecho de que las estimaciones por mínimos cuadrados b_0, b_1, \dots, b_p difieren de los verdaderos parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ y la diferencia con μ_t por la componente de error aleatorio ϵ_t . Aunque se supone que el valor de la esperanza de ϵ_t es cero, no siempre es cierto para un periodo particular t .

2.2 SUAVIZAMIENTO EXPONENCIAL.

La implementación de un procedimiento de suavizamiento exponencial consiste de varios pasos. Uno de ellos es el de escoger una constante de suavizamiento apropiada. Si una estimación con suavizamiento exponencial no produce predicciones adecuadas quizás es por un cambio en el patrón de la serie. Este cambio deberá detectarse para tomar una acción correctiva. Esta acción involucra siempre un cambio en la constante de suavizamiento. Puede efectuarse un análisis de los errores predichos para determinar cuando es necesario tomar una acción correctiva [2].

PREDICCIÓN DE SERIES DE TIEMPO SIN TENDENCIA.

Se supondrá en esta sección que el nivel promedio de la serie no cambia en el tiempo, o bien, cambia muy lentamente. Entonces un modelo apropiado sería:

$$y_t = \beta_0 + \epsilon_t$$

donde β_0 es el valor promedio de la serie, el cual no cambia (o lo hace muy lentamente) en el tiempo, combinados con una componente de error ϵ_t .

Aproximación por Suavizamiento Exponencial.

Tómese como suposición que al final de algún periodo, por ejemplo, $T-1$ se ha obtenido un conjunto de observaciones para la Serie de Tiempo, la cual se denota como y_1, y_2, \dots, y_{T-1} . Dadas estas observaciones se desea estimar β_0 para la cual su estimación por mínimos cuadrados será $b_0(T-1)$ (se denota de esta manera para indicar que $T-1$ fue la última observación obtenida).

$$b_0(T-1) = \bar{y} = \frac{\sum_{t=1}^{T-1} y_t}{(T-1)}$$

donde se ve que la estimación de β_0 es simplemente el promedio de la serie hasta el periodo $T-1$, por lo que la estimación para algún otro periodo, digase, $T-1+\tau$ (con τ un entero positivo) es $b_0(T-1)$.

Si se obtuviera una nueva observación y_T la estimación para β_0 sería entonces:

$$b_0(T) = \bar{y} = \frac{\sum_{t=1}^T y_t}{(T)}$$

con lo que se incorpora la nueva observación.

Otra forma de incorporar esta nueva observación, sería el utilizar el Suavizamiento Exponencial Simple. Este método, genera la predicción $b_0(T)$ en forma diferente. Intuitivamente, se puede ver como la suma de la estimación anterior $b_0(T-1)$ más una fracción del error predicho, que resulta del utilizar la estimación $b_0(T-1)$ para predecir el valor de la serie del período actual, esto es:

$$y_T - b_0(T-1)$$

siendo esta la diferencia entre el valor observado en el período T y la predicción hecha para el período T , en el período $T-1$. Si la fracción utilizada es α la estimación será:

$$b_0(T) = b_0(T-1) + \alpha [y_T - b_0(T-1)]$$

con $0 < \alpha < 1$.

Si la estimación anterior, producida para el período T estuvo muy por debajo de la observación, entonces la nueva estimación $b_0(T)$ será alta; si por el contrario la estimación estuvo por arriba de la observación, la nueva observación será baja. La magnitud del ajuste hacia arriba o hacia abajo es determinada por la magnitud del error predicho. Un error grande da como resultado un ajuste grande y viceversa.

Para simplificar la notación definase a $S_T = b_0(T)$. Entonces la ecuación queda:

$$\begin{aligned} S_T &= S_{T-1} + \alpha(y_T - S_{T-1}) \\ &= S_{T-1} + \alpha y_T - \alpha S_{T-1} \\ &= \alpha y_T + (1-\alpha) S_{T-1} \end{aligned}$$

La ecuación definida es llamada Suavizamiento Exponencial Simple.

Llamando a S_T la estimación suave o estadística suave. La fracción α es llamada la constante suave.

Examinando la Ecuación de Suavizamiento:

$$S_T = \alpha y_T + (1-\alpha)S_{T-1}$$

se puede ver que la estimación S_T esta basada en las observaciones y_1, y_2, \dots, y_T . Esto es cierto, mientras S_{T-1} sea el promedio de las observaciones y_1, y_2, \dots, y_{T-1} . Si se cambia el período origen de tal forma que ahora sea el período cero (llamese S_0). En la práctica la estimación S_0 podría obtenerse calculando el promedio de un conjunto inicial de observaciones en la Serie de Tiempo. Si no se tiene este conjunto inicial, S_0 es comúnmente tomada como la primera observación de la serie. Asumiendo que la ecuación

$$S_T = \alpha y_t + (1-\alpha)S_{T-1}$$

ha sido utilizada para actualizar las estimaciones para cada periodo t , con $1 < t < T$. Para este caso la estimación del periodo T , es decir, S_T puede mostrarse como una combinación lineal de las observaciones pasadas.

Considere

$$S_T = \alpha y_T + (1-\alpha)S_{T-1}$$

de donde

$$S_{T-1} = \alpha y_{T-1} + (1-\alpha)S_{T-2}$$

sustituyendo en la anterior

$$\begin{aligned} S_T &= \alpha y_T + (1-\alpha)[\alpha y_{T-1} + (1-\alpha)S_{T-2}] \\ &= \alpha y_T + \alpha(1-\alpha)y_{T-1} + (1-\alpha)^2 S_{T-2} \end{aligned}$$

si se sustituye recursivamente $S_{T-2}, S_{T-3}, \dots, S_1$ se obtiene

$$S_T = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots + \alpha(1-\alpha)^{T-1} y_1 + (1-\alpha)^T S_0$$

Entonces S_T , la estimación de β_0 en el periodo T puede expresarse en términos de las observaciones y_1, y_2, \dots, y_T y la estimación inicial S_0 . Los coeficientes $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \dots, \alpha(1-\alpha)^{T-1}$ son "pesos" que les son dados a las observaciones.

Los coeficientes decrecen en forma exponencial, haciendo que las observaciones más recientes, tengan una mayor contribución a la actual estimación de β_0 , mientras que para las observaciones más antiguas, la contribución será menor.

Entre más cercano a 1 sea el valor de α , las observaciones lejanas al tiempo T irán disminuyendo rápidamente; y para valores cercanos a 0 dichas observaciones disminuirán lentamente.

Como se ve el escoger la constante de suavizamiento no es fácil, y debe hacerse con cuidado, ya que esta constante determina en que grado las observaciones anteriores afectarán la predicción.

PREDICCIÓN DE SERIES DE TIEMPO CON TENDENCIA LINEAL.

Ahora se considerará la situación en la que el nivel promedio de la serie varía en el tiempo, de forma lineal. De donde el modelo apropiado para este tipo de series quedaría:

$$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

La expresión $\beta_0 + \beta_1 t$ indica la relación que guardan el nivel promedio de la serie y el tiempo. Además se agrega al modelo un término de error ϵ_t .

Si β_1 es mayor que cero esto reflejará que el nivel promedio de la serie se incrementa conforme el tiempo pasa, mientras que si β_1 es negativa implicará que el nivel de promedio de la serie se decrementa al pasar el tiempo.

Aproximación por Doble Suavizamiento Exponencial.

Suponiendo igualmente que para el periodo $T-1$, se cuenta con un conjunto de observaciones y_1, y_2, \dots, y_{T-1} que conforman la Serie de Tiempo, y se desea estimar los parámetros β_0 y β_1 con los cuales se generaran las predicciones para la serie.

Las estimaciones por mínimos cuadrados de β_0 y β_1 se denotarán por $b_0(T-1)$ y $b_1(T-1)$ respectivamente y son:

$$b_1(T-1) = \frac{(T-1) \sum_{t=1}^{T-1} y_t - \begin{bmatrix} T-1 \\ \sum_{t=1}^{T-1} t \end{bmatrix} \begin{bmatrix} T-1 \\ \sum_{t=1}^{T-1} y_t \end{bmatrix}}{(T-1) \sum_{t=1}^{T-1} t^2 - \begin{bmatrix} T-1 \\ \sum_{t=1}^{T-1} t \end{bmatrix}^2}$$

y

$$b_0(T-1) = \frac{\sum_{t=1}^{T-1} y_t}{T-1} - b_1(T-1) \begin{bmatrix} T-1 \\ \sum_{t=1}^{T-1} t \end{bmatrix} \frac{1}{T-1}$$

Estas estimaciones serán utilizadas para generar predicciones en el periodo $T-1$ para cualquier otro periodo deseado, digase $(T-1) + \tau$ donde τ es un entero positivo, esto es,

$$Y_{T-1+\tau}(T-1) = b_0(T-1) + b_1(T-1)(T-1+\tau)$$

Supongase que se desea incorporar una nueva observación y_T a la serie, las nueva estimaciones $b_0(T)$ y $b_1(T)$ podrán calcularse con el procedimiento de Doble Suavizamiento Exponencial. Esta aproximación da una estimación de β_1 como:

$$b_1(T) = \frac{\alpha}{1-\alpha} (S_T - S_T^{(2)})$$

donde S_T es la estimación por Suavizamiento Simple, que se obtiene de la ecuación anteriormente vista:

$$S_T = \alpha y_T + (1-\alpha)S_{T-1}$$

La expresión $S_T^{(2)}$ es la estadística de doble suavizamiento. La constante α se encuentra entre 0 y 1, siendo nuevamente la constante de suavizamiento.

Para β_0 la estimación queda:

$$\begin{aligned} b_0(T) &= 2S_T - S_T^{(2)} - T b_1(T) \\ &= 2S_T - S_T^{(2)} - T \left[\frac{\alpha}{1-\alpha} (S_T - S_T^{(2)}) \right] \end{aligned}$$

Para comenzar con el procedimiento de Doble Suavizamiento Exponencial es necesario tener los valores de S_0 y $S_0^{(2)}$. Dichos valores pueden encontrarse aplicando regresión a cierto número n de observaciones y obteniendo así estimaciones para los coeficientes β_0 y β_1 . Denotando a estas estimaciones iniciales de β_0 y β_1 por $b_0(0)$ y $b_1(0)$ y sabiendo que para cualquier período en el tiempo T

$$b_1(T) = \frac{\alpha}{1-\alpha} (S_T - S_T^{(2)})$$

Entonces para el período $T = 0$ se tiene

$$b_1(0) = \frac{\alpha}{1-\alpha} (S_0 - S_0^{(2)})$$

También para cualquier período T

$$b_0(T) = 2S_T - S_T^{(2)} - T \left[\frac{\alpha}{1-\alpha} (S_T - S_T^{(2)}) \right]$$

Por lo que para $T = 0$

$$b_0(0) = 2S_0 - S_0^{(2)}$$

Así para $T = 0$ las ecuaciones

$$b_1(0) = \frac{\alpha}{1-\alpha} (S_0 - S_0^{(2)}) \quad \text{y} \quad b_0(0) = 2S_0 - S_0^{(2)}$$

pueden resolverse para S_0 y $S_0^{(2)}$ en términos de las estimaciones iniciales

$$S_0 = b_0(0) - \left[\frac{1-\alpha}{\alpha} \right] b_1(0)$$

$$S_0^{(2)} = b_0(0) - 2 \left[\frac{1-\alpha}{\alpha} \right] b_1(0)$$

La solución de estas ecuaciones producirá los valores iniciales de las estadísticas de suavizamiento S_0 y $S_0^{(2)}$, necesarias para iniciar el procedimiento. Si las estimaciones de β_0 y β_1 , no se podrán determinar S_0 y $S_0^{(2)}$. Por lo que para este caso puede asignarse el valor de la primera observación para ambas S_0 y $S_0^{(2)}$.

Dadas S_0 y $S_0^{(2)}$ puede efectuarse el suavizamiento en cada periodo $T + \tau$ usando las ecuaciones:

$$S_T = \alpha Y_T + (1-\alpha)S_{T-1} \quad \text{y} \quad S_T^{(2)} = \alpha S_T + (1-\alpha)S_{T-1}^{(2)}$$

y pueden generarse predicciones para cualquier periodo $T+\tau$ usando la ecuación

$$\hat{Y}_{T+\tau}(T) = \left[2 + \frac{\alpha\tau}{(1-\alpha)} \right] S_T - \left[1 + \frac{\alpha\tau}{(1-\alpha)} \right] S_T^{(2)}$$

Con la simulación de una Serie de Tiempo puede determinarse una α adecuada para utilizarse en el procedimiento.

2.3 MODELOS ARIMA.

2.3.1 CONCEPTOS FUNDAMENTALES.

Este trabajo se enfocará primordialmente a este tipo de análisis desarrollado por Box y Jenkins en su libro "Time Series Analysis, Forecasting and Control" [1] el cual consta de modelos matemáticos que permiten representar Series de Tiempo que contengan en su patrón de comportamiento ciclos, variaciones estacionales, fluctuaciones y tendencias dando así una gran flexibilidad de modelaje para las Series de Tiempo.

El punto de partida en el análisis de Series de Tiempo es el concepto de serie que es una secuencia de observaciones en el tiempo [2]. Llamando Z_1, Z_2, \dots, Z_N a la secuencia de observaciones igualmente espaciadas en el tiempo $1, 2, \dots, N$ tomadas como la ocurrencia de una distribución de probabilidad $P_{1..N}(Z_1, Z_2, \dots, Z_N)$. Por lo que el objetivo se traduce en utilizar la distribución conjunta con la que se establecerán predicciones para futuras observaciones, por lo que se tendrá una función de distribución condicional para Z_{T+1} (la observación futura).

$$P_{T+1/1..T}(Z_{T+1}/Z_1, Z_2, \dots, Z_T)$$

Ahora, se construirán predicciones e intervalos de confianza para estas observaciones con ayuda del primer y segundo momentos de la distribución condicional, con el fin ésto último, de fundamentar el procedimiento.

Lo que se busca con este método es inferir de datos anteriores, tratando de encontrar el patrón que siguen, permitiendo con ésto ver la evolución que dichos datos han tenido a través del tiempo y utilizar así este conocimiento para finalmente predecir las ocurrencias futuras.

Es necesario analizar algunos conceptos que deberán contemplarse dentro de la serie en cuestión.

Estacionaridad.

Este concepto es fundamental en una distribución conjunta, ésto quiere decir que la serie deberá ser invariante en su desplazamiento por el tiempo, es decir,

$$P(Z_t, \dots, Z_{t+k}) = P(Z_{t+m}, \dots, Z_{t+m+k})$$

donde $t \in$ tiempo
 $k, m \in \mathbb{Z}$

a este concepto se le conoce como estacionaridad estricta, sin embargo existe otro tipo de estacionaridad que es aquella en la que sólo se requiere que los momentos hacia cierto orden sean invariantes en el tiempo.

Si $k \neq 0$ se tiene que

$$P(Z_t) = P(Z_{t+m})$$

donde $m = t_1, t_2 \dots$

ésto es la función de distribución marginal será la misma para ambas. De donde sus esperanzas son iguales

$$E(Z_t) = E(Z_{t+m})$$

y sus varianzas son también iguales

$$V(Z_t) = V(Z_{t+m}).$$

Así, para cualquier par de observaciones separadas por cierto número de periodos j

$$P(Z_t, Z_{t+j}) = P(Z_{t+m}, Z_{t+m+j})$$

de donde la covarianza de cualquier par de observaciones depende tan sólo de j .

$$C(Z_t, Z_{t+j}) = C(Z_{t+m}, Z_{t+m+j}) = \Gamma_j$$

Llamando a Γ_j la autocovarianza en el periodo j , de donde el término "auto" se utiliza para saber que ambas observaciones se refieren a la misma serie.

Notese que la estacionaridad refleja varias implicaciones generales sobre el comportamiento de la serie. Se sabe que la esperanza de Z_t que no depende de t aunque es una constante sirve para situar la serie, es decir, aunque la serie se encuentre fluctuando siempre lo hará cerca del valor de esa media, ésto, teniendo en cuenta las observaciones pasadas de la serie. Si se tiene una distribución Normal, el contar con estacionaridad podrá ofrecer un intervalo de confianza en el que seguramente (con algún porcentaje de confianza) se encontrará la siguiente observación.

Autocorrelación.

Una implicación de estacionaridad que tiene una interpretación importante en el comportamiento de una Serie de Tiempo se deriva de que la autocovarianza entre dos observaciones depende sólo del número de periodos que las separa, por lo que

$$\Gamma_j = C(Z_t, Z_{t+j}) = E[(Z_t - E Z_t)(Z_{t+j} - E Z_{t+j})] = E[(Z_t - \mu)(Z_{t+j} - \mu)]$$

De lo anterior se puede decir que existirá una autocovarianza positiva o negativa (positiva si una

observación por arriba de la media es seguida por otra igual j periodos después, negativa en caso contrario).

Esto nos dice que la estacionariedad sugiere un comportamiento similar en el patrón de la serie sin importar el periodo en el que se observe.

Las autocovarianzas dependiendo de la serie tendrán valores muy distintos por lo que con el fin de evitar esto y tener un punto de referencia se suelen estandarizar dividiendo cada una de ellas entre Γ_0 obteniendo así las autocorrelaciones. Si se denota entonces por ρ_j la correlación entre Z_t y Z_{t+j} , la función de autocorrelación será:

$$\rho_0 = \Gamma_0 / \Gamma_0 = 1$$

$$\rho_1 = \Gamma_1 / \Gamma_0 \dots$$

La gráfica que las contiene es llamada correlograma.

Estimación de la Función de Autocorrelación para una Muestra de Datos.

Una estimación natural de Γ_j , el producto esperado, será la diferencia entre Z_t y Z_{t+j} con la media de la serie, que será precisamente el producto promedio de la diferencia entre Z_t y Z_{t+j} con la media muestral. Denotando la estimación de Γ_j por C_j se tendrá:

$$C_j = 1/T \sum_{t=1}^{T-j} (Z_t - \bar{Z})(Z_{t+j} - \bar{Z})$$

$$\text{con } j = 1, 2, \dots$$

y donde

$$\bar{Z} = 1/T \sum_{t=1}^T Z_t$$

Por lo que $r_j = C_j / C_0$ con $j = 1, 2, \dots$. A una gráfica de r_j se le llama correlograma muestral, éste será más adelante de gran ayuda para escoger el modelo que sugieran los datos.

Análisis de Series de Tiempo no Estacionarias.

Existen series en la práctica que no muestran un comportamiento estacionario, pero que sin embargo muestran una tendencia determinada (creciente o decreciente). Para éstos casos pueden utilizarse las diferencias de los datos $(Z_t - Z_{t-1})$ que harán que la serie refleje estacionariedad para muchos de los casos; este método se verá más adelante.

2.3.2 MODELOS PARA SERIES DE TIEMPO ESTACIONARIAS.

En la práctica los parámetros de la distribución deben inferirse de las observaciones que se tienen disponibles, por lo que se requiere una estructura tal que la inferencia estadística se confine sólo a algunos parámetros fundamentales. Esto acompañado de la suposición de que los datos se generaron por un proceso estocástico lineal de forma autoregresiva, de promedios móviles o una mezcla de ambos (mostrados a continuación). Su flexibilidad hace de ellos una muy utilizada clase de modelos para muchos tipos de Series de Tiempo encontradas en la práctica.

Procesos Estocásticos Lineales Discretos.

Un proceso estocástico es un proceso lineal discreto si cada observación Z_t puede expresarse en la forma

$$Z_t = \mu + U_t + \phi_1 U_{t-1} + \phi_2 U_{t-2} \quad (1)$$

donde μ y ϕ_i son fijas y la Serie de Tiempo $(\dots U_{t-1}, U_t, \dots)$ es una secuencia de errores aleatorios independientes e idénticamente distribuidos con media cero y varianzas σ_u^2 o llamado también "ruido blanco". El proceso (1) es discreto debido a que las observaciones Z_t son tomadas en intervalos discretos e igualmente espaciados, y lineal porque las Z_t 's son una combinación lineal de los errores pasados y actuales.

Proceso de Promedios Móviles.

Aunque cualquier proceso lineal puede escribirse como (1) es de interés para propósitos prácticos expresar a los pesos ϕ_i en términos de un cierto número de parámetros fundamentales. Uno de ellos puede ser el de

$$\phi_1 = 0 \quad \text{para } i > q$$

esto es

$$Z_t = \mu + U_t + \phi_1 U_{t-1} + \phi_2 U_{t-2} + \dots + \phi_q U_{t-q} \quad (2)$$

el cual se llama un proceso de promedios móviles de orden q debido a que las observaciones son un promedio móvil sobre los errores acumulados hacia atrás durante q periodos. Con el fin de distinguir el proceso de promedios móviles del proceso lineal en general, se escribirá como:

$$Z_t = \mu + U_t - \theta_1 U_{t-1} - \dots - \theta_q U_{t-q}$$

el signo (-) se introdujo por convención. Esto es necesario tenerlo muy en cuenta.

Una forma de abreviar un proceso de promedios móviles de orden q es $MA(q)$, notese que el proceso está descrito por $q+2$ parámetros $\sigma_u^2, \mu, \theta_1, \dots, \theta_q$. Este proceso será estacionario mientras la condición de que

$$\sum_{i=0}^{\infty} \phi_i \quad \text{o bien} \quad 1 - \sum_{i=0}^{\infty} \theta_i$$

sea finita (Esto se cumplirá debido a que existe un número finito en la suma).

El Proceso de Promedios Móviles de 1^{er} Orden.

Un proceso importante es el $MA(1)$ o bien

$$Z_t = \mu + U_t - \theta_1 U_{t-1}$$

donde la media del proceso es

$$E(Z_t) = \mu + E(U_t) - \theta_1 E(U_{t-1}) = \mu$$

y la varianza

$$\begin{aligned} \Gamma_0 &= E(U_t - \theta_1 U_{t-1})^2 \\ &= E(U_t^2 - 2\theta_1 U_t U_{t-1} + \theta_1^2 U_{t-1}^2) \\ &= \sigma_u^2 + \theta_1^2 \sigma_u^2 \\ &= \sigma_u^2 (1 + \theta_1^2) \end{aligned}$$

la autocovarianza del primer periodo

$$\begin{aligned} \Gamma_1 &= E[(U_t - \theta_1 U_{t-1})(U_{t-1} - \theta_1 U_{t-2})] \\ &= \sigma_u^2 (-\theta_1) \end{aligned}$$

y autocovarianzas para periodos mayores

$$\begin{aligned} \Gamma_j &= E[(U_t - \theta_1 U_{t-1})(U_{t-j} - \theta_1 U_{t-j-1})] \\ &= 0 \quad \text{para } j > 1. \end{aligned}$$

Y la función de autocorrelación es

$$\rho_1 = -\theta_1 / (1 + \theta_1^2) = \Gamma_1 / \Gamma_0$$

$$\rho_j = 0 \quad \text{con } j > 1$$

Una propiedad importante del $MA(1)$ es vista en la función de autocorrelación que sólo guarda un periodo de "memoria", en otras palabras la observación Z_t está

correlacionada con su predicción Z_{t-1} y su sucesor Z_{t+1} , pero no con ningún otro miembro de la serie.

Para este proceso el correlograma correspondiente presentará un pico en el primer período (ρ_1).

Proceso de Promedios Móviles de Alto Orden.

Los momentos para cualquier MA(q) son:

La esperanza $E(Z_t) = \mu$

y la varianza $\Gamma_0 = \sigma_u^2 \sum_{i=0}^q \theta_i^2$ donde $\theta_0 = 1$

las autocovarianzas

$$\Gamma_j = \begin{cases} \sigma_u^2(-\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q) & j=1..q \\ 0 & j > q \end{cases}$$

y la función de autocorrelación

$$\rho_j = \begin{cases} -\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q & j=1..q \\ 0 & j > q \end{cases}$$

así el correlograma consiste en q picos del período 1..q y ningún pico después, la memoria de MA(q) será de q períodos, esto es, Z_t no tendrá errores comunes con miembros de la serie separados por más de q períodos.

Proceso Autogresivo.

Recordando la ecuación

$$Z_t = \mu + U_t + \phi_1 U_{t-1} + \phi_2 U_{t-2} + \dots$$

donde se expresa Z_t en términos del error actual y los pasados. Otra forma alternativa de expresar Z_t es en términos del error actual y de las observaciones pasadas. Despejando entonces U_t queda:

$$U_t = Z_t - \mu - \phi_1 U_{t-1} - \phi_2 U_{t-2} - \dots \quad (4)$$

además para t+1

$$U_{t+1} = Z_{t+1} - \mu - \phi_1 U_{t+1} - \phi_2 U_{t+1} - \dots \quad (5)$$

sustituyendo (5) en (4) para eliminar U_{t+1} queda:

$$Z_t = \mu(1-\phi_1) + \phi_1 Z_{t-1} + U_t + (\phi_2 - \phi_1^2) U_{t-2} + (\phi_3 + \phi_1 \phi_2) U_{t-3} + \dots \quad (6)$$

pueden irse eliminando así sucesivamente de (6) hasta llegar a:

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \dots + \delta + U_t$$

donde los pesos π_i en observaciones pasadas son funciones de los pesos ϕ_i ; δ es una constante la cual es también una función de μ y ϕ_1 .

Si $\pi_1 = 0$ para $i > p$ entonces tenemos

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \dots + \pi_p Z_{t-p} + \delta + U_t$$

la cual es llamada proceso autoregresivo de orden p o bien AR(p), igualmente para distinguirlo del modelo general se denotará:

$$Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_p Z_{t-p} + \delta + U_t \quad (7)$$

El llamarlo autoregresivo viene del hecho de que la ecuación (7) es esencialmente de regresión en la cual Z_t es obtenido por sus propios valores anteriores en vez de obtenerse de un conjunto de variables independientes.

Modelo Autoregresivo de Primer Orden.

El modelo más simple y de gran utilidad en la práctica es el AR(1) el cual se denota como:

$$Z_t = \theta_1 Z_{t-1} + \delta + U_t \quad (8)$$

que sólo depende de la observación previa.

Para saber si es estacionario puede escribirse a (8) en términos de errores pasados, para lo cual bastará sustituir Z_{t-1} , Z_{t-2} y así sucesivamente lo que resultará:

$$Z_t = \delta / (1 - \theta_1) + U_t + \theta_1 U_{t-1} + \theta_1^2 U_{t-2} + \dots$$

Por lo que podría verse como un proceso de promedios móviles de orden infinito en donde $\theta_1 = \theta$ y $\delta / (1 - \theta_1) = \mu$ por lo que de los resultados anteriores se requerirá que $|\theta_1| < 1$ y su media por lo tanto es de:

$$E(Z_t) = \delta / (1 - \theta_1)$$

y la varianza y autocovarianzas están dadas por

$$\Gamma_j = \theta_1^j [\sigma_u^2 / (1 - \theta_1^2)] \quad (9)$$

La función de autocorrelación para el AR(1) de la ecuación (9) es simplemente

$$\rho_j = \theta_1^j$$

indicando que la correlación entre observaciones declina exponencialmente dependiendo del número de periodos que se encuentran separados.

El correlograma entonces está dado por Γ_j (ecuación 9).

Otra forma de calcular los momentos del AR(1), que será muy útil para calcular la de los procesos autoregresivos de orden mayor es:
Para la esperanza

$$\begin{aligned} E(Z_t) &= E(\theta_1 Z_{t-1} + \delta + U_t) \\ &= \theta_1 E(Z_{t-1}) + \delta + 0 \end{aligned}$$

Si el proceso es estacionario $E(Z_{t-1}) = E(Z_t)$ y entonces

$$E(Z_t) = \delta / (1 - \theta_1)$$

que es el resultado obtenido anteriormente.

La varianza se muestra mas fácilmente si se denota a Z_t como la desviación del proceso a la media, esto es,

$$\tilde{Z}_t = Z_t - \delta / (1 - \theta_1)$$

lo que implica que $Z_t = \theta_1 Z_{t-1} + U_t$.

$$\text{La media } (E\tilde{Z}_t) = 0$$

y la varianza es:

$$\begin{aligned} \Gamma_0 &= E(\tilde{Z}_t^2) \\ &= E[\tilde{Z}_t(\theta_1 \tilde{Z}_{t-1} + U_t)] \\ &= \theta_1 E(\tilde{Z}_t \tilde{Z}_{t-1}) + E(\tilde{Z}_t U_t) \\ &= \theta_1 \Gamma_1 + E[(\theta_1 \tilde{Z}_{t-1} + U_t) U_t] \\ &= \theta_1 \Gamma_1 + \theta_1 E(\tilde{Z}_{t-1} U_t) + E(U_t^2) \\ &= \theta_1 \Gamma_1 + \sigma_u^2 \end{aligned} \tag{10}$$

el cual involucra la autocovarianza del primer periodo y la varianza de los errores.

Notese que $E(\tilde{Z}_{t-1} U_t) = 0$ debido a que \tilde{Z}_{t-1} involucra sólo errores que ocurrieron antes del t-ésimo periodo, en el que todos son independientes de U_t .

Ahora

$$\begin{aligned} \Gamma_0 &= E(\tilde{Z}_t \tilde{Z}_{t-1}) \\ &= \theta_1 E(\tilde{Z}_{t-1}^2) + E(U_t \tilde{Z}_{t-1}) \\ &= \theta_1 \Gamma_0 \end{aligned} \tag{11}$$

resolviendo (10) y (11) para Γ_0 y Γ_1 y sustituyendo (11) en (10)

$$\begin{aligned}\Gamma_0 &= \theta_1^2 \Gamma_0 + \sigma_u^2 \\ &= \sigma_u^2 / (1 - \theta_1^2) \quad y \\ \Gamma_1 &= \theta_1 (\sigma_u^2 / (1 - \theta_1^2))\end{aligned}$$

La autocovarianza para los periodos 2, 3, etc. se obtiene recursivamente para la relación

$$\begin{aligned}\Gamma_j &= E(\tilde{Z}_t \tilde{Z}_{t-j}) \\ &= \theta_1 E(\tilde{Z}_{t-1} \tilde{Z}_{t-j}) + E(U_t \tilde{Z}_{t-j}) \\ &= \theta_1 \Gamma_{j-1} \quad j = 1, 2, 3, \dots\end{aligned}$$

que implica $\Gamma_j = \theta_1^j \Gamma_0 \quad j = 1, 2, 3, \dots$

$$\begin{aligned}\theta_j &= \theta_1 \theta_{j-1} \\ &= \theta_1^j \quad \text{derivadas del primer método.}\end{aligned}$$

Finalmente el que $|\theta_1| < 1$ tenga sentido puede verse intuitivamente ya que suponiendo que $\theta_1 = 1$ se tendría una caminata aleatoria no estacionaria. Si ahora $\theta_1 > 1$, Z_t sería no sólo no estacionaria sino crecería explosivamente. Esto sugiere que existe algo especial sobre el valor de 1 para θ_1 . Es decir, que podría sugerirse el procedimiento de una diferencia entre observaciones seguidas para convertir a la caminata aleatoria de no estacionaria en estacionaria.

Proceso Autoregresivo de Orden Mayor.

Los resultados obtenidos para el proceso de orden 1 serán utilizados para procesos en donde $p > 1$. Un proceso $AR(p)$ puede escribirse en términos de los errores solamente, y entonces la estacionaridad se tiene sólo si la suma infinita de los coeficientes ϕ_i converge.

La media de un proceso $AR(p)$ es:

$$\begin{aligned}E(Z_t) &= \theta_1 E(Z_{t-1}) + \dots + \theta_p E(Z_{t-p}) + \delta + E(U_t) \\ &= \frac{\delta}{1 - \theta_1 - \dots - \theta_p}\end{aligned}$$

Las ecuaciones

$$\begin{aligned}\Gamma_0 &= \theta_1 \Gamma_1 + \dots + \theta_p \Gamma_p + \sigma_u^2 \\ \Gamma_1 &= \theta_1 \Gamma_0 + \dots + \theta_p \Gamma_{p-1} \\ \dots &\dots \dots \dots \dots \dots \dots \\ \Gamma_p &= \theta_1 \Gamma_{p-1} + \dots + \theta_p \Gamma_0\end{aligned} \quad (1)$$

dados los parámetros $\theta_1, \dots, \theta_p$ y σ_u^2 pueden ser resueltas ($p+1$ ecuaciones lineales) para los $p + 1$ parámetros

$\Gamma_0, \dots, \Gamma_p$ desconocidos. Para periodos mayores a p , la covarianza Γ_j puede calcularse recursivamente de

$$\Gamma_j = \theta_1 \Gamma_{j-1} + \dots + \theta_p \Gamma_{j-p} \quad j > p.$$

La función de autocorrelación puede calcularse directamente dividiendo (1) entre Γ_0

$$\begin{aligned} \rho_1 &= \theta_1 + \theta_2 \rho_1 + \dots + \theta_p \rho_{p-1} \\ \dots & \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \rho_p &= \theta_1 \rho_{p-1} + \theta_2 \rho_{p-2} + \dots + \theta_p \end{aligned}$$

un sistema de mp ecuaciones con p $\rho_1 \dots \rho_p$ desconocidas. Los coeficientes de autocorrelación para periodos mayores pueden calcularse recursivamente para

$$\rho_j = \theta_1 \rho_{j-1} + \dots + \theta_p \rho_{j-p} \quad j > p.$$

Proceso Autoregresivo de Promedios Móviles.

Una extensión natural de los procesos AR y MA sería una nueva clase de modelos de la forma:

$$Z_t = \theta_1 Z_{t-1} + \dots + \theta_p Z_{t-p} + \delta + U_t - \theta_1 U_{t-1} - \dots - \theta_q U_{t-q}$$

al que se le llama un proceso autoregresivo de promedios móviles de grado p y q respectivamente o ARMA(p, q).

Para muchas de las series en la práctica, el tener esta combinación proporciona un modelo más satisfactorio que el en los casos anteriores.

El Proceso ARMA(1,1).

Este proceso es de la forma

$$Z_t = \theta_1 Z_{t-1} + \delta + U_t - \theta_1 U_{t-1}$$

aunque pudiera escribirse como un proceso MA de orden infinito

$$Z_t = \delta / (1 - \theta_1) + U_t + (\theta_1 - \theta_1) U_{t-1} + \theta_1 (\theta_1 - \theta_1) U_{t-2} + \theta_1^2 (\theta_1 - \theta_1) U_{t-3} + \dots$$

puede aproximarse con un proceso MA de orden finito quitando términos en el momento en que los coeficientes $\theta_1^i (\theta_1 - \theta_1)$ sean "arbitrariamente chicos". De igual forma el proceso ARMA(1,1) puede escribirse en forma autoregresiva, la cual también podría truncarse cuando $\pi_1 = \theta_1 / (1 - \theta_1)$ sean lo "suficientemente pequeños". Así, como pueda verse para cada uno de éstos modelos, se tendrían un número grande de parámetros, hasta que los coeficientes parezcan lo suficientemente pequeños, y en cambio con un modelo ARMA(1,1) el número de coeficientes será mucho menor.

La media del proceso ARMA(1,1) esta dada por:

$$\begin{aligned} E(Z_t) &= \theta_1 E(Z_{t-1}) + \delta + E(U_t) - \theta_1 E(U_{t-1}) \\ &= \frac{\delta}{1-\theta_1} \end{aligned}$$

el mismo resultado que se obtuvo para un AR(1).
Y la varianza

$$\begin{aligned} \Gamma_0 &= E(Z_t^2) \\ &= \theta_1 E(Z_t Z_{t-1}) + E(Z_t U_t) - \theta_1 E(Z_t U_{t-1}) \end{aligned}$$

Ahora la $E(U_t U_{t-1})$ es la covarianza entre Z_t y el error previo, y

$$\begin{aligned} E(\bar{Z}_t U_{t-1}) &= \theta_1 E(\bar{Z}_{t-1} U_{t-1}) + E(U_t U_{t-1}) - \theta_1 E(U_{t-1}^2) \\ &= \theta_1 \sigma_u^2 - \theta_1 \sigma_u^2 \end{aligned}$$

Entonces

$$\Gamma_0 = \theta_1 \Gamma_1 + \sigma_u^2 - \theta_1 (\theta_1 - \theta_1) \sigma_u^2 \quad (1)$$

Para

$$\begin{aligned} \Gamma_1 &= E(\bar{Z}_t \bar{Z}_{t-1}) \\ &= \theta_1 E(\bar{Z}_{t-1}^2) + E(U_t \bar{Z}_{t-1}) - \theta_1 E(U_{t-1} \bar{Z}_{t-1}) \\ &= \theta_1 \Gamma_0 - \theta_1 \sigma_u^2 \end{aligned} \quad (2)$$

$$\begin{aligned} \Gamma_2 &= E(\bar{Z}_t \bar{Z}_{t-2}) \\ &= \theta_1 E(\bar{Z}_{t-1} \bar{Z}_{t-2}) + E(U_t \bar{Z}_{t-2}) - \theta_1 E(U_{t-1} \bar{Z}_{t-2}) \\ &= \theta_1 E(\bar{Z}_{t-1} \bar{Z}_{t-2}) \\ &= \theta_1 \Gamma_1 \end{aligned}$$

Finalmente para

$$\Gamma_j = \theta_1 \Gamma_{j-1} \quad j = 2, 3, \dots$$

Para calcular las autocovarianzas dados los parámetros del proceso se resuelven (1) y (2) resultando

$$\Gamma_0 = \frac{(1+\theta_1^2 - 2\theta_1\theta_1)}{1-\theta_1^2} \quad \text{y} \quad \Gamma_1 = \frac{(1-\theta_1\theta_1)(\theta_1-\theta_1)}{1-\theta_1^2}$$

el resto se calcula recursivamente de (3).

La autocorrelación al período 1 es :

$$\rho_1 = \frac{r_1}{r_0} = \frac{(1-\theta_1\theta_2)(\theta_1-\theta_2)}{1+\theta_1^2-2\theta_1\theta_2}$$

y dividiendo entre (3) obtenemos para los siguientes períodos $\rho_j = \theta_1 \rho_{j-1}$ para $j = 2, 3, \dots$

En resumen, la función de autocorrelación combina características tanto del proceso AR(1) como del MA(1).

Para poder analizar las funciones de autocorrelación, se tienen los correlogramas, en los que puede observarse si ésta es, positiva o negativa dentro del modelo.

Procesos Autoregresivos de Promedios Móviles de Orden Mayor.

Extendiendo ahora los procesos ARMA a procesos de orden mayor, éstos pueden escribirse como un proceso MA de orden infinito. Así la condición de estacionaridad se cumple cuando

$$\sum_{i=0}^{\infty} \phi_i$$

converga. Similarmente si el proceso se escribe como un AR de orden infinito, la condición para que se tenga estacionaridad será la de

$$\sum_{i=0}^{\infty} \pi_i$$

La media de este proceso estará dada por

$$E(Z_t) = \theta_1 E(Z_{t-1}) + \dots + \theta_p E(Z_{t-p}) + \delta + E(U_t) - \dots - \theta_q E(U_{t-q})$$

$$= \frac{\delta}{1 - \theta_1 - \dots - \theta_p}$$

Las covarianzas son evaluadas por un conjunto de ecuaciones de la forma:

$$r_j = E(\bar{Z}_t \bar{Z}_{t-j})$$

$$= \theta_1 E(\bar{Z}_{t-1} \bar{Z}_{t-j}) + \dots + \theta_p E(\bar{Z}_{t-p} \bar{Z}_{t-j}) + E(U_t \bar{Z}_{t-j}) - \theta_1 E(U_{t-1} \bar{Z}_{t-j}) - \dots - \theta_q E(U_{t-q} \bar{Z}_{t-j})$$

Ahora, si $j < q$ entonces los términos que contienen Z_{t-j} y a los errores pasados U , serán distintos de cero; mientras que \bar{Z}_{t-j} estará correlacionado con los errores ocurridos

hacia el periodo $t-j$. Entonces la ecuación para $\Gamma_0, \Gamma_1, \dots, \Gamma_q$ contemplará los parámetros del promedio móvil $\theta_1, \theta_2, \dots, \theta_q$. Sin embargo para $j > q$, se tiene:

$$\Gamma_j = \theta_1 \Gamma_{j-1} + \dots + \theta_q \Gamma_{j-q} \quad j > q$$

porque entonces los términos que contienen a Z_{t-j} y los errores son cero. Consecuentemente, el ρ_j para periodos lejanos están dados por:

$$\rho_j = \theta_1 \rho_{j-1} + \dots + \theta_q \rho_{j-q} \quad j > q$$

En resumen, entonces la autocorrelación para periodos $1 \dots q$ será afectada por la parte de promedios móviles del proceso.

Estos modelos mixtos constituyen una muy poderosa clase de procesos para Series de Tiempo Estacionarias que se encuentran con mayor frecuencia en la práctica, debido a su gran flexibilidad para explicar el comportamiento de la serie.

2.3.3 MODELOS PARA SERIES DE TIEMPO NO ESTACIONARIAS.

La estacionaridad es una condición muy fuerte en las Series de Tiempo aunque casi no se encuentra en la práctica. Así pueden tenerse diversos conceptos de estacionaridad, uno de ellos puede ser el de la "no estacionaridad homogénea" en el sentido de que la serie se mueve libremente pero con un cierto patrón, con su comportamiento para ciertos períodos en el tiempo esencialmente igual, se puede decir entonces que la no "estacionaridad homogénea" se presentan en series que muestran cambios sucesivos (o diferencias estacionarias). Al obtener las diferencias de la serie, los modelos ARMA darán una clase altamente flexible para trabajar con las Series de Tiempo.

Diferenciación y no Estacionaridad Homogénea.

Considerando el proceso autoregresivo

$$Z_t = \theta_1 Z_{t-1} + U_t$$

y analizando los posibles valores para θ_1 , puede verse que si $|\theta_1| < 1$ el proceso es estacionario; si $|\theta_1| > 1$ el modelo crecerá sin cota; y de ser $\theta_1 = 1$ se tendrá una caminata aleatoria. Para este último caso la distribución de los cambios (o diferencias) en el proceso no cambia, es decir, se comporta homogéneamente. Por esto es que las diferencias de la Serie de Tiempo son estacionarias:

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

y la distribución de U_t es fija.

Una generalización natural de la caminata aleatoria, es considerar ahora a los procesos ARMA estacionarios, como un mecanismo potencial modelador para las series de tiempo no estacionarias. Lo anterior es posible si se define W_t como la diferencia de dos observaciones; esto es,

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

con lo que el modelo general quedaría:

$$W_t = \theta_1 W_{t-1} + \dots + \theta_p W_{t-p} + U_t - \theta_1 U_{t-1} - \dots - \theta_q U_{t-q}$$

sustituyendo W_t por $(Z_t - Z_{t-1})$ la serie observada Z_t estaría dada por

$$Z_t = Z_{t-1} + \theta_1 (Z_{t-1} - Z_{t-2}) + \dots + \theta_p (Z_{t-p} - Z_{t-p-1}) + U_t - \theta_1 U_{t-1} - \dots - \theta_q U_{t-q}$$

además

$$Z_t = W_t + W_{t-1} + W_{t-2} + \dots$$

Como Z_t es una integración de la serie W_t , se le llama al proceso Proceso Integrado Autoregresivo de Promedios Móviles, o bien, ARIMA.

Cuando el proceso no se convierte en estacionario al obtener las primeras diferencias, pueden hacerse unas nuevas, que serán las diferencias de las primeras diferencias, es decir, si se llama a y_t las segundas diferencias de la serie, entonces:

$$y_t = W_t - W_{t-1} = (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2}) = Z_t - 2Z_{t-1} + Z_{t-2}$$

pudiendo así ejecutar un cierto número de diferencias si la serie lo requiere, aunque para la mayoría de los casos con esto será suficiente.

Finalmente, si se denota con d el orden de diferencia de un proceso, se dirá que un modelo Integrado Autoregresivo de Promedios Móviles tendrá dimensiones p, d y q ; y se denotará ARIMA(p, d, q).

Es necesario hacer notar que en general, el Análisis de Regresión y el Suavizamiento Exponencial tienen una fuerte suposición, ésta es que los componentes de error $\epsilon_1, \epsilon_2, \dots$ no están relacionados unos con otros. Sin embargo si los términos de error están relacionados, por ejemplo, cada 7 u 8 periodos, entonces las técnicas de Análisis de Regresión y Suavizamiento Exponencial podrán producir una predicción aceptablemente exacta. Pero si los términos de error están muy relacionados una predicción más exacta podría ser dada por una técnica que utilizará precisamente esa relación, como es la propuesta por Box y Jenkins debido a que y_t es expresado como una función de los componentes de error $\epsilon_1, \epsilon_2, \dots$.

Así, aunque se centrará este trabajo en el análisis de modelos ARIMA propuesto por Box y Jenkins, es necesario tener en cuenta que existen estas 2 alternativas como Análisis de Series de Tiempo y que pueden utilizarse en caso de no tener la posibilidad de efectuar el modelado de Box y Jenkins con la precaución de efectuar una prueba de correlación entre los errores para evitar problemas con la predicción.

2.3.4 MODELOS DE SERIES DE TIEMPO ESTACIONALES.

En la práctica la estacionalidad es un fenómeno muy común, siendo la repetición de cierto patrón o cierto comportamiento sobre determinados periodos de tiempo (anuales generalmente). Las observaciones de la serie tienen una fuerte correlación serial de acuerdo a dichos periodos de tiempo.

Proceso Estacional de Promedios Móviles.

Considerese un proceso de Promedios Móviles con las siguientes especificaciones:

- S es el número de observaciones por periodo estacional, por lo que el orden de promedios móviles es un entero múltiplo de s
- Los únicos coeficientes diferentes de cero son aquellos con subíndice entero múltiplo de s.

Sería entonces :

$$Z_t = U_t - \theta_1 U_{t-s} - \dots - \theta_{Qs} U_{t-Qs}$$

donde Q es el múltiplo mayor de s, de donde el orden del proceso es Qs . Cambiando la notación $\theta_j = \theta_{js}$

$$Z_t = U_t - \theta_1 U_{t-s} - \dots - \theta_Q U_{t-Qs}$$

al que se le llama proceso SMA.

Así la función de autocorrelación será diferente de cero sólo en $s, 2s, \dots, Qs$ en particular

$$\rho_s = \frac{-\theta_1 + \theta_1 \theta_2 + \dots + \theta_{Q-1} \theta_Q}{1 + \theta_1^2 + \dots + \theta_Q^2}$$

$$\rho_{2s} = \frac{-\theta_Q}{1 + \theta_1^2 + \dots + \theta_Q^2}$$

correspondiendo a los picos en $s, 2s, \dots, Qs$ de la correlación solamente.

Es importante notar que la función de autocorrelación implica que una observación dada está correlacionada sólo con la observación siguiente y precedente por s periodos.

Proceso Autoregresivo Estacional.

De igual forma

$$Z_t = \theta_1 Z_{t-s} + \dots + \theta_{Ps} Z_{t-Ps} + U_t$$

donde P es el mayor múltiplo de s presente en el modelo. De nuevo $\theta_{j_s} = \alpha_j$

$$Z_t = \alpha_1 Z_{t-s} + \dots + \alpha_P Z_{t-Ps} + U_t$$

llamado un proceso SAR de orden P .

La autocovarianza distinta de cero sólo se da para períodos múltiplos de s

$$\Gamma_0 = \frac{\sigma_U^2}{1 - \alpha_1^2}$$

$$\Gamma_s = \alpha_1 \Gamma_0$$

$$\Gamma_{2s} = \alpha_1^2 \Gamma_0$$

.....

$$\Gamma_{js} = \alpha_1^j \Gamma_0$$

y las autocorrelaciones son

$$\rho_{js} = \alpha_1^j$$

$$j = 0, 1, 2, 3, \dots$$

Es importante notar las diferencias y semejanzas que existen entre un proceso de promedios móviles estacional y uno autoregresivo estacional. Primero, la correlación en los períodos estacionales persiste indefinidamente, aunque va declinando en el caso del autoregresivo y se corta en cierto período estacional para uno de promedios móviles. Por otro lado, su similitud consiste en que las series de $1, s+1, 2s+1$ y $2, s+2, 2s+2$ etc. son independientes entre ellas.

Proceso Autoregresivo Integrado de Promedios Móviles Estacional.

La natural generalización de los procesos autoregresivo y de promedios móviles estacionales es por supuesto el modelo mixto que tiene la forma:

$$Z_t = \alpha_1 Z_{t-s} + \dots + \alpha_P Z_{t-Ps} + U_t - \beta_1 U_{t-s} - \dots - \beta_Q U_{t-Qs}$$

Puede considerarse la posibilidad de que la serie sea no estacionaria y que las diferencias estacionarias de la serie generen un proceso estacionario. Denotando las diferencias estacionarias de Z_t con x_t

$$x_t = Z_t - Z_{t-s}$$

y el modelo integrado es

$$x_t = \alpha_1 x_{t-s} + \dots + \alpha_p x_{t-p} + u_t - \alpha_1 u_{t-s} - \dots - \alpha_q u_{t-q}$$

La estructura de la autocorrelación del proceso ARMA estacional es análoga a la del proceso ARMA no estacional con correlación distinta de cero que ocurre sólo en los períodos $s, 2s, \text{etc.}$

Es útil para propósitos futuros el utilizar el operador hacia atrás B definido como

$$Bz_t = z_{t-1} \quad B^s z_t = z_{t-s}$$

para reescribir la ecuación.

$$(1 - \alpha_1 B^s - \dots - \alpha_p B^p)(1 - B^s)^d z_t = (1 - \alpha_1 B^s - \dots - \alpha_q B^q) u_t$$

Más general, para modelos estacionarios en la D th diferencia estacional, el modelo se convierte en:

$$(1 - \alpha_1 B^s - \dots - \alpha_p B^p)(1 - B^s)^D z_t = (1 - \alpha_1 B^s - \dots - \alpha_q B^q) u_t$$

III PARAMETROS NECESARIOS PARA LA IDENTIFICACION DE UN MODELO.

3.1 PROBLEMAS EN LA SELECCION DE UN MODELO DE PREDICCIÓN.

Sea cual fuera la Serie de Tiempo que se va a utilizar para generar las predicciones, será necesario desarrollar un modelo que "explique" el movimiento de dicha serie, este modelo será utilizado en base a los datos anteriores para obtener predicciones de la variable de interés en el futuro.

Estos modelos serán ecuaciones que contienen coeficientes fijos estimados, que representan una estructura estocástica que se supone no cambiará en el tiempo.

Una base para generar predicciones futuras, de una variable dependiente del tiempo, será un modelo matemático que incorpora componentes determinísticos (variaciones cíclicas, tendencias, fluctuaciones, etc.) y estocásticos del presente comportamiento de la serie.

Box y Jenkins han desarrollado sus modelos autoregresivo y de promedios móviles, mezclando también formas estacionales y no estacionales de comportamiento.

Una importante guía en el Análisis de Series de Tiempo, es el cómo seleccionar la forma del modelo introduciendo el mínimo número de parámetros que sea posible. Box y Jenkins han propuesto un proceso iterativo de 3 pasos para determinar el modelo ideal para una Serie de Tiempo:

- Identificación.
- Estimación.
- Diagnóstico del modelo.

Revisión de los datos.

Previo al proceso iterativo, es necesario y muy importante echar un vistazo a los datos para evitar al máximo la posibilidad de errores en dichos datos, debidos en primer lugar a como fueron recolectados, evitando con esto que factores externos afecten a la predicción. Un ejemplo claro de este problema sería que cierto fenómeno fuera medido en cada observación con un diferente aparato, quedando afectada la predicción con la variación de aparato a aparato. Otro importante factor es el que no se tenga la totalidad de los datos, esto aunque no es fatal, podría no dejar ver la existencia de alguna fluctuación o algún cambio de tendencia, etc. Otro problema más podría ser que la serie tuviera observaciones aberrantes, que bien podrían provenir de una observación errónea o que significarían un posible punto de atención que sería necesario analizar con el debido cuidado.

Este punto aunque no pertenece al proceso mencionado es muy importante tomarlo en cuenta.

Identificación.

En la fase de identificación, son utilizadas para seleccionar un modelo tentativo ciertas estadísticas apropiadas para la identificación de la Serie de Tiempo.

Es importante analizar, que es lo que indican cada una de las estadísticas tomadas en cuenta para la identificación del modelo de la serie en cuestión, esto requerirá de una cierta experiencia de parte del analista, que será el que decida que coeficientes y de que grado serán los que intervengan en el modelo representativo de la serie.

Para determinar éstos coeficientes, será necesario encontrar con base en los datos, si se tienen o no los elementos de ciclo, tendencia, fluctuaciones irregulares o variaciones cíclicas o bien alguna combinación de ellos.

Estimación.

Cuando un modelo es especificado, el siguiente paso es el de estimar los parámetros que se decidió formen parte del modelo. Es en esta parte en la que intervendrá el uso de los paquetes estadísticos para evitar lo tedioso de efectuar las operaciones y dar así rapidez y exactitud a esta fase.

Diagnóstico del Modelo.

El siguiente paso es tomar el modelo y ver que efectivamente represente a la serie en estudio. Esto puede hacerse mediante la revisión de la función de autocorrelación de los residuales del modelo observado. Es decir, la serie resultante de restar la serie actual de la serie estimada.

De no pasar esta prueba, este modelo deberá ser modificado de manera que vuelvan a repetirse los pasos de identificación y estimación, hasta lograr que el modelo sea realmente adecuado a los datos que se tienen disponibles.

Si el modelo resultó adecuado deberá ser evaluado y analizado, esto es, se determinará el poder del modelo para predecir con exactitud los datos futuros y determinar las propiedades de predicción del modelo, como pueden ser las de tener estadísticas apropiadas que no limiten la utilidad para la predicción. Una de estas estadísticas podría ser que los parámetros estimados tuvieran errores estándar grandes, produciendo con esto que el error de la predicción se incrementara.

Un método de evaluación y análisis es el de hacer una "simulación histórica" del modelo, empezando en diferentes puntos del tiempo. Es decir, simular el modelo sobre el período en el que se hizo la observación pero empezando en distintos puntos dentro de ese intervalo.

Este tipo de análisis puede ayudar al investigador a decidir que tan grande podrá ser el período de predicción utilizando el modelo propuesto, tomando como base la exactitud que reflejen las predicciones hechas, contra las observaciones que se tengan. Generalmente los modelos de Series de Tiempo dan mejores resultados para muy cortos intervalos de predicción, es decir, se espera que para períodos cercanos al período de observaciones se tenga mayor exactitud que para los lejanos a ese período.

Una vez seleccionado el modelo que se utilizará se podrán efectuar las predicciones para la variable de interés.

Para calcular la mejor predicción es necesario especificar su exactitud, ésta puede ser expresada mediante el cálculo de límites de probabilidad tanto inferiores como superiores a la predicción. Los límites pueden calcularse para cierto conjunto conveniente de probabilidades (por ejemplo 50% o 95%). Estos límites serán tales, que el valor realizado de la serie al ocurrir, deberá estar dentro del intervalo descrito por los límites arriba mencionados con la probabilidad establecida.

Este valor será entregado al decisor quien sabrá que trato deberá dársele según sea el caso.

Finalmente, es importante hacer notar la estrecha relación que debe haber entre la persona que predice y aquella que tomará la decisión, que no en todos los casos es la misma. Cuando se escoge una técnica algo sofisticada esta relación suele no ser muy fácil, sin embargo es necesario hacerle ver al decisor una estructura general del proceso. Además como la estructura, desarrollo y objetivos del cambio en una organización llegan a modificarse y nuevas técnicas de predicción y fuentes de información se vuelven disponibles, la política y estrategia en el área de predicción deberá revisarse periódicamente por ambos.

3.2 IDENTIFICACION DE UN MODELO.

Recordando ante todo el objetivo de los modelos ARIMA, que es el tratar de predecir los valores de una Serie de Tiempo, con base en los valores pasados de la variable de interés que se tengan disponibles. Se buscará resolver las incógnitas de ¿Cómo se estimará la función de autocorrelación de la Serie de Tiempo?, ¿Cómo esa función de autocorrelación servirá para identificar un modelo ARIMA apropiado para la serie?, es decir, los valores apropiados para p , d y q . Y una vez identificado el modelo ¿cómo se identificarán los parámetros θ 's y Θ 's?, para finalmente como será utilizado el modelo obtenido para predecir a la Serie de Tiempo.

Estimación de la Función de Autocorrelación.

El primer paso, es el de dar una estimación para la función de autocorrelación, que será de gran ayuda para la identificación del modelo.

Recordando que el coeficiente de autocorrelación está dado por:

$$\rho_j = \Gamma_j / \Gamma_0$$

donde Γ_j es la autocovarianza al periodo j y Γ_0 al periodo cero.

La autocovarianza Γ_j es en turno, la esperanza del producto cruzado de la diferencia con la media del proceso de un par de observaciones separadas por j periodos, esto es,

$$\Gamma_j = E[(Z_t - \mu)(Z_{t+j} - \mu)]$$

Una estimación de Γ_j , sería el promedio de tales productos cruzados, donde μ se remplazaría por la media muestral.

Denotando por C_j a dicha estimación, se tiene,

$$C_j = \frac{1}{T} \sum_{t=1}^{T-j} [(Z_t - Z)(Z_{t+j} - Z)]$$

donde T es el número de observaciones disponibles y Z está dada por

$$Z = \frac{1}{T} \sum_{t=1}^T Z_t$$

finalmente la estimación de ρ_j , denotada por r_j , será

$$r_j = C_j / C_0.$$

Así, dado un conjunto de observaciones, la función de autocorrelación se obtiene calculando r_j para $j = 1, 2, 3, \dots$, donde, el mayor subíndice j será $T-1$, aunque en la práctica dicho cálculo se hace para un número mucho menor de periodos. Una forma de presentarlo, es como un correlograma muestral, el cual es una gráfica de las autocorrelaciones muestrales.

El propósito de conocer las autocorrelaciones muestrales, es el de poder ver en ellas un patrón típico de un proceso ARIMA. Por lo que el utilizar el correlograma muestral facilita dicha inspección.

Es importante recordar que las autocorrelaciones muestrales son sólo estimaciones de las verdaderas autocorrelaciones del proceso en cuestión.

Una aproximación para la varianza de r_j y la covarianza entre r_j y r_{j+k} para procesos Normales en el caso particular de un proceso de promedios móviles de orden q sería,

$$V(r_j) \approx \frac{1}{T} \left[1 + 2 \sum_{i=1}^q \rho_i^2 \right] \quad j > q \quad (1)$$

En general, la covarianza entre r_j y r_{j+k} será

$$\text{cov}(r_j, r_{j+k}) = \frac{1}{T} \sum_{i=1}^q \rho_i \rho_{i+k}$$

Esto sugiere que deben verse en los correlogramas las características generales que lo componen, y no tomar con demasiada importancia los detalles de dicho correlograma. Por ejemplo, las autocorrelaciones muestrales producidas por un proceso autoregresivo de primer orden, declinarán aproximadamente, aunque no exactamente de forma exponencial.

Enfrentándose con el problema de distinguir que es importante y que no en un correlograma, puede hacerse una prueba de significancia de la autocorrelación muestral. La aproximación para la varianza arriba mostrada (1), sugiere esta prueba. En particular, la varianza de r_j para $j > q$ está dada por dicha fórmula bajo la hipótesis de que el verdadero orden del proceso sea q . Y si se supone que la distribución para r_j con $j > q$ es Normal, puede verse una autocorrelación muestral en valor absoluto como

$$|r_j| > (1.96) \frac{1}{T^{1/2}} \left[1 + 2 \sum_{i=1}^q \rho_i^2 \right]^{1/2} \quad j > q$$

siendo significativamente distinta de cero al nivel .05. En la práctica como no se tiene el valor de ρ_i se sustituye por su estimación r_i quedando

$$SE(r_j) = \frac{1}{\sqrt{1/2}} \left[1 + 2 \sum_{i=1}^q r_i^2 \right]^{1/2} \quad j > q$$

y

$$|r_j| > 2 \frac{1}{\sqrt{1/2}} \left[1 + 2 \sum_{i=1}^q r_i^2 \right]^{1/2} \quad j > q$$

que podría utilizarse para probar si un r_j en cierto periodo distinto de q es razonablemente cero.

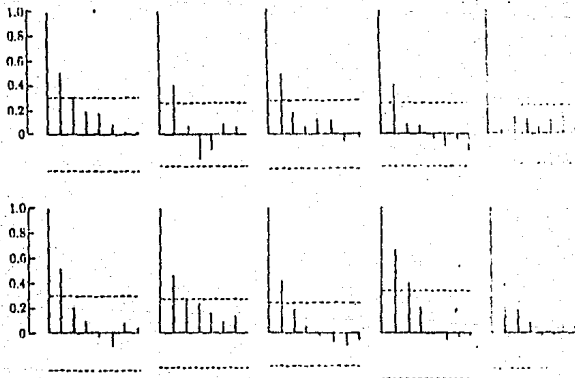
Debe tenerse cuidado al aplicar la prueba de significancia, ya que si se prueban los suficientes coeficientes de autocorrelación se encontrarían varios (5 de 100 aprox.) que serían significantes, lo cual complicaría la elección de q . La filosofía básica en la construcción de los modelos ARIMA es que el modelo final deberá tener una forma relativamente simple. En otras palabras, q deberá ser pequeña.

Para ilustrar esta sección tomese una función de autocorrelación teórica para el proceso AR(1)

$$Z_t = .5 Z_{t-1} + U_t$$

para el cual se produjeron artificialmente 10 realizaciones independientes del proceso. (Pag 72, [2])

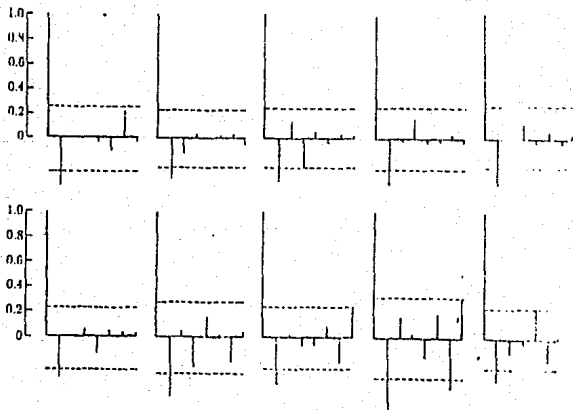
figure 5.1 Sample correlograms for 10 realizations of length $T = 100$ for the process $Z_t = .5Z_{t-1} + U_t$.



Similarmente para un proceso MA(1)

$$Z_t = U_t - .5 U_{t-1}$$

Figure 5.2 Sample correlograms for 10 realizations of length $T = 100$ of the process $Z_t = U_t - .5U_{t-1}$.



Determinación del Grado Apropriado de Diferencias.

Generalmente en la práctica, los datos no reflejan estacionaridad sino hasta después de aplicarles la primera o segunda diferencia. El problema entonces, es el de determinar el grado de diferencia que los datos necesitan, y como la autocorrelación puede indicar la diferencia requerida.

Las series no estacionarias presentan una gran autocorrelación muestral a lo largo del tiempo, debido ésto, a que la serie tiende a estar a un lado o a otro de la media muestral durante muchos periodos.

Para un proceso ARMA(1,q) la autocorrelación para $j > q$ es

$$p_j = \theta_1 p_{j-1} \quad j = q+1, q+2, \dots$$

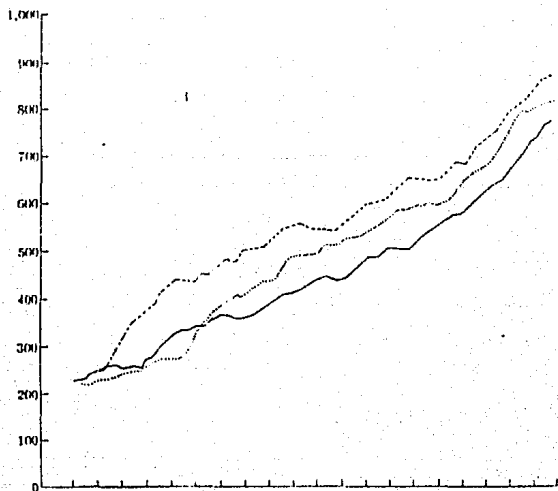
si θ_1 es muy cercana a uno, p_j declinará lentamente, por lo que en una estimación muestral de p_j , se esperaría que se

repitease lo mismo, esto es, que la autocorrelación muestral declinará lentamente

Si se cree que los datos son no estacionarios pueden examinarse las primeras diferencias, de no ser esto suficiente se examinarán las segundas diferencias, etc. Es necesario notar que sólo se necesita identificar el grado menor de diferencia para el cual la estacionaridad en el modelo es evidente, esto es debido a que las diferencias de series estacionarias son también estacionarias y no se gana nada aumentando el grado de diferenciación.

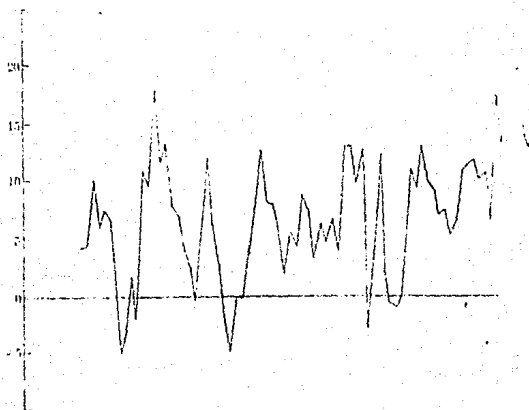
Si se tiene por ejemplo la siguiente gráfica de cierta serie:

figure 4.1 Actual GNP in billions of current dollars (solid line) and two simulations of ARI (1,1) model, quarterly 1947-01 through 1966-04.



puede verse que la serie no muestra afinidad respecto a una cierta media. En la siguiente figura se muestran las primeras diferencias de la serie anterior, y como puede verse, esta serie parece fluctuar alrededor de una media, de donde podría sugerirse que las diferencias de la serie original son ya estacionarias, por lo cual en el modelo d valdría 1.

Figure 4.2 Actual first differences of GNP in billion of current dollars in quarter 1947:01 through 1966:04.



Identificación del Orden de un Proceso Autoregresivo:
La Función de Autocorrelación Parcial.

Se espera distinguir en las autocorrelaciones muestrales la decaída exponencial como patrón de un proceso AR(1), pero para modelos más complicados, éste es, de orden mayor, es necesario ayudarse de otras estadísticas muestrales conocidas como autocorrelaciones parciales, que ayudan a hacer algunas distinciones.

El uso de las autocorrelaciones parciales, surge cuando se considera un modelo autoregresivo puro y el orden de dicho modelo está en duda.

Tomando ahora, las ecuaciones de Yule-Walker para un proceso AR(p)

$$\rho_1 = \theta_1 + \theta_2 \rho_1 + \dots + \theta_p \rho_{p-1}$$

.....

$$\rho_p = \theta_1 \rho_{p-1} + \theta_2 \rho_{p-2} + \dots + \theta_p$$

Si se conoce p y los valores de ρ_1, \dots, ρ_p , puede resolverse el sistema de p ecuaciones para las p

desconocidas θ_i , con $i=1\dots p$. En la práctica sin embargo, se ignoran los verdaderos valores de ρ_i como también p , lo cual es el problema.

Si se supone a $p=1$, la estimación de θ_1 puede obtenerse resolviendo la ecuación de Yule-Walker para $p=1$ en la que ρ_1 es reemplazada por la estimación r_1 , es decir,

$$r_1 = \theta_1$$

donde θ_1 es la estimación resultante de θ_1 . Ahora, si dicha estimación es razonablemente distinta de cero, podría concluirse que el proceso es por lo menos de orden 1. Si se desea saber si el proceso es de orden 2 o es mayor, pueden resolverse las ecuaciones de Yule-Walker para $p=2$

$$r_1 = \theta_1 + \theta_2 r_1$$

$$r_2 = \theta_1 r_1 + \theta_2$$

Si la estimación θ_2 resultante difiere razonablemente de cero, puede concluirse que el proceso es por lo menos de orden dos. Si se repite dicho procedimiento sucesivamente, cuando la estimación θ_{p-1} sea aproximadamente cero podrá decirse que p es el orden del modelo, es decir,

$$\theta_j \approx 0 \quad \text{para } j > p$$

Pero las autocorrelaciones parciales son estadísticas muestrales y por lo tanto, se tiene asociado a ésto un cierto error, por lo cual se necesita un estándar para decidir cuando se llega a "aproximadamente" cero. Bajo la hipótesis de que el orden del modelo es p , el error estándar estará dado por

$$SE(\theta_j) \approx 1/(T)^{1/2} \quad j > p$$

Entonces se infiere que p es el orden del modelo si θ_{p-1} es pequeño con respecto a $1/(T)^{1/2}$.

Si el proceso fuese MA, el poder escribirlo como un proceso AR de orden infinito y conociendo que θ_j es estimación de θ_j , se tendría que el θ_j debe declinar en magnitud, conforme se incrementan los valores de j , pero nunca llegar a cortarse en algún periodo, como sería característica de un proceso AR de orden finito.

El siguiente ejemplo, muestra también un proceso no estacionario, la gráfica, las autocorrelaciones, autocorrelaciones parciales y primeras diferencias están dadas en las cuatro figuras siguientes.

figure 4.3 Actual expenditures on producers' durables in billions of current dollars and two simulations of IMA (1,1) model, quarterly 1947-01 through 1966-04.

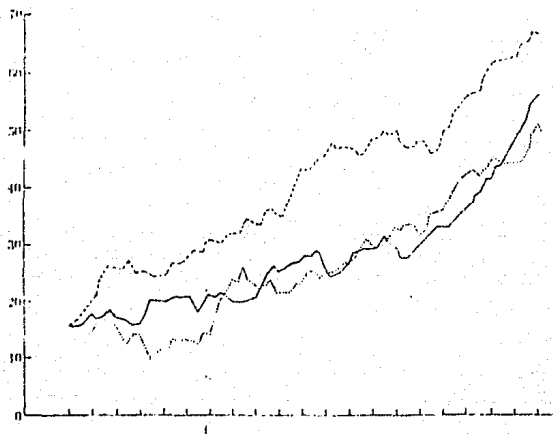


figure 4.4 Actual first differences of expenditures on producers' durables in billions of current dollars, quarterly 1947-01 through 1966-04.

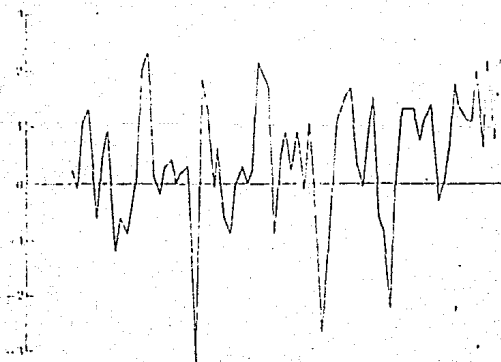


table 5.3 Sample autocorrelations and partial autocorrelations for EPD 1947-01 through 1966-04, PDO output

EXPENDITURES PRODUERS DURABLES										
AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTIONS OF ORIGINAL DATA										
UNDIFFERENCED DATA										
LAGS	AUTOCORRELATIONS									
0	1.00									
1-12	0.91	0.85	0.79	0.73	0.67	0.62	0.56	0.51	0.45	0.40
13-24	0.35	0.31	0.27	0.24	0.20	0.17	0.14	0.11	0.08	0.05
1-12	0.16	0.15	0.13	-0.12	0.11	0.09	0.07	0.04	0.03	0.02
13-24	0.01	0.01	-0.01	-0.01	0.00	-0.01	0.01	0.00	-0.01	-0.01
LAGS	PARTIAL AUTOCORRELATIONS									
0	1.00									
1-12	0.91	-0.03	-0.02	0.01	0.02	-0.01	0.02	-0.02	0.01	-0.03
13-24	0.35	-0.33	0.00	-0.03	0.33	0.02	0.30	-0.26	0.10	0.12

Figure 5.3 Sample correlogram for EPD, undifferenced series, 1947-01 through 1966-04, PDO output.

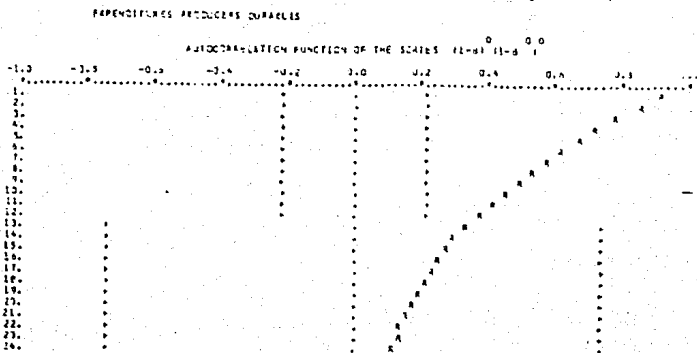
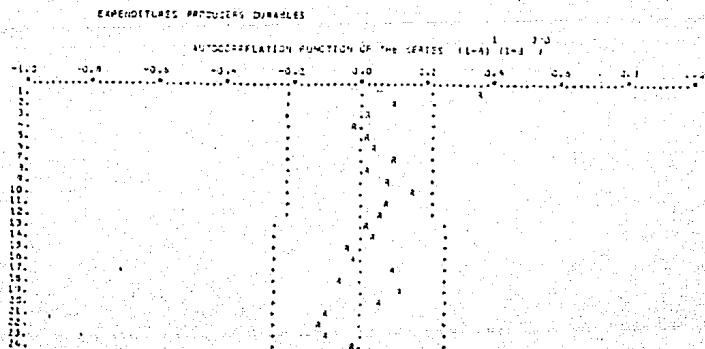


Figure 5.4 Sample correlogram of first differences of EPD, 1947-01 through 1966-04, PDO output.



Las autocorrelaciones de los datos originales decaen lentamente, en contraste, las autocorrelaciones de las primeras diferencias decaen rápidamente. Esto sugiere que un modelo apropiado para las primeras diferencias sería AR(1). Las autocorrelaciones parciales apoyan esta conclusión, cortándose de repente después de $p = 1$.

El modelo deberá incluir un término constante mientras que las primeras diferencias de la serie deben tener una media positiva como es de esperarse al ver la gráfica *

Así el modelo tentativo es uno de la forma

$$W_t = \theta_1 W_{t-1} + \delta + U_t$$

$$Z_t = Z_{t-1} + \theta_1 (Z_{t-1} - Z_{t-2}) + \delta + U_t$$

el siguiente paso será entonces, utilizar los datos para estimar valores de los parámetros θ_1 , δ y σ_U^2 .

Identificación de un Proceso Mixto.

El problema de la identificación de un modelo mixto es más complejo que el tener sólo un modelo AR o uno MA simples. Para ver el porqué, supóngase se hicieran diferencias para quitar la estacionaridad de algún conjunto de datos que hubiera sido generado por un proceso mixto. Bajo estas condiciones se esperan ciertas características para las autocorrelaciones y autocorrelaciones parciales muestrales.

Las primeras q autocorrelaciones, son determinadas por parámetros de ambos modelos AR y MA a través de cierta relación algo compleja. Las autocorrelaciones para periodos mayores a q están dadas por

$$\rho_j = \theta_1 \rho_{j-1} + \dots + \theta_p \rho_{j-p} \quad j > q$$

Esto sugiere que se espera un patrón regular en las autocorrelaciones hasta un período inicial asumido como q .

Se sabe que un modelo mixto cuando se escribe de forma autorregresiva tiene un orden infinito, entonces las autocorrelaciones parciales, como estimaciones de esos coeficientes autorregresivos, decaerán mientras se corta definitivamente, consecuentemente un modelo mixto tiende a demostrar una declinación gradual tanto en las autocorrelaciones como en las autocorrelaciones parciales muestrales.

Haciendo entonces un resumen de lo anterior, puede mostrarse la siguiente tabla:

CLASE DE PROCESO	AUTOCORRELACIONES	AUT. PARCIALES
MA	picos en los periodos 1 a q, luego se corta.	se va desvaneciendo.
AR	desvanece con $\rho_j = \theta_1 \rho_{j-1} + \dots + \theta_p \rho_{j-p}$	picos en los periodos 1 a p luego se corta.
ARMA	patrón irregular de los periodos 1 a q, luego desvanece con $\rho_j = \theta_1 \rho_{j-1} + \dots + \theta_p \rho_{j-p}$	se desvanece

Estimación Preliminar de Parámetros.

Habiendo hecho un modelo tentativo para una particular Serie de Tiempo, pueden utilizarse las autocorrelaciones muestrales para obtener estimaciones preliminares de los parámetros. Estas estimaciones son útiles, ya que, sugieren un bosquejo del modelo final, además, dan los valores iniciales para el procedimiento iterativo, utilizado para calcular las estimaciones de máxima verosimilitud para los parámetros.

Las estimaciones preliminares son obtenidas, resolviendo para los valores de los parámetros, la relación que une parámetros y autocorrelaciones.

El procedimiento para un proceso AR, es el de dar soluciones a las p ecuaciones de Yule-Walker sustituyendo r_j en lugar de ρ_j , es decir, resolver el sistema

$$r_1 = \theta_1 + \dots + \theta_p r_{p-1}$$

.....

$$r_p = \theta_1 r_{p-1} + \dots + \theta_p$$

con el que se obtienen las estimaciones de los p parámetros $\theta_1, \dots, \theta_p$.

La relación entre autocorrelaciones y parámetros no es lineal para un proceso MA o uno mixto, y entonces, el cálculo de la estimación preliminar no es tan directa.

Para un MA(1) se tiene

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

entonces una estimación de θ_1 puede ser obtenida de

$$r_1 = -\theta_1 / (1 - \theta_1^2) \quad (2).$$

Notese que, (2) es una ecuación cuadrática en θ_1 y tiene 2 soluciones

$$\theta_1 = - \frac{1}{2r_1} \pm \left[\frac{1}{(2r_1)^2} - 1 \right]^{1/2}$$

y cada solución es el recíproco de la otra. Entonces las estimaciones alternativas pueden ser por ejemplo .2 y 5. Así si una alternativa satisface la condición requerida de $|\theta_1| < 1$ la otra la violará. Entonces, la alternativa escogida será aquella que cumpla con esta condición.

Para procesos MA de orden mayor las q ecuaciones

$$r_1 = \frac{-\theta_1 + \theta_1\theta_2 + \dots + \theta_{q-1}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}$$

$$r_q = \frac{-\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}$$

pueden en principio ser resueltas para estimar los parámetros $\theta_1, \dots, \theta_q$. En general, se tendrán soluciones múltiples y la estimación escogida se basará en la satisfacción de los requerimientos.

Para continuar con los procesos mixtos, un procedimiento general es el siguiente. De las ecuaciones

$$p_{q+1} = \theta_1 p_q + \dots + \theta_p p_{q-p+1}$$

$$p_{q+p} = \theta_1 p_{q+p-1} + \dots + \theta_1 p_q$$

pueden estimarse $\theta_1, \dots, \theta_p$ si el r_j es sustituido por p_j .

Entonces con la relación entre p_1, \dots, p_q , las θ 's y las θ 's, puede resolverse la relación para $\theta_1, \dots, \theta_p$ sustituyendo r_j en lugar de p_j y las estimaciones de las θ 's se obtendrán directamente de su relación con r_j ; por ejemplo para un ARMA(1,1)

$$\theta_1 = r_1 / r_2$$

y θ_1 estará dada por la solución de

$$r_1 = \frac{(1 - \theta_1\theta_1)(\theta_1 - \theta_1)}{1 + \theta_1^2 - 2\theta_1\theta_1}$$

que nuevamente se escogerá el valor que cumpla con las condiciones pedidas anteriormente.

Una estimación preliminar del término constante en un modelo ARIMA, puede obtenerse de la relación entre la media del proceso y los demás parámetros usando la media muestral de la serie como estimación.

Aunque las estimaciones de los parámetros se derivan de los momentos muestrales de la Serie de Tiempo, son útiles para un primer bosquejo del modelo. Las estimaciones por máxima verosimilitud, son generalmente eficientes, al menos para muestras de gran tamaño.

Estimación por Máxima Verosimilitud para Modelos ARIMA.

Habiendo identificado uno o más modelos tentativos para una Serie de Tiempo, se desea obtener la "mejor" o "más eficiente" estimación de los parámetros antes de predecir. Para ello puede mostrarse que las estimaciones que maximizan a la función de máxima verosimilitud, son las estimaciones más eficientes, si el número de observaciones que se utiliza es grande. Por lo anterior, será conveniente echar un vistazo a las estimaciones por máxima verosimilitud, para las que existe un procedimiento con el que pueden establecerse estimaciones para cualquier modelo ARIMA.

Tomese un conjunto de observaciones Z_1, \dots, Z_T , y un modelo que se supone aproxima a las observaciones con que se cuenta. Para este caso, las observaciones son variables aleatorias que siguen una distribución conjunta definida por el modelo ARIMA y sus parámetros desconocidos $\theta_1, \dots, \theta_n, \delta_1, \dots, \delta_m, \delta$ y σ^2 . Esto es, las observaciones seguirán una distribución conjunta $p(w | \theta, \delta, \sigma^2)$ donde w denota al vector de observaciones en la diferencia para la cual la serie es estacionaria. θ el vector de las θ 's, etc.

Teniendo las observaciones disponibles, se querrá saber el valor de los parámetros que pueden estimar el comportamiento de dichos datos. Así, la función de verosimilitud $L(\theta, \delta, \sigma^2 | w)$ que es de la misma forma que la función de densidad de probabilidad para w , considera ahora a los parámetros como variables y a las observaciones como dadas.

Haciendo una ligera revisión de la técnica de estimación por máxima verosimilitud, y asumiendo que las observaciones se distribuyen como Normal y las U_t 's se distribuyen de igual forma, se comenzará el análisis.

La función de distribución para un error U_t está dado por:

$$p(U_t | \sigma^2) = (2\pi)^{-1/2} (\sigma^2)^{-1/2} \exp[-(U_t^2) / (2\sigma^2)]$$

Si los U_t 's son independientes su función conjunta de distribución es precisamente el producto de las distribuciones marginales, ésto es,

$$p(U_1 \dots U_T | \sigma_u^2) = (2\pi)^{-T/2} (\sigma_u^2)^{-T/2} \exp[-1/(2\sigma_u^2) \sum_{t=1}^T U_t^2]$$

y como U_t puede expresarse

$$U_t = w_t - \theta_1 w_{t-1} - \dots - \theta_p w_{t-p} - \delta + \theta_1 U_{t-1} + \dots + \theta_q U_{t-q}$$

llamando $U(\theta, \theta, \delta)_t$ a la estimación de U_t en la ecuación anterior.

Sustituyendo U_t en $p(U_1 \dots U_T | \sigma_u^2)$ queda:

$$p(w | \theta, \theta, \delta, \sigma_u^2) = (2\pi)^{-T/2} (\sigma_u^2)^{-T/2} \exp[-1/(2\sigma_u^2) \sum_{t=1}^T U(\theta, \theta, \delta)_t^2]$$

Considerando ahora el logaritmo de la función se tiene

$$l(\theta, \theta, \delta, \sigma_u^2 | w) = -T \ln \sigma_u^2 - \frac{S(\theta, \theta, \delta)}{2 \sigma_u^2}$$

donde $S(\theta, \theta, \delta)$ es $S(\theta, \theta, \delta) = \sum_{t=1}^T U(\theta, \theta, \delta)_t^2$ la suma de cuadrados

Puede verse que los parámetros θ, θ, δ sólo están presentes en la suma de cuadrados, por lo que para maximizar la función de verosimilitud es necesario minimizar la función de suma de cuadrados sobre los valores de los parámetros.

Se puede mostrar que

$$\sigma_u^2 = S(\theta, \theta, \delta) / T,$$

donde θ, θ y δ son los parámetros estimados por máxima verosimilitud.

Predicciones Hacia Atrás (Backforecast).

Algunas veces es necesario calcular los valores

$$w_0, w_{-1}, w_n$$

de la serie, que ocurrieron antes de la primera observación que se hizo. Esto sucede porque los "valores iniciales" son necesarios como base para los cálculos recursivos usados para estimar los parámetros del modelo.

El valor w_t tiene exactamente la misma relación de probabilidad de la secuencia w_1, w_2, \dots, w_n como el valor w_{n+1} de la secuencia $w_n, w_{n-1}, w_{n-2}, \dots, w_1$. Entonces para estimar un valor $l+1$ periodos antes de que las observaciones empiecen, puede primero considerarse cual podría ser la estimación opcional o predicción $l+1$ periodos después del fin de la serie, y entonces aplicar este procedimiento en forma inversa. En otras palabras "se predice" la serie al

revés. A este procedimiento se le llama *predicción hacia atrás*.

Inferencia Estadística:
Error Estándar para Estimaciones Máximo Verosimiles.

Después de obtener la estimación por máxima verosimilitud de los parámetros para el modelo ARIMA, sería bueno saber la precisión de dichos estimadores, o bien poder tener un intervalo que contenga al verdadero valor de el parámetro en cuestión.

Es necesario recordar que las estimaciones por máxima verosimilitud de los parámetros, son variables aleatorias y función de los datos. En muestras de gran tamaño se tiene una distribución Normal conjunta con media igual al verdadero parámetro y matriz de varianzas y covarianzas dada por

$$V(\beta) \approx 2\sigma_u^2 \begin{bmatrix} \frac{\partial^2 S(\beta)}{\partial \beta_1^2} & & & & \\ & \dots & & & \\ & & \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_{p+q+1}} & & \\ & & & \frac{\partial^2 S(\beta)}{\partial \beta_{p+q+1}^2} & \\ & & & & \dots \end{bmatrix}$$

esto es, dos veces la varianza de U_t por la inversa de la matriz de segundas derivadas de la función de sumas de cuadrados, donde β es el vector de los $p+q+1$ parámetros ϕ, θ, δ .

Con el fin de hacerlo operacional se sustituye a σ_u^2 por $\hat{\sigma}_u^2$ y las segundas derivadas se evalúan con los valores de las estimaciones máximo verosimiles.

Tomando la raíz cuadrada de los elementos de la diagonal de la matriz estimada de varianzas y covarianzas, se estimará la desviación estándar de los parámetros estimados denotados por $SE(\hat{\beta}_i)$.

Los errores estándar permiten hacer enunciados probabilísticos de los parámetros. Bajo la hipótesis de que $\beta_i = \beta_{i0}$ se tiene aproximadamente que

$$\frac{\hat{\beta}_i - \beta_{i0}}{SE(\hat{\beta}_i)} \sim \mathcal{N}(0, 1)$$

de la que puede entonces tenerse la probabilidad con que se puede o no rechazar la hipótesis de que $\beta_i = \beta_{i0}$.

Diagnóstico.

Una vez identificado el modelo y habiendo estimado sus parámetros se desea checar la adecuación de ese modelo a los datos que se tienen. Por ejemplo, si el modelo es un $ARI(1,d)$ se desearía saber si un $ARI(2,d)$ pudiera ser un modelo más apropiado, o si se debería añadir un término de promedios móviles para formar un $ARIMA(1,d,1)$.

Una forma de diagnóstico podría ser el agregar el parámetro y probar la hipótesis de que el parámetro añadido es cero.

Si se deseará expandir el modelo tanto en dirección autoregresiva como de promedios móviles podría caerse en la redundancia de parámetros.

Otra forma podría ser la de examinar la autocorrelación muestral de los residuales. Si la autocorrelación es evidente en los residuales, esto sugiere en que forma el modelo deberá ser modificado. Por ejemplo, un valor grande de r , puede indicar que un orden de promedios móviles deberá agregarse al modelo.

Finalmente, es útil una inspección a la gráfica de residuales para ver si ésta refleja una adecuación del modelo. Por ejemplo, si no se tiene homogeneidad en una serie con tendencia positiva, esto se reflejará en la gráfica por una amplitud general de la escala de residuales conforme el periodo muestral avanza.

IV CRITERIOS A CONSIDERAR PARA UN "BUEN" PAQUETE ESTADISTICO.

Antes de ver, el como un paquete computacional estadístico puede ayudar en el cálculo y estimación de los parámetros de los modelos ARIMA, se procederá a dar algunos criterios con los cuales se podrá escoger, de entre los paquetes disponibles, uno que se adapte a las necesidades del usuario.

Los paquetes de computadora para análisis estadísticos han proliferado en gran manera. Sin embargo, esto no implica que dichos paquetes hayan sido entendidos en su totalidad por los usuarios, que necesitan de ellos para facilitar el proceso de toma de decisiones.

Por otro lado, éstos paquetes han sido de gran ayuda para los estadísticos (y algunos otros usuarios), debido a que les ha permitido manipular un mayor número de datos, ofreciéndoles mayor rapidez en los cálculos, así como la disminución de errores en los mismos, permitiendo además una mayor flexibilidad y versatilidad en el análisis.

Una primera aproximación a la evaluación de este tipo de paquetes, es la de considerar el rango de funciones ofrecidas, probarlas una a una observando velocidad de procesamiento, la capacidad límite y la precisión de resultados, comparando el rendimiento de cada paquete basándose en este tipo de análisis. O bien, puede tomarse una visión más extensa del programa, considerando aspectos como la calidad de interacción con el usuario, el rango de funciones ofrecidas, etc. Se tratará entonces de combinar ambas aproximaciones, para que con esto se tenga oportunidad de estimar el valor de los paquetes y de identificar los elementos que se esperarían tengan los productos existentes.

Generalmente, el usuario que requiere un paquete estadístico utiliza el que tiene a su alcance o el que "alguien" le dijo que era bueno. Por desgracia, al trabajar con la mayoría de los paquetes dan la impresión de no haber

sido hechos o revisados por un equipo de estadísticos y expertos en computación. Como éstos paquetes están disponibles cada vez para un mayor número de personas, se puede tener como consecuencia el que dichos paquetes contengan errores, o bien, que sean utilizados para propósitos diferentes a los que el autor contempla, o, que las técnicas contenidas no ofrezcan resultados que cumplen con los estándares establecidos.

A continuación se presentan, algunos de los criterios y consideraciones, útiles para la evaluación de los paquetes estadísticos, que según Ivor Francis, Richard M. Heiberger y Paul F. Velleman en su artículo "Criteria and Considerations in the Evaluation of Statistical Program Packages" [5] y donde se expone la descripción de lo que sería un paquete ideal, y por supuesto sin esperar que ninguno cumpla con todas estas características, se intentará buscar aquel que cumpla con el mayor número de éstas.

CRITERIOS Y CONSIDERACIONES.

Es necesario notar, que un sólo paquete no puede ser óptimo para todas las situaciones, y que no sólo un criterio es importante para la evaluación de un paquete. Así la importancia de cierto criterio podrá cambiar junto con las expectativas del usuario.

Si se divide a esos criterios en tres grandes bloques, se tendrá:

- Interface con el usuario.
- Efectividad Estadística
- Implementación.

El primer bloque, contempla los tópicos de interés inmediato para el usuario, como el manejo del paquete y la comprensión del manual. El segundo bloque, incluye las funciones y capacidades del paquete estadístico. Y finalmente, el tercer bloque, incluye las características computacionales del paquete.

Los diseñadores del paquete, con el fin de cubrir los criterios arriba mencionados, deberán balancear la versatilidad requerida por los usuarios a los que se enfocará el modelo, pudiendo así buscar, la facilidad de uso, la forma de presentación de resultados, la forma de efectuar los cálculos, la eficiencia, la extensión deseada por los estadísticos o, la facilidad en la estructura del lenguaje necesaria para los usuarios ocasionales. Siendo por todo lo anterior muy difícil el poder dar una definición de "facilidad de uso", en un paquete estadístico.

Es importante que el programador conozca y tome en cuenta los problemas y limitaciones del usuario al diseñar un sistema. Henry Simpson en su artículo "A Human-Factors

Style Guide for Program Design" define seis factores para tres áreas del diseño : Introducción de Datos, Diseño de Pantallas y Control del Procedimiento.

Sugiriendo para ayudar a que la introducción de datos sea mejor manejada el que deban validarse los datos antes de ser guardados en las bases de datos, y que se permita la corrección de ellos si es necesario. En cuanto al diseño de pantallas, éstas deberán tener un título que las identifique, y los mensajes un cierto formato. El control de procedimientos en los programas interactivos de hoy en día se maneja a base de menus, o programas de pregunta-respuesta en los que el usuario sólo tiene que escoger una opción o contestar ciertos datos que le son solicitados evitando con ello el que el usuario olvide el cómo continuar un procedimiento.

Para ello es necesario conocer las necesidades del usuario, la diversidad en los niveles de conocimiento entre ellos, la complejidad del programa y las consecuencias de los posibles errores del operador.

Los seis principios básicos que muestra Simpson son:

- *Responder al Usuario.*

El usuario al ejecutar un comando espera respuesta del paquete, la cual, puede no ser siempre inmediata, pudiéndose pensar que la acción tomada no ha tenido efecto. Esto puede causar que el usuario reintente el comando perjudicando con ello el funcionamiento del paquete. Para evitar lo anterior, el paquete deberá siempre responder al menos una leyenda que indique lo que en ese momento se está haciendo.

- *Ser Consistente.*

Es necesario definir un conjunto de reglas que el propio programador siga. Estas reglas permitirán al usuario (a base de operar el paquete) el saber como son desplegados menus, mensajes de error, etc.

- *Minimizar la Demanda de Memoria Humana.*

Cuando el usuario requiere seleccionar un subprograma, es más fácil el que se le muestre un menú o panel que le haga recordar la forma de llamar a dicho procedimiento.

- *Hacer el Programa Fácil.*

Lográndose esto después de un concienzudo diseño, quitando lo innecesario y reorganizando cuantas veces se requiera.

- *Revisar el Nivel de los Usuarios.*

Es necesario tener en cuenta los objetivos del programa, las decisiones que se deben tomar dentro de una sesión y el nivel de conocimiento que se requiere.

- *La Orientación del Usuario.*

Debe darse al usuario la posibilidad de saber en que módulo se está, y como regresar al paso anterior.

INTERFACE CON EL USUARIO.

Documentación del Usuario.

Un punto determinante en el uso de un paquete estadístico, es la documentación para el usuario, esta documentación, deberá incluir un manual elemental y otro de referencia, así como también es de esperarse un manual técnico y otro de procedimientos numéricos (pudiendo estar incluidos todos en uno mismo).

Todos los manuales deberán estar bien organizados, contarán con un índice general, deberán estar impresos en forma legible y de preferencia ser capaces físicamente de soportar una constante utilización.

Un manual elemental es reconocido por ser conciso y claro, en el cual deberán ilustrarse ejemplos elementales antes de introducir complicaciones. Cada ejemplo, deberá satisfacer la demanda del problema, mostrando además listados de como el paquete muestra la entrada y salida de los datos.

El manual de referencia deberá incluir todo lo que el usuario necesita saber sobre el paquete. Como su exactitud y organización. Debe documentar exactamente lo que el paquete hace, esto es, que funciones realiza y con que método lo hace. Debe estar bien organizado, contando con un índice general, una sección con el uso y la sintaxis brevemente explicados para cada uno de los comandos del paquete. Deberá contener información general acerca de los requerimientos mínimos como son, la versión de la cual se trata, los requerimientos de hardware y si se pueden o no agregar rutinas extras. También será deseable que se cuente con ayuda para el manejo del paquete dentro de la misma corrida y en cualquier momento de ésta.

En el manual de procedimientos numéricos se tendrá información acerca de como, poder probar la exactitud de los resultados de ciertos procedimientos, o bien, dar información de si esto fué probado y con que técnica se hizo. Sería deseable también la referencia bibliográfica de la técnica utilizada.

En el manual técnico deberá indicarse la referencia de donde se obtuvo la información y las técnicas que se utilizaron para los procedimientos del paquete.

Todos los manuales, deberán contener las convenciones de sintaxis para el manejo del paquete, describiendo además los posibles errores.

Para cada procedimiento del paquete, el manual de referencia deberá incluir una descripción completa del procedimiento utilizado, siendo ésta lo suficientemente clara, pero evitando profundizar demasiado, esto con el fin de no convertir al manual en un libro de estadística. Por otro lado para el usuario interesado, o el que desconoce sobre el tema deberá contener referencias en literatura

tanto estadística como de técnicas numéricas empleadas. Además, una documentación completa de los ejemplos utilizados para ilustrar los procedimientos, una especificación de valores de "default" y del procedimiento a seguir en cuanto se presentan datos faltantes, así como una lista de los mensajes de error.

Adicionalmente los tópicos deseables serían, desde el tener a la mano un procedimiento extra que muestre los pasos a seguir para cada uno de los métodos cubiertos por el paquete, hasta el resultado que se obtiene de cada uno de ellos. Además una clara y exacta descripción de los comandos y de las opciones disponibles, documentación de las limitaciones en tamaño y precisión aritmética, así como estimaciones del tiempo de ejecución para los procedimientos. Para ciertos usuarios que sólo desean utilizar un sólo procedimiento sería deseable que la documentación para cada uno de ellos fuera independiente de los demás. Sería también deseable que para cada punto de decisión se tuviera una lista de las posibles opciones de entre las cuales pudiera elegirse.

La documentación completa deberá constar entonces de, los manuales elemental, de referencia, técnico y de procedimientos numéricos, las impresiones de salida de resultados, instrucciones de instalación y una lista de las fuentes del paquete.

Lenguaje de Control. (Comandos)

Todos los paquetes crean a su alrededor un cierto ambiente que afecta siempre al entendimiento del usuario y al análisis de sus datos, debido a que el usuario debe aprender antes que nada a manejar el paquete que ha decidido utilizar. Se ha visto que muchas veces lo que el usuario piensa hacer con sus datos en cierto paquete, se ve limitado por lo que el puede hacer con ese paquete, debido ésto, a lo poco que conoce acerca del lenguaje del paquete.

Los paquetes más eficientes, muestran con números de código o menús la especificación del procedimiento, variables utilizadas y opciones disponibles. Para ciertos procedimientos, cuando los pasos a seguir son similares o se repiten frecuentemente, se puede tener una cierta rutina mostrada por un panel en la pantalla. De no existir, el usuario puede verse forzado a acudir demasiado al manual, tanto para investigar comandos, como para interpretar resultados, produciendo que el análisis se vuelva más tardado. Cuando se utiliza una sola variable en varios procedimientos, es deseable, el poder asignar etiquetas o términos descriptivos, que permitan la especificación opcional a las variables, ésto con el fin de evitar confusiones, además de lograr una representación más clara y legible de los resultados.

Además se esperaría, que los comandos estén organizados de tal manera que los procedimientos puedan ser llamados con términos que describan su función, esto, para ayudar a recordar el comando deseado, así el usuario, con un poco de experiencia y el panel de control podrá hacer el análisis mucho más rápido.

En cuanto a la transformación de variables, estaría bien que se especificarán con un formato fijo y con una sintaxis algebraica que sea familiar al usuario.

En referencia a los errores, deberá existir una cierta seguridad para el usuario, sea cual sea el error éste no deberá ser fatal y habrá de tenerse además un mensaje informativo del error cometido.

Estructura de los Datos.

Mucho del trabajo que es ofrecido por los paquetes estadísticos es el de facilitar la manipulación de los datos. También es de esperarse un buen manejo de los archivos de datos, el permitir la importación y exportación de datos tanto de archivos (ver anexo B) ASCII, DIF, como de hojas electrónicas o manejadores de bases de datos (LOTUS o DBASE), hasta el manejo de datos dentro del paquete permitiendo graficarlos en pantalla para la localización de puntos aberrantes o valores faltantes dentro de los datos. En caso de aparecer datos faltantes, saber cuales son los valores de default que se utilizan o que procedimiento se sigue.

Es de esperar entonces, que si existe algún problema con el método que se está siguiendo, como por ejemplo, los valores faltantes, se indique claramente el tratamiento que se les dará, para este caso, si se excluirán automáticamente o no.

Impresión de Resultados.

La impresión de resultados es otro atributo a considerar. Todo lo que ayude a que la impresión sea más clara será de gran importancia, como el poder utilizar graficadores y otras herramientas útiles para este efecto. Además, esta impresión deberá ser completa, no voluminosa y sobre todo ordenada. Un paquete deberá imprimir toda la información necesaria, siendo además, esta impresión opcional.

Los resultados por impresora deberán presentarse en forma estándar y clara, con el fin de que cualquier persona que conozca el procedimiento estadístico pueda leerlo y comprenderlo sin que sea necesario que conozca el paquete. Los resultados deberán ser lo suficientemente claros para minimizar la necesidad de acudir al manual para identificar variables y procedimientos.

Dentro de la impresión de cualquier resultado sería conveniente incluir algunos datos, con el fin de poder distinguir el paquete origen por el que fué producido, éstos podrían ser, la fecha de procesamiento, la versión y el nombre del proceso estadístico.

Gráficas.

Las gráficas son necesarias, tanto para todo el período de análisis, como para la conclusión del procedimiento. Estos elementos como gráficas, histogramas, correlogramas, etc. son de gran ayuda y casi imposibles de sustituir por elementos no gráficos.

La evaluación de capacidad de graficación es algo complicada, debido tanto a la reciente proliferación de graficadores como a las múltiples áreas donde se les puede juzgar, éstas pueden ser, su claridad, si se tienen o no etiquetas, si se efectuó correctamente la decisión de escala, si se pueden o no hacer correcciones a la gráfica (como reetiquetar), si los residuales se grafican en la misma escala en la que los datos fueron graficados, y si al graficarse la información puede ser guardada para procedimientos posteriores, o si a elección del usuario puede ser utilizado un graficador.

Costo.

Dentro de la accesibilidad de un paquete es también importante considerar el costo que genera el efectuar una corrida. Regularmente cuando un paquete es bueno en términos de su lenguaje, capacidad y exactitud es muy caro. Sobre todo se sabe que al pedir todas las características deseadas se tendrá asociado un cierto costo, es importante también tomar este factor en cuenta, ya que muchas veces aunque un programa tenga muchas de las características deseadas, exige contar con recursos que no se tienen, éstos pueden ir desde el número de diskettes del paquete, hasta ciertos requerimientos de hardware. Sin embargo el paquete de menor costo, podría no ser el mejor si se toma en cuenta el tiempo utilizado por el usuario para aprender a utilizar el paquete. Es también necesario considerar que el paquete escogido haga lo que el usuario desea, es decir, contenga las funciones requeridas y con el método deseado. Aunque este tipo de costo es también importante y deberá ser considerado, no se tomará en cuenta en este trabajo.

La documentación del procedimiento deberá incluir fórmulas que permitan estimar el costo de los componentes en una corrida en términos de las dimensiones del problema.

Alcance.

Usualmente el paquete se diseña para cierto tipo de personas. Así, las capacidades, vocabulario y lenguaje de control son especiales. Algunos diseñadores, esperan de sus usuarios un considerable conocimiento estadístico, mientras que otros asumen un conocimiento sólo en un cierto tópic. Por lo que al evaluar un paquete es importante tener este aspecto en cuenta.

Ayuda.

Un paquete requiere siempre de un nivel de ayuda adecuado, que permita en todo momento recordar la forma de efectuar los diversos procedimientos que contiene, por lo que es de esperarse que el paquete cuente con ayuda lo suficientemente clara para que el usuario no tenga que recurrir en forma excesiva al manual.

ECTIVIDAD ESTADISTICA.Versatilidad.

El análisis de los datos no consiste sólo en buscar la técnica estadística apropiada, si no que, por ser dicha técnica una secuencia de procedimientos, es necesario revisar los resultados haciendo una pausa al final de cada uno de ellos, para considerar entonces el siguiente paso. Por esto, es deseable que un paquete estadístico permita hacer dicha pausa para la revisión de los resultados.

Para iniciar un análisis es importante primero el preparar los datos. Refiriéndose esto, a que el paquete cuente con un conveniente sistema de archivos, como ya se había mencionado. Pueden también quererse gráficas, distribuciones de frecuencia, pruebas de normalidad o de otras suposiciones, así como poder efectuar transformaciones a dichos datos.

Así cuando los datos han sido preparados el usuario busca de entre los procedimientos disponibles aquel que más se adapte a sus necesidades.

Exactitud.

Es de esperarse que en un paquete los cálculos sean exactos, las fórmulas correctas y apropiadas, los algoritmos bien definidos, numéricamente estables y programados apropiadamente. Y que las técnicas utilizadas estén documentadas con referencias literarias.

Para poder evaluar la exactitud de los datos, procedimientos y estadísticas, como errores estándar, matrices, etc. Algunos procedimientos pueden presentar inestabilidad numérica tan grande que puede perjudicar de gran manera los resultados. Los paquetes deberán monitorear cálculos que detecten estos problemas, para luego, reportarlo al usuario en vez de dar resultados inexactos o provocar otras violaciones como asumir normalidad o independencia, o bien, aplicar alguna acción correctiva sin dar ningún aviso al usuario. Todas estas suposiciones y acciones correctivas deberán contemplarse en el manual, especificándose claramente para cada uno de los procedimientos.

IMPLEMENTACION.

Documentación del Programador.

Sería bueno poder contar con el manual del programador y el listado fuente del programa. Con este listado se sabría con exactitud lo que hace el paquete y como lo hace, así como los estándares que sigue. Las nuevas técnicas estadísticas serían más fáciles de entender, si junto con el listado se tuviera una guía de la descripción algebraica de los cálculos, con la cual se permitiría al usuario localizar y sugerir correcciones a errores en el paquete, ya que a veces un error puede ser detectado rápidamente por el usuario, más que por el programador. Sin embargo, a pesar de las ventajas que esto tendría, los propietarios de los paquetes sólo distribuyen el código objeto y casi nunca el código fuente del programa, debido a su propia seguridad, ya que así, sin tener los usuarios el código fuente del paquete, se tiene el problema de que se hacen copias piratas y el autor tiene grandes pérdidas, ahora si se tuviera, las copias piratas se harían con mayor frecuencia pudiendo hacerse modificaciones al paquete y distribuyéndose de forma pirata, perjudicando todavía más al autor, debido a que las modificaciones pueden ser erróneas, de lo cual se le culpaba, o bien, podrían venderse bajo el nombre del usuario ya corregidas perjudicando así también al autor. Por otro lado, si el autor decidiera distribuir también el código fuente, el costo de éste se incrementaría de gran manera, y sería esto bajo ciertas condiciones.

De no ser posible lo anterior, sería muy conveniente el poder tener contacto con el autor, para corregir errores o dar sugerencias acerca del paquete.

Por desgracia el alto costo del software, propicia su copia "pirata". Si estos costos disminuyeran, mayor número de personas tendría acceso a ellos siendo esto una gran ventaja, ya que toda persona desea disfrutar las ventajas de tener un software original, como son la comunicación con el

autor, el tener un manual original, o tener la información de cualquier modificación al paquete.

Flexibilidad.

Un paquete estadístico no puede estar 100% actualizado ni tampoco contener todas las nuevas técnicas estadísticas. Si no se ofrece un procedimiento estadístico en particular sería deseable que un paquete estadístico hiciera alguna de las siguientes cosas: Podría permitir, como ya se mencionó, que los datos fueran exportados del paquete para ser utilizados después, dejándolos en forma legible para la máquina (por ejemplo en ASCII), o bien, podría permitir al usuario añadir sus propias rutinas al sistema. Los paquetes que cuentan con este último elemento tienen la ventaja de dar al usuario una mayor flexibilidad y facilidad al manejo de sus datos.

Distribución.

Los paquetes son desarrollados para cierto tipo de hardware. Cuando en alguna instalación se cuenta con la misma computadora y sistema operativo principalmente, puede transferirse una copia del modulo ejecutable, de no tenerse la misma instalación el código fuente será el que se utilice para adaptar el paquete a la nueva instalación. Esta forma de transferencia puede traer problemas, debido a que en la adaptación de un lenguaje a otro puede haber algunas ambigüedades serias con las que debe tenerse cuidado.

Algunos programas se transfieren a través de sistemas sin ningún problema, mientras que otros, debido a su dependencia con el sistema (como convenciones de I/O, diferencias de hardware o segmentación del programa) no pueden transferirse fácilmente. Sin embargo, el crecimiento de redes en computadora ha facilitado en gran manera éstos problemas de transferencia, aunque debe cumplirse con cierta compatibilidad de máquinas (todavía no es posible que un paquete de PC corra en alguna máquina grande, aunque la transferencia pueda hacerse).

Lenguaje Fuente.

El lenguaje en el que el paquete está escrito puede afectar el trabajo ofrecido por dicho paquete. Por ejemplo, el alto costo de la interpretación del control del formato libre. Otro ejemplo es el de las operaciones de manejo de cuerdas requeridas, que toman significativamente menos tiempo en ciertos lenguajes.

V SOFTWARE DISPONIBLE.

5.1 PAQUETES ESTADÍSTICOS CON SERIES DE TIEMPO.

Una vez habiendo visto el vínculo que existe entre los paquetes estadísticos y el uso de las microcomputadoras, y habiendo establecido algunos criterios para seleccionar un paquete para cierto propósito, se analizará ahora a cada paquete estadístico que contenga el análisis de Series de Tiempo, dándosele mayor énfasis a aquellos paquetes que contienen resultados y rutinas que permiten modelar mediante procesos ARIMA. Mencionando sólo aquellos que como alternativa tienen Análisis de Regresión o modelos de Suavizamiento Exponencial recordando los errores que pueden resultar utilizando este análisis con los supuestos de los modelos anteriores (como se vio en el capítulo 2.3.1). Esta información será útil de dos formas, una, ayudará a aclarar cuál de los paquetes mostrados podría ajustarse mejor a cada necesidad, y otra, como guía para poder llevar a cabo posteriormente un análisis de los nuevos paquetes estadísticos que surgan al mercado.

Es importante hacer notar que aunque se sabe que dentro de los criterios a considerar, es necesario tomar en cuenta el análisis numérico de cada procedimiento para conocer la exactitud del paquete, en este trabajo, este aspecto no será considerado como un criterio para la comparación de cada uno de los paquetes estadísticos disponibles. Esto es, debido a la imposibilidad de poder revisar las técnicas numéricas empleadas por cada paquete, ya que ninguno de ellos presenta documentación acerca de esto, además, no se cuenta con el código fuente de ninguno con el que se podría ver el algoritmo utilizado.

Antes que nada se efectuará un resumen de todos los paquetes estadísticos a los cuales se ha podido tener acceso dentro de la Universidad, con el fin de separar aquellos que

cuentan con análisis de Series de Tiempo con el modelado ARIMA de Box y Jenkins, y de no contenerlo, que es lo ofrecen como análisis opcional de Series de Tiempo (Análisis de Regresión o Suavizamiento Exponencial).

A continuación se enlistan todos los paquetes estadísticos que incluirá este trabajo:

NOMBRE DEL PAQUETE	VERSION
ABSTAT DOS	4.08
STATISTICAL PROCESSING SYSTEM	4.0
EPISSTAT	3.0
NWA STATPAK	3.1
STATEMENT/PLUS	1.0
STATPAC	1984
STATPRO	1.3
LSMLMW	1987
NUMBER CRUNCHER	1.0 y 2.0
SPSS/PC	1.0
SPSS	9.0
BASIS	1975
ESP	1.0
LIONHEART	2.0
STATGRAFH	1.0
TSP	4.0 y 4.1
FORECAST PLUS	
SYSTAT	3.0
RATS	1.2

Empezando entonces el análisis de cada uno de los paquetes y mencionando de aquellos paquetes que no contengan análisis de Series de Tiempo, (Modelos ARIMA) tan sólo, si cuentan o no con Análisis de Regresión o Suavizamiento Exponencial como alternativa (Teniendo el debido cuidado con los posibles errores).

Para aquellos paquetes que si contienen Series de Tiempo se hará un análisis más profundo, mostrando los resultados que se obtienen con cada uno de ellos al aplicarlos sobre un conjunto de datos específico.

Abstat.

Así, revisando el contenido de este paquete para el análisis estadístico, se puede ver que ABSTAT para realizar el análisis de Series de Tiempo sólo se cuenta con Regresión Simple y Regresión Múltiple (comando REGR).

Statistical Processing System.

En este paquete sucede lo mismo que con el anterior, no se cuenta más que con Análisis de Regresión Simple y Regresión Múltiple.

Epistat.

Epistat no cuenta con Análisis de Series de Tiempo. Listando las pruebas y funciones disponibles (opción 1 del menú principal) puede verse que sólo contiene Regresión Lineal, en el programa LINREGRES (10).

NWA_Statpak.

Al igual que en los paquetes anteriores no contiene Series de Tiempo, dentro del programa SPMENU en la operación de REGRESSION & CORRELATION puede verse que se cuenta con varios tipos de Análisis de Regresión. Y dentro de MATHEMATICAL UTILITIES tiene una opción para modelaje con suavizamiento.

Statmate/Plus.

Cuenta con Análisis de Regresión Lineal y Múltiple como puede verse utilizando el comando de ayuda, en su primera pantalla donde se muestran los posibles comandos que se pueden utilizar dentro de este paquete [4].

StatPac.

Al entrar al menú principal y escoger la opción de ANALYSIS (3) aparecerá un nuevo menú que tras seleccionar la opción de EXECUTE AN ANALYSIS mostrará el menú de todos los tipos de análisis que contiene. Aquí puede verse que sólo se contempla la Regresión Lineal Simple y la Regresión Múltiple para el interés de este trabajo [4].

Statpro.

En este paquete se puede contemplar mediante el uso de sus menús, que se tiene un módulo especial en el que se maneja el Análisis de Series de Tiempo, aunque por desgracia no existe el procedimiento para modelos ARIMA, y sólo se tienen promedios móviles, regresión y suavizamiento exponencial en varias de sus formas.

LSMLMW.

(Mixed Model Least-Squares and Maximum Likelihood Computer Program).

LSMLMW está diseñado según explica su manual [12] para manejar variables continuas y discretas independientes directamente. Los tipos de constantes que pueden calcularse por mínimos cuadrados o máxima verosimilitud son los

siguientes: Clasificación cruzada, Efectos anidados, interacción de dos factores de efectos de clasificación cruzada, Interacción de dos factores de efectos anidados, Interacción de dos factores de efectos de clasificación cruzada y anidados, combinaciones de regresión parcial para variables independientes continuas (lineal, lineal y cuadrática o lineal, cuadrática y cúbica) e Interacción de efectos de clasificación cruzada o anidados con variables continuas independientes.

El programa está diseñado para contemplar seis tipos de modelos mixtos (entre los que no se encuentran modelos de interés para este trabajo). En el manual del usuario puede encontrarse la descripción detallada de cada uno de estos seis tipos de modelos.

Number Cruncher.

En la sección de Series de Tiempo se pueden construir predicciones de series con tendencia o con tendencia y estacionalidad, o bien, sin tendencia. Pero sólo con la técnica de mínimos cuadrados, suavizamiento exponencial simple, doble suavizamiento exponencial, tendencia multiplicativa con suavizamiento estacional y tendencia aditiva con suavizamiento estacional [13].

SPSS/PC.

Para gran sorpresa y a pesar de su gran tamaño, este paquete en su versión para PC no contiene Análisis de Series de Tiempo, más que Análisis de Regresión, como en los casos anteriores.

Mediante la utilización del menú de ayuda HELP REGRESSION pudo verse las posibilidades para el Análisis de Regresión [18].

A continuación se tienen los menús principales y/o las funciones donde se muestran los alcances de cada uno de los paquetes arriba mencionados (exceptuando a LSM/LMW), en ellos puede verse que no existe el Análisis de Series de Tiempo mediante Modelos ARIMA.

Los menús se encuentran en el orden en que fueron mencionados los paquetes estadísticos.

Es necesario hacer notar que la información sobre los paquetes Statmate/Plus, Statpro, Statpac y SPSS/PC se obtuvo del libro "Paquetes Estadísticos para la familia IBM PC y Compatibles" de P. Seibold, L. O'keeffe y J. Klagge, debido a que no se cuentan con los diskettes disponibles.

Los demás paquetes mencionados en la lista contienen el Análisis de Series de Tiempo, por medio de modelos ARIMA.

Welcome to ABstat rel 4.08 02/27/85

Copyright 1984 Anderson-Bell Co.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

INSTALLED FOR ANSII TERMINAL

NOTE: This program is in DEMO mode so data sets are limited to 12 cases

The following commands will display names of commands available in each of the following categories. The ESCape key can be used to cancel any command. You may type ? for HELP at any prompt.

DATA - Data manipulation
STAT - Statistical analysis
GRAPH - Graphic functions
REPORT - Report writing
MISC - Miscellaneous: QUIT, HELP, COMM, MISS, etc.

WHICH COMMAND?

[25;01HS;01HS;02HTS;03HA5;04HTS;05H

Statistical commands:

ANOVA Up to 4-way analysis of variance w/wo replications
CHIFIT Chi Square goodness of fit and probability
CORR Correlation coefficients (r) matrix
DESC,DESC1,DESC2: Means, standard deviations, modes, etc.
FREQ Lists values, frequencies, %, and Z scores
KS2 Kolmogorov-Smirnov 2-sample test
KW Kruskal-Wallis 1-way ANOVA by ranks
MANN Mann-Whitney U test - U and probability
TIND t test for independent observations - "t" and probability
TPAIR t test for paired observations - "t" and probability
TPOP Population mean test - "t" and probability
PROB CHI, TTEST, FTEST, POIS, BIN, ZPROB probabilities
REGR Simple and multiple linear regressions (also ESTI & PROJ)
RUN1 1-sample runs test
RUN2 Wald-Wolfowitz runs test
SIGN Sign test
SRANK Spearman rank correlation matrix
XTAB Cross tabulation - chi square, prob, & Fishers exact test
ZSCOR Lists values, frequencies, %, and Z scores

WHICH COMMAND?

[25;01HS;01H

(c) 1983 Southeast Technical Associates: ALL RIGHTS RESERVED
Portions copyrighted by: DATA BASIC, Inc. of Mt. Pleasant, MI 48058
Authors: G.J. Buhoff, R.C. Kirk, H.M. Fauscher, R.B. Hull IV, E.E. McIenne

- 1 FILE CREATION AND MANIPULATION
 - 2 DESCRIPTIVE STATISTICS AND PLOTS
 - 3 PARAMETRIC AND NONPARAMETRIC CORRELATION
 - 4 REGRESSION ANALYSIS
 - 5 MEAN TESTS, T TESTS AND ANOVA
 - 6 TEST DISTRIBUTIONS: T, F, X²
 - 7 CROSS TABS: K X C CONTINGENCY TABLES
 - 8 MONTE CARLO DISTRIBUTIONS
 - 9 MULTIVARIATE ANALYSES
 - 10 UNIVAR, BIVAR, TRIVAR (QUALITATIVE ANALYSES)
 - 11 PREPARE LINK S.P.S. FILES
 - 12 EXIT TO BASIC INTERPRETER
 - 13 EXIT TO DISK OPERATING SYSTEM
 - 14 User Definable
 - 15 CHANGE PROGRAM DISKS
- SELECTION ? 4

Conversion FC4.0 by R. C. Kirk and G. J. Buhoff

S.P.S. GROUP # 4 REGRESSION ANALYSIS

SUB-GROUP CHOICES:

- 0 Return to main menu
 - 1 Simple linear regression models
 - 2 Multiple regression models
 - 3 General linear models
 - 4 Bartlett's test for homogeneity of variances
 - 5 Test for equality of regression models
- SELECTION ?

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

EPISTAT STATISTICAL PACKAGE

Would you like to:

- 1.) List the tests and functions available.
- 2.) Determine the best test for your data.
- 3.) Exit to run a specific test program.

Enter choice: 1

TESTS AVAILABLE IN EPISTAT

<u>TEST OR FUNCTION</u>	<u>PROGRAM NAME</u>	<u>TEST OR FUNCTION</u>	<u>PROGRAM NAME</u>
Analysis of variance.....	ANOVA*	McNemar's test.....	MCNEMAR
Bayes' theorem.....	BAYES	Mean, Median and S.Dev.....	DATA-ONE*
Binomial distribution.....	BINOMIAL	Normal distribution.....	NORMAL*
Chi-square test.....	CHISQR	Poisson distribution.....	POISSON
Correlation coefficients...	CORRELAT*	Random sample generator...	RANDOMIZ
Crosstab reports.....	XTAB*	Rank sum test.....	RANKTEST*
F Distribution.....	ANOVA*	Rates adjusted.....	RATEADJ*
Fisher's exact test.....	FISHERS	Sample size determination...	SAMPLESIZ
Graph histogram.....	HISTOGRM*	Select specific records...	SELECT*
Graph scattergram.....	SCATRGRM*	Signed rank test.....	RANKTEST*
Linear regression.....	LNREGRES*	Student's T test.....	T-TEST*
Mantel-Haenszel Chi-square..	MHCHISQR	Transfer data between	
Mantel-Haenszel for		two EPISTAT files.....	FILETRAN*
multiple controls/case...	MHCHIMLT*	Transfer data from	
		FORTRAN datafiles.....	FORTRANS*

*Starred programs can evaluate data entered and saved using DATA-ONE.
Press any key to return to main menu:

STATPAK 1.1
Copyright © 1982 Northwest Analytical, Inc.
Portland, Oregon

COLUMN Mode active. Report Date: 01/08/85

Enter: Program Name for CUR File Mode Menu
Program: SPMENU

STATPAK Main Menu

- A = File Utilities
- B = Mathematical Utilities
- C = Probability Distributions
- D = Single-Variable Statistics
- E = Distribution Functions
- F = Regression & Correlation
- G = Means Testing & Chi-square Analysis
- H = Non-Parametric Statistics
- I = Analysis of Variance

- S = SP (Statpak entry program)
- X = Exit to System

Enter selection letter (no CUR needed):

Regression & Correlation Menu

- A = AUTOCORR (Auto-Correlation)
- B = CORREL (Correlation & Covariance)
- C = CROSSCOR (Cross-Correlation)
- D = FOURIER (Fourier Analysis)
- E = MULTREG (Multiple-Linear, Stepwise, Interact.)
- F = UNIVREG (Single-Variable Regression)
- G = POLYREG (Polynomial Regression)
- R = Return to Main Menu
- S = SP (Statpak entry program)
- X = Exit to System

STATMATE/PLUS
Copyright (C), 1984
The Software Mill

ENTER ID-JEK

Command: HELP

---COMMANDS---

COMPUTE	CORRELATE	CROSSTAB	CURVE	EDIT	ERASE
EXECUTE	EXIT	GIVE	HELP	HISTOGRAM	INPUT
LET	NONWAY	ONEWAY	PLOT	POLYNOMIAL	PRINT
QUERY	RECODE	REGRESSION	REMARK	SET	STATISTICS
TTEST	TWOWAY				

PAUSE. HIT CARRIAGE RETURN TO CONTINUE.

STATFAC - STATISTICAL ANALYSIS PACKAGE
COPYRIGHT DAVID S. WALONICK, 1981, 1982, 1983, 1984

1. CODEBOOK
2. DATA FILE
3. ANALYSIS
4. UTILITY PROGRAMS
5. END PROGRAM

WHICH ?

* * ANALYSIS MANAGEMENT * *

1. CREATE A NEW ANALYSIS CONTROL FILE
2. EDIT AN EXISTING ANALYSIS CONTROL FILE
3. PRINT AN ANALYSIS CONTROL FILE
4. CHANGE THE ANALYSIS PARAMETER TABLE
5. EXECUTE AN ANALYSIS
6. RETURN TO MAIN MENU

WHICH ?

ANALYSIS TYPE :

0. UTILITIES (SRT LST TEL)
1. WRITE SUBFILE
2. FREQUENCY DISTRIBUTION
3. DESCRIPTIVE STATISTICS
4. CROSSTABS & LINEAR REGRESSION
5. CORRELATION & LINEAR REGRESSION
6. T-TEST
7. MULTIPLE REGRESSION
9. ANALYSIS OF VARIANCE
10. BREAKDOWN

WHICH ?

5.2 ANALISIS DE CADA PAQUETE.

Aunque este trabajo está enfocado a paquetes estadísticos disponibles para PC's, se mencionaran algunas características de dos paquetes disponibles para la Burroughs 7800 de la UNAM. Estos paquetes son BASIS y SPSS.

BASIS.

Es necesario especificar que en este paquete no se trabajó, y sólo fue revisado el manual que se encuentra disponible para las personas que así lo deseen en la DGSCA, UNAM [9].

BASIS, significa Burroughs Advanced Statistical Inquiry System, se encuentra escrito en ALGOL y dividido por secciones, dentro de las cuales se encuentra la de TSA que es en la que se realizan los análisis de Series de Tiempo.

Ofrece como análisis el de Promedios Móviles (MA) y el Suavizamiento Exponencial.

Para el caso de Promedios Móviles, se da a cada observación un peso igual, permitiendo además, escoger al usuario el número de términos a utilizar.

Da como resultado de un análisis de cierta serie, una gráfica de los datos originales de la serie, cada uno identificado por su fecha correspondiente (sólo se muestra por impresora). Se puede escoger entre un modelo de Promedios Móviles o de Suavizamiento Exponencial. El método seleccionado predice un valor para cada dato dando el error que se tiene.

Existe la opción de TERMS sólo para Promedios Móviles. Consiste en poder dar al programa un entero mayor que cero que representa el número de periodos para usarse en la predicción.

Como restricción se tiene que pueden procesarse hasta 40 variables.

Se especifica también en el manual lo que se hace con los datos faltantes, que dependiendo del caso se sustituyen por el promedio de los dos valores más cercanos al dato por cada uno de sus datos, o bien, se dejan fuera del análisis especificando esto en el reporte de salida, o también, especifica que el trabajo no puede realizarse.

SPSS.

En la versión de 1975 disponible en la B7800 de la UNAM no se tienen rutinas para el análisis de Series de Tiempo con modelos ARIMA, según Normand H. Nie en su libro "Statistical Package for the Social Sciences" [8] en la cual se tiene información hasta la versión número 8, no se tiene todavía dicho análisis. Por fortuna en la versión 9 como puede verificarse en el libro "SPSS Update 7-9 New

Procedures and Facilities for Realease 7-9" [7] del mismo autor [7], ya se cuenta con este análisis.

En este libro se muestra la utilización del paquete mediante un ejemplo (la serie G) del libro de Box y Jenkins "Time Series Analysis: Forecasting and Control" [1].

Como puede observarse en el ejemplo, puede obtenerse de SPSS lo siguiente para un análisis de Series de Tiempo.

- Una gráfica de la Serie.
- Una gráfica de la función de autocorrelación (Por default con la anterior).
- Un sumario estadístico(1) que contiene el nombre de la variable, la media y desviación estándar del proceso.
- Un sumario estadístico(2) que contiene el nombre de la variable, el número de observaciones, el grado de diferencia estacional y no estacional, la media y la desviación estándar del proceso.
- Una gráfica de autocorrelación parcial.
- Una gráfica de residuales.
- Los resultados de la estimación:
 - Estimación inicial de los parámetros (constante, valor de cada parámetro y varianza de los residuales)
 - Valores iniciales de los datos (tolerancia de convergencia, máximo de iteraciones)
 - Resultados estimados (para cada parámetro su nombre, estimación, error estándar, estadística T, además, varianza, media y desv. estándar de residuales, etc.)
 - Puede dar también gráficas de las series transformadas (logaritmos, diferencias).

En SPSS puede especificarse entre otras cosas:

El número de periodos para la gráfica de autocorrelación (Default 25).

El grado del modelo y cada uno de los parámetros que lo formaran.

La ausencia o presencia de la constante en el modelo.

El número de iteraciones (Default 40).

El número de predicciones hacia atrás de la última observación (Default 0).

La estimación inicial de los parámetros (de tenerse algún conocimiento).

El nivel de confianza de la predicción.

Todas las gráficas y resultados arriba mencionados pueden desplegarse por impresora.

Entre otros de los paquetes que si contienen el análisis de Series de Tiempo con modelos ARIMA, están los de ESP y Lionheart de los cuales no se tienen diskettes pero se pudo obtener información por medio del libro anteriormente mencionado de P. Seybold, L. O'keeffee y J. Klagge [4].

ESP.

Como dicen P. Seybold, L. O'keeffee y J. Klagge, las posibilidades estadísticas del ESP están basada en técnicas probabilísticas de Series de Tiempo. Por lo tanto, ESP no es un paquete estadístico de propósito general, sino que está específicamente diseñado para proporcionar técnicas de predicción, tanto simples como avanzadas, a los usuarios de computadoras personales.

La utilidad de este paquete se ve incrementada por su flexibilidad. Construido en torno a un lenguaje de órdenes, el ESP permite que el usuario construya multitud de fórmulas, y utiliza multitud de técnicas para intentar describir el pasado y predecir el futuro. Las posibilidades gráficas incluyen diagramas de dispersión, gráficos de residuos y gráficos de comparación de datos históricos reales con datos estimados.

Si se dispone de pantalla a color e impresora de cierta calidad, se puede apreciar mejor la estética de los gráficos del ESP.

La flexibilidad de las funciones estadísticas del ESP se refleja en sus previsiones en cuanto a entrada de datos y manejo de archivos, como especifican los autores del libro [4]. La entrada de datos se puede realizar por teclado, mediante un procesador de textos o mediante la lectura de archivos externos. Los archivos de LOTUS, la familia de programas VISI y de Microsoft, así otros archivos ASCII, se pueden leer con el ESP (ver anexo B de este trabajo). También se puede introducir como archivo de datos una sesión de trabajo (.WRS), así como ser salvada para un uso futuro. (Asegurandose de usar la extensión .WRS cuando se salve una sesión de trabajo no se perderá el archivo de datos .DAT) ESP proporciona varios niveles de ayuda para directorios. Estos directorios se pueden ver en su forma abreviada (sólo nombre del archivo) o en su forma detallada (nombres, observaciones, fecha de la última entrada, datos faltantes y si esta o no salvado).

ESP también facilita el requerimiento de las Series de Tiempo de introducir los datos en forma "periódica". Mediante la orden de muestreo (SMPL) se puede identificar el tipo de periodicidad de los datos que se vayan a introducir. La orden SMPL puede usar datos diarios, semanales, mensuales, trimestrales, anuales, especificados por el usuario, o bien, no periódicos. Si por ejemplo, se fueran a introducir datos trimestrales entre 1976 y 1984, se introduciría: SMPL 76Q1 84Q4. Entonces, el ESP irá pidiendo los datos para cada trimestre entre 1976 y 1984.

Según P. Seybold, L. O'keeffee y J. Klagge tres elementos del ESP proporcionan ayuda al usuario del programa, estos son, el procedimiento de ejemplos guiados, el procedimiento de ayuda y los mensajes de error. El procedimiento de ejemplos guiados se compone de ocho lecciones, comenzando con una sencilla sesión de trabajo, siguiendo con la función de datos y estadísticas para terminar con una lección de gestión de tareas repetitivas del ESP. El único aspecto curioso del procedimiento de ejemplos guiados es que la primera lección, "Getting Started" no explica al usuario los procedimientos de entrada de datos al ESP. Parece ser un fallo en el planteamiento de lecciones de los ejemplos guiados.

A P. Seybold, L. O'keeffee y J. Klagge las funciones de "ayuda", parecieron no proporcionar prácticamente ninguna ayuda debido a la escasa y poco profunda información que proporcionan. Los productores del ESP parecen haber encontrado una solución proporcionando una función de ayuda a dos niveles. Aunque ninguno de los dos niveles es sensible al contexto, juntos ayudan al usuario que momentáneamente haya tenido un lapsus de memoria. AL primer nivel se tiene acceso mediante la palabra HELP. Con ello se obtiene una pantalla que muestra todas las órdenes de manejo de datos y archivos. Si se mueve el cursor hasta una orden determinada y se pulsa la tecla RETURN, se verá una detallada descripción de dicha orden. La ayuda del segundo nivel se usa tecleando la palabra HELP seguida del nombre de un comando. Con ello se obtiene directamente la descripción de la orden especificada. La limitación de esta ayuda de segundo nivel es que el usuario debe conocer el nombre de la orden antes de poder obtener la información precisa. Se trata de una seria limitación, ya que a menudo es el nombre de la orden lo que el usuario necesita conocer.

Los mensajes de error del ESP parecen ser específicos al contexto, precedidos por una señal audible (un "bip") y aparejado en la parte inferior de la pantalla. Aunque estos mensajes parece que no siempre indican como corregir lo que se ha hecho mal, siempre indican cuál es el error cometido. Normalmente, una simple incursión por la función HELP más ORDEN será suficiente para descubrir el remedio.

Las órdenes BATCH/NOBATCH, los archivos RUN y los comandos DEFINE/CALL permiten un procesamiento multi-tarea. El proceso BATCH permite concatenar varias órdenes juntas para ser ejecutadas secuencialmente. Una vez que se activa la secuencia de órdenes no es precisa la intervención del usuario. Incluso aunque se den errores durante su ejecución, la secuencia continua hasta que se termina.

Los archivos RUN (de ejecución) están compuestos por órdenes del ESP introducidas secuencialmente en ellos y que se leen de un archivo de texto. Estos archivos, al igual que con la orden BATCH, permiten al usuario estar ausente durante la ejecución de las tareas. A diferencia de la orden

BATCH, los archivos RUN permiten la creación de subrutinas modulares que faciliten la "depuración".

Las órdenes DEFINE/CALL permiten la construcción de secuencias de tareas "comunes" como rutinas de análisis. La ventaja de estas órdenes es que permiten usar cualquier archivo de datos y cualquier variable que se especifique cuando se activa con la orden de llamada (CALL).

Este elemento presenta algunos problemas: hay que recordar el nombre dado a la secuencia de tareas que se quiere usar, hay que recordar el orden secuencial de los parámetros requeridos como "argumentos" cuando se hace la llamada CALL, así como que el número de variables en la secuencia de la orden DEFINE ha de ser el mismo que el número de variables que se usen en la orden CALL. Todas estas desventajas de las órdenes DEFINE/CALL se podrían eliminar si los "nombres" de las tareas aparecieran en un directorio y si el programa permitiera listar las secuencias DEFINE de las tareas junto con sus parámetros.

Los autores del análisis de este paquete, resumen diciendo que la fuerza del ESP reside en su potente colección de técnicas probabilísticas de Series de Tiempo, que son directamente accesibles mediante el flexible formato conducido por órdenes. Dicen que esta espléndida oferta se ve aún mejorada con las previsiones de indicaciones sensibles al contexto cuando se van complicando las subórdenes. ESP es muy específico, ya que sólo contiene Series de Tiempo.

Aunque la documentación parece ser muy buena en la sección que se refiere a las órdenes analíticas, se hechan de menos algunos ejemplos, problemas de muestra e información sobre las técnicas estadísticas usadas. También se dice que la presentación en dos volúmenes (Guía del Usuario y Manual de Referencia) es repetitiva y confusa. La documentación sobre las técnicas "avanzadas" no está en orden alfabético, ni en ningún orden que parezca lógico.

Para el Análisis de Series de Tiempo, además del procedimiento ARIMA se tiene Análisis de Regresión, sin embargo no se cuenta con Suavizamiento Exponencial.

LIONHEART.

De igual manera documentándose del libro arriba mencionado [4], se pudo ver que Lionheart ofrece, junto con los correspondientes discos, una serie de libros (no sólo manuales) con la intención de proporcionar un conocimiento de las técnicas estadísticas usadas. Los títulos de la serie son: Estadística Experimental, Estadística en los Negocios, Programación Lineal y No Lineal, Predicción y Series de Tiempo, Predicción Económica y de Negocios, Optimización, Técnicas de Camino Crítico y de Pert, y de Técnicas de Análisis de Decisiones.

Los libros (al parecer de P. Seybold, L. O'keeffee y J. Klagge) están completos, con prefacio y listado de

notaciones. El índice es generoso y contiene referencias cruzadas. Hay una única tabla de contenidos, desgraciadamente sin número de página. Sería útil disponer además de tablas de contenidos por capítulos. Se incluye una interesante lista de referencias a textos sobre los diferentes temas estadísticos tratados, y la última parte consiste en un glosario de términos de 16 páginas y ocho apéndices adicionales sobre temas especiales, como el manejo de sucesos fuera de rango y una descripción de cada uno de los programas para la computadora.

Una sección extra de páginas amarillas cubre los procedimientos básicos de carga y ejecución del programa.

El tono de explicación del libro según P. Seybold, L. O'keeffe y J. Klagge es sencillo (no condescendiente), práctico e informativo. El asunto es que el libro está generosamente animado con hechos interesantes y detalles que hacen que se reduzca la ansiedad del usuario.

La introducción de datos parece ser rápida y fácil (programa MREGDATA), ya que aparece una indicación que explica cómo han de ser introducidos los datos. Una vez que se ha creado un archivo de datos, se puede cambiar, transformar, ver o imprimir.

Lo que P. Seybold, L. O'keeffe y J. Klagge echaron de menos fue una aproximación al uso de los programas a base de ejemplos. Por otro lado, observaron los archivos de datos que vienen con los programas y obtuvieron una idea de los formatos de los archivos de datos.

El programa no incluye procedimientos de gráficas, aunque esto se considera un punto crucial para un programa. Mientras que, desde el punto de vista pedagógico, es deseable poder graficar los datos, en el mundo real de los ejecutivos y profesionales el acceso a un generador de gráficos es algo esencial.

En el paquete de Lionheart se incluyen varios métodos de Análisis de Series de Tiempo. Las formas más sencillas, media móvil y Suavizamiento Exponencial, son procedimientos que se encontraron menos flexibles en este paquete. La única participación del usuario que se permite al usuario en estas funciones viene dada por la especificación de "aproximaciones" simples o dobles, la selección del número de periodos y la determinación del parámetro de aproximación. Los métodos probabilísticos, regresión estacional y autorregresivo, son fáciles de usar y proporcionan un resultado estadístico adecuado. La única debilidad de estos modelos es la imposibilidad de incluir un término constante; por tanto, todas las ecuaciones de Series de Tiempo están forzadas desde el origen. El proceso más potente y flexible en cuanto a Series de Tiempo de las incluidas en el paquete es el método ARIMA. Sin embargo, la flexibilidad de este caso requiere que el usuario proporcione varias iteraciones de coeficientes estimados antes de llegar a un modelo satisfactorio.

Así, como dicen P. Seybold, L. O'keeffee y J. Klagge, la fuerza del producto de Lionheart reside en su explicación de las funciones estadísticas, en la facilidad de la entrada de datos (se puede ver, cambiar o añadir datos, y se pueden salvar los cambios en un nuevo archivo o en el antiguo), y en el buen formato de los resultados impresos.

Es buena la forma en que la serie de libros ha sido diseñada y terminada, permitiendo mezclar y ajustar las posibilidades estadísticas. Sin embargo, por otro lado, algunos de los programas contienen funciones solapadas.

Entre sus debilidades se incluyen la falta de gráficas. No se da ningún problema de ejemplo que guie paso a paso los procedimientos con los que se cuenta. Algunas órdenes no se entienden según se escriben; hay que introducir las en mayúsculas.

El éxito de su autor está en haber proporcionado programas estadísticos profesionales y útiles para no estadísticos, y se piensa que eso fué lo que se propuso. Lionheart ha hecho un buen trabajo teniendo en cuenta la audiencia a la que destina esta familia de programas: estadística experimental para investigadores, predicción y Series de Tiempo para académicos y programación lineal y estadística de negocios para los ejecutivos o los analistas.

Habiendo dejado al final el análisis de los paquetes para los que si se cuenta con el software y con los que si se trabajo, se procederá a revisar cada uno de acuerdo a los criterios establecidos en el capítulo IV.

MicroTSP.

Para este paquete producido por McGraw-Hill y que significa Procesador de Series de Tiempo, se cuentan con las versiones 4.0 y 4.1. Según P. Seybold, L. O'keeffe y J. Klagge [4] TSP fue diseñado por economistas, razón por la cual está enfocado al análisis de Series de Tiempo.

- Documentación del Usuario.

En cuanto a su interface con el usuario puede decirse que la documentación es adecuada [17], aunque la organización del manual es confusa. Por ejemplo, mientras que la información sobre el equipo requerido aparece en el capítulo 1, no se dan instrucciones sobre el modo de configurar microTSP en el sistema hasta el capítulo 3. En el capítulo 6, 8 y 11.4 aparece como título "Forecasting" (predicción), lo que hace confuso el tema que ahí se tratará. Por fortuna, las explicaciones en general, son dadas en forma muy detallada.

Por otro lado, el manual contiene ejemplos correspondientes a datos que aparecen en archivos suministrados con el programa. El manual no sólo guía a través de los procedimientos, sino también muestra listados de como aparecen los resultados en pantalla, ayudando a comprenderlos e interpretarlos.

El manual contiene al principio un índice con el contenido de cada uno de los capítulos, así como un índice alfabético al fin del manual. Cuenta también con una lista de los mensajes de error en la que se da una breve explicación de la causa de los mismos, y para algunos de ellos incluso dice como solucionar la condición de error. Además, tiene una sección donde se muestra en forma sintetizada el uso y la sintaxis de cada comando así como para algunos la página donde se encuentra el procedimiento que usa dicho comando.

Dentro del manual se mencionan con claridad la información general que va desde la versión de que se trata, los requerimientos de hardware: 256 K de memoria, al menos un drive y/o disco duro, un monitor monocromático o color (para gráficas un adaptador), una impresora y cuando menos el sistema operativo PC-DOS 2.0. El como debe instalarse, y hasta los nuevos y principales cambios que tiene esta versión con respecto a la anterior.

Aunque no pueden añadirse rutinas extras, pueden escribirse archivos BATCH que al constar de una serie de instrucciones reconocidas por TSP con parámetros opcionales permite la ejecución de ciertos procedimientos para

diferentes conjuntos de datos que tienen una misma secuencia de pasos. Este archivo puede ser escrito en cualquier procesador de textos, por lo que no existen tantas restricciones para su creación.

En particular para el procedimiento de Análisis de Series de Tiempo mediante modelos ARIMA se cuenta con un capítulo especial.

En dicho capítulo se menciona que la técnica utilizada es precisamente la de Box y Jenkins. Con un ejemplo va mostrando el procedimiento a seguir para obtener un modelo de predicción. Primero da una breve explicación de lo que son los modelos ARIMA y los parámetros que contendrá (autoregresivos, promedios móviles y diferencias). Después muestra como efectuar la identificación indicando el significado de las funciones de autocorrelación y autocorrelación parcial, además de como obtener las diferencias de una serie. Posteriormente muestra la forma de efectuar la estimación del modelo, explicando el procedimiento a seguir y los criterios que es necesario considerar; finalmente muestra una revisión del comportamiento de los residuales para un ejemplo específico, indicando brevemente que es lo que sería deseable que se obtuviera.

Para cada paso se da un ejemplo de como se obtendrá el resultado en la pantalla.

El procedimiento es muy flexible, debido a que la ecuación se debe dar en forma manual, por ejemplo:

```
LS <VARIABLE> C MA(1) MA(6) AR(2)
```

indica que el modelo contendrá una constante, un parámetro de promedios móviles no estacional de orden uno, otro de promedios móviles de orden seis y otro parámetro autoregresivo de orden dos. Por esto mismo, se espera del usuario cierto conocimiento de la técnica estadística.

Se muestran todas las funciones posibles dentro de este procedimiento por medio de ejemplos.

Se muestra además, con claridad en otro capítulo (11 del manual) el procedimiento a seguir en caso de existir datos faltantes así como el valor de default que se les asigna.

En el capítulo 8 del manual se muestran las referencias bibliográficas estadísticas de diversos procedimientos utilizados, entre los que se encuentran los modelos ARIMA.

La documentación sobre Series de Tiempo se encuentra independiente de las demás, lo cual es una ventaja. Cuenta además la versión 4.1 con un diskette de ejemplos guiados que explica de forma exacta varios de los procedimientos con el uso de los comandos y las respuestas que se obtienen de éstos.

Es importante hacer notar que la versión 4.0 no contiene la posibilidad de modelar con series estacionales, mientras que la 4.1 permite modelar los parámetros SMA y SAR

para promedios móviles y autoregresivos estacionales respectivamente.

- *Lenguaje de Control.*

Aunque no se cuenta con ayuda dentro de los procedimientos, los mensajes de petición de datos de cada comando y el panel que aparece en la parte inferior de la pantalla en forma permanente permiten que exista una muy buena interfase con el usuario. Existe además, un panel superior con información acerca de las variables que se están utilizando, el número de ellas (especificando un máximo de cien, entre las cuales estarán todas aquellas transformaciones que se efectúen a las variables dentro del paquete), el rango de tiempo que se ha especificado y el drive en el que se está trabajando.

En cuanto a limitaciones de tamaño, precisión, tiempo y costo de ejecución el diskette de ejemplos guaidos da los tiempos en segundos para cierto número de observaciones, si es máquina PC o AT, y para ciertos procesadores. Especifica también que correrá más rápido si se carga en disco duro, o bien, la máquina cuenta con más de 256k de memoria. Dentro del manual no se dice nada de esto.

Para el lenguaje de control se cuenta, como ya se mencionó, con una serie de menús asociados a las teclas de funciones, las cuales se encuentran divididas por la función que realizan y no por procedimientos.

Los comandos pueden ser introducidos en forma manual o utilizando las teclas de función.

Puede asignarse a las variables (por medio de la instrucción LABEL), cierta información como la fecha de la última alteración al archivo, frecuencia y rango de la serie, nombre y ciertos comentarios propios.

La transformación de variables se hace mediante la instrucción GENR, el formato es muy familiar para quienes han utilizado lenguajes de programación, se encuentra explicado de forma exacta en el manual. Por ejemplo:

$$\text{GENR RESULTADO} = (\text{LOG}(X) + X(-1)) * (Y^2)$$

que significa que en la variable RESULTADO se dejará el valor del logaritmo de la variable X de la observación actual, sumado con el valor de la observación inmediata anterior de la misma variable, y todo esto multiplicado por el cuadrado de la variable Y.

En general los comandos sugieren la función que realizan. Como por ejemplo, CREATE sirve para crear una serie, READ para leerla, PLOT para graficarla, GENR para generar una nueva, etc.

- Estructura de Datos.

TSP permite la importación y exportación de archivos (ver anexo B) de datos de tipo ASCII, DIF (y LOTUS para la versión 4.1), o bien, pueden producirse dentro de una sesión de TSP. Además permite la importación de archivos BATCH de cualquier procesador de textos como Wordstar por ejemplo, los cuales están compuestos de una serie de instrucciones reconocidas por TSP que pueden crearse por el usuario con el fin de agilizar un análisis.

La importación de los datos se hace preguntando al usuario: si los datos están ordenados por series, por observaciones, si es de tipo DIF, DIF invertido, si fue escrito en el editor de TSP (DATA) y para la versión 4.1 si es de LOTUS con las series por renglón o por columna.

- Impresión de Resultados.

La impresión de los resultados es clara, se encuentra especificada en forma estándar, las funciones de autocorrelación y autocorrelación parcial son mostradas en una sola gráfica por medio de la instrucción IDENT, dejando al lado derecho de esta gráfica los valores numéricos de las funciones.

La tabla de parámetros estimados contiene el número de iteraciones en las que se efectuó la estimación (máximo 20), el nombre de cada parámetro, su estimación, error estándar y estadística T. La media y desviación estándar de la serie, así como el error estándar y la suma de cuadrados de los residuales, la estadística Durbin-Watson. La R-cuadrada ajustada y la estadística F para la hipótesis de que todos los coeficientes, excepto la ordenada al origen, son cero. Existe en esta parte un pequeño problema, este es, que los valores de los parámetros estimados para MA y SMA son dejados con el signo invertido, habiendo tomado tal vez por los autores de TSP el modelo con los signos invertidos y no exactamente igual al del modelo propuesto por Box y Jenkins en su libro [1], aunque éstos últimos pusieron un signo negativo sólo por convención (ver sección 2.3.2 "Proceso de Promedios Móviles", de este trabajo).

Muestra como resultado de los procedimientos anteriores (IDENT, LS) el nombre del comando que los generó, el nombre de la serie utilizada, el número de observaciones y el rango de tiempo que cubre la serie (Dado con la instrucción SMPL).

- Gráficas.

En cuanto a los gráficos, se puede decir que son buenos, pueden añadirse etiquetas para clarificar los resultados así como elegir entre una serie de opciones para fijar las escalas: manual, automática, automática con entramado (líneas punteadas sobre la gráfica, que forman cuadrados) o residuales, por ejemplo. Pudiendo así,

escogerse los límites de escala, o dejando que TSP los ajuste automáticamente. Las gráficas pueden salvarse después de haber sido hechas con la instrucción GRAPH o PLOT por medio de la opción S y el nombre del archivo, o bien, imprimirse con la opción P.

Las gráficas de residuales las muestra en forma vertical llegando a ocupar el número de pantallas que sean necesarias, éstas son dadas con la ejecución de la instrucción LS junto con la tabla de parámetros estimados y la matriz de covarianzas.

- *Costo.*

En cuanto al número de diskettes, estas versiones de TSP constan de un sólo diskette y en el caso de la 4.1 viene otro diskette de ejemplos guiados adjunto, el cual vale la pena revisar.

TSP se encuentra dirigido a personas con cierto conocimiento de los procedimientos estadísticos, debido a que su gran flexibilidad hace que el usuario sea el que proporcione las fórmulas del modelo, así como la forma de efectuar las diferencias.

A continuación se muestran las principales instrucciones con las que puede efectuarse el análisis, dando para cada una de ellas su uso y lo que se obtiene del paquete como respuesta.

TSP *Este comando permite introducirse al paquete. Da información introductoria y muestra el panel de comandos en la parte inferior de la pantalla.*

El primer paso a seguir será el introducir los datos al paquete. Para ello, es indispensable definirla previamente:

CREATE *Permite definir el archivo preguntando para ello algunos datos sobre la serie. Primeramente, pregunta el tipo de periodicidad de la que se trata.*

U UNDATED
A ANNUAL
Q QUARTERLY
M MONTHLY

Como respuesta sólo espera la inicial de la opción deseada. Preguntando después, por la fecha inicial y final de la serie (si se contesta A, Q o M), o bien, el número de observaciones que contiene la serie (si se contesta U).

Existen diversos formatos en los que pueden escribirse las fechas: si es anual (A) el año 1981 o 2000 por ejemplo. Si la serie es del siglo XX será suficiente dar las últimas dos cifras. Para el caso de trimestres (Q) habrá que especificar el año y el trimestre de que se trata separados por un punto decimal, por ejemplo, 1981.1 para el primer trimestre de 1981 o 40.3 para el tercero de 1940. En el caso de que la serie sea mensual se tiene el mismo formato que para el caso en que sea trimestral (1881.7 para julio de 1881).

Una vez definida la serie los datos pueden introducirse de varias formas, una de ellas es el de introducirlos de un archivo ASCII previamente guardado bajo la extensión .DB (como se mostrará a continuación).

READ *Permite leer de un archivo en disco una o más series. Preguntará primero por el nombre del archivo (con extensión).*

<NOMBRE DEL ARCHIVO>

Al dar el nombre, TSP preguntará por la forma en que la serie o las series están ordenadas desplegando el siguiente menú.

S DATA ORDERED BY SERIES
 O DATA ORDERED BY OBSERVATION
 R LOTUS .PRN - SERIES IN ROWS
 C LOTUS .PRN - SERIES IN COLUMNS
 D DIF (DATA INTERCHANGE FORMAT)
 I INVERTED DIF
 H FORMAT BY HEADER FILE

Al teclear la opción deseada se preguntará el nombre de la serie o las series para uso interno en el paquete

<NOMBRE INTERNO DE LA(S) SERIE(S)> *Así se reconocerán para el análisis. Finalmente pedirá que se confirmen los datos.*

Y/N *Si están correctos teclear Y, de lo contrario al teclear N regresará a pedir los datos nuevamente.*

Ahora, si se deseara hechar un vistazo a la serie, TSP permite desplegarla en pantalla.

SHOW <NOMBRE SERIE> *Desplegará la serie nombrada por pantalla.*

Ya con los datos en el paquete puede efectuarse el primer paso que es el de identificación del modelo. Las herramientas utilizadas para este paso se muestran a continuación:

PLOT *Mostrará la gráfica de la serie que se especifique, preguntando antes por las opciones que se ofrecen en el menú.*

S SET OPTIONS
 A AUTOMATIC - SINGLE SCALE
 M MANUAL - SINGLE SCALE
 N NORMALIZED
 D DUAL SCALE - NO CROSSING
 X DUAL SCALE - LINES CROSS
 R RESIDUALS WITH S.E. BANDS

Al seleccionar (S) se podrá escoger entre cuatro tipos de gráficas que dependerán del número de series que se deseen graficar simultáneamente (a lo más seis), y de como se desea la gráfica, preguntará además si se desea que se muestren las leyendas, y finalmente regresará al menú anterior.

En el caso de seleccionar (M), pedirá los valores extremos deseados. Y en cuanto a la opción de (R), se obtendrá la gráfica de los residuales concernientes a la estimación más reciente.

Al seleccionar A, M, N, D, X o R se preguntará el nombre o nombres de la variable que se desea graficar.

<NOMBRE SERIE> *Desplegara en la pantalla la gráfica con las características deseadas, preguntando si la gráfica se desea salvar, imprimir o continuar.*

S,P,C *Salva, imprime o continua respectivamente.*

Para el cálculo de las funciones de autocorrelación y autocorrelación parcial se tiene una sola instrucción

IDENT(N) <NOMBRE SERIE> *Produce ambas funciones de la serie especificada para N periodos. Preguntará también si se desea salvar o imprimir.*

Para transformar la serie:

GENR <ECUACION> *Permite generar a una nueva variable por medio de una fórmula. Pudiendo utilizar operadores matemáticos (+, -, *, /, ^) y funciones (SQRT(), LOG(), EXP() o ABS()).*

Por ejemplo, cuando se desean hacer diferencias de una serie

GENR DIFERENCIA = SERIE - SERIE(-1)

utilizando el comando GENR, debe tenerse cuidado de que al efectuar la diferencia existan los datos requeridos en la serie, de lo contrario no podrá efectuarse la diferencia, por contener datos faltantes. Para fijar el periodo para el cual se desea efectuar la diferencia, se utiliza,

SMPL <PERIODOS> *La cual especifica cuales observaciones serán incluidas en las operaciones posteriores. SMPL 87.1 88.12.*

Con esto se tienen las herramientas mínimas para efectuar la identificación del modelo, ahora, para la estimación se tiene

LS <VAR> <PARAMETROS> *Permite la estimación de los parámetros, deben dársele el nombre de la variable y los parámetros que se deseen incluir en el modelo, así como el grado en que se desean.*

TSP preguntará si se quiere ver la matriz de covarianzas de los coeficientes estimados y la gráfica de residuales.

Se guarda además (por default), en una variable llamada RESID el valor de los residuales de la última estimación.

Una vez habiendo elegido el modelo de acuerdo a la serie mediante la utilización de los comandos anteriores, se deseará predecir valores de esta variable.

Normalmente con SMPL sólo puede modificarse el período utilizado dentro del rango que fue especificado cuando se creó la serie, sin embargo, para hacer la predicción (por ejemplo al futuro) es necesario contemplar períodos posteriores.

EXPAND

Expande el rango de la serie y tiene la misma sintaxis que el comando SMPL.

FORCST <PREDICCIONES>

Predice para la variable de la última estimación efectuada, el período especificado con el último SMPL efectuado, dejando los valores en la variable que acompaña al comando <PREDICCIONES>.

EXIT

Sale del paquete.

A continuación se presenta un ejemplo, mostrándose en forma sencilla el análisis de una serie tomada del libro de Box y Jenkins "Time Series Analysis, Forecasting and Control" que contiene datos del total de pasajeros (en miles) que viajaron en una aerolínea internacional de enero de 1949 a diciembre de 1960 (serie G). Los datos se encuentran listados al final de la tesis en el anexo A.

TSP puede guiarse a través de pequeños menús, esto es, para ciertos comandos CREATE, PLOT, READ por ejemplo, se muestra en pantalla un menú de opciones o preguntas con las que el paquete obtiene la información necesaria para procesar los datos de acuerdo al comando en cuestión. Para los usuarios más familiarizados con el paquete se tiene la posibilidad de dar para este tipo de comandos, todas las respuestas y opciones en un sólo renglón, o bien, el crear un programa BATCH que consta de un grupo de instrucciones familiares a TSP, las cuales pueden ser comandos como CREATE, READ, GENR, LS, etc. que al irse efectuando pueden producir la salida a impresora (mediante el comando PDN), por desgracia no puede incluirse el comando PLOT en éstos programas.

Para llamar al paquete se utiliza el comando TSP, una vez adentro el primer paso efectuado fue el de introducir la serie de datos. Se utilizó para esto, un programa BATCH, listado a continuación:

```
CREATE M 49.01 60.12
READ(D) A:SERIEG.DB SER
SHOW SER
```

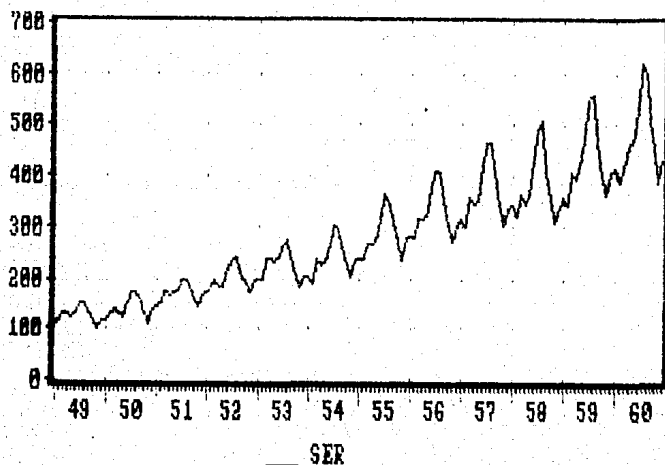
Con esta serie de comandos se logró crear una serie definida como mensual que va de enero de 1949 a diciembre de 1960 los cuales habían sido guardados en un archivo ASCII con extensión .DB ordenado por observación (un dato por renglón) y a la cual se llamará internamente SER, finalmente la serie es mostrada en pantalla.

Ya con la serie en TSP, pudo efectuarse el análisis. Para ver los datos graficados:

```
PLOT(A) SER
IDENT(20) SER
```

Con esta opción se obtuvo la gráfica en pantalla obteniéndose una gráfica como la mostrada a continuación al utilizar la opción P (Print).

GRAFICA DE LA SERIE



En la gráfica puede observarse una tendencia creciente, además, tiene un componente periódico consistente en un patrón estacional que ocurre aproximadamente en el verano de cada año. También se muestra una amplitud en la serie para los años más recientes. Estos factores deberán tomarse en cuenta para identificar el modelo.

A continuación se muestran también las funciones de autocorrelación y autocorrelación parcial para veinte periodos.

DNFL 1949.01 - 1960.12
 144 Observations
 IDENT SER

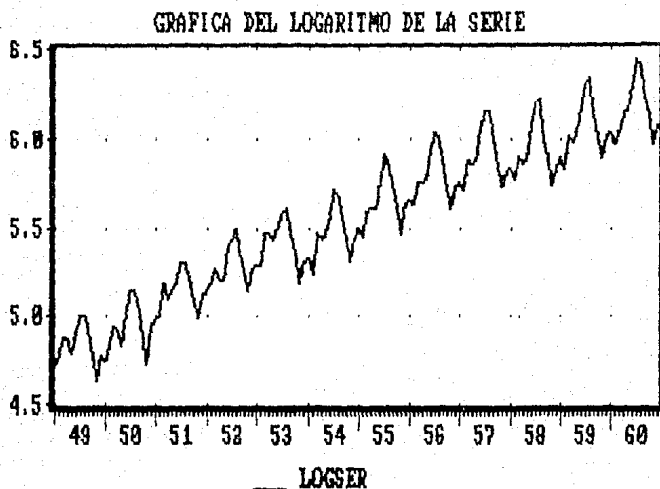
Autocorrelations		Partial Autocorrelations		ac	pac	
	*****	*****		1	0.9420	0.9420
	*****	**		2	0.8756	-0.2094
	*****			3	0.8067	0.0391
	*****			4	0.7329	0.0911
	*****			5	0.7103	0.0771
	*****			6	0.6817	0.0677
	*****			7	0.6627	0.1252
	*****			8	0.6552	0.0911
	*****			9	0.6700	0.2100
	*****	**		10	0.7007	0.1511
	*****	**		11	0.7472	0.1711
	*****	*		12	0.7501	-0.1024
	*****			13	0.7127	-0.5727
	*****			14	0.6463	-0.0111
	*****			15	0.6087	0.0911
	*****			16	0.5351	0.0111
	*****			17	0.4997	0.0711
	*****			18	0.4687	0.0711
	*****			19	0.4487	0.0424
	*****			20	0.4411	0.0424
=====						
S.E. of Correlations		9.000734E-02	0-Stat. (20 lags)	1313.500		
=====						

La no estacionaridad de la serie puede observarse también en la función de autocorrelación, para los primeros periodos la función es alta, decreciendo conforme avanza. Es sugerida una periodicidad de doce debido a que la función crece nuevamente en el periodo 12.

La amplitud puede removerse al transformar la serie obteniendo el logaritmo de cada observación.

GENR LOGSER = LOG(SER)
 PLOT(A) LOGSER

Dejando en la variable LOGSER el logaritmo de la serie SER y graficándola se obtuvo



Notese que la amplitud de la serie es ahora constante en el tiempo. Sin embargo, la serie transformada muestra no estacionaridad en forma anual y en forma creciente. Para remover estos factores, lo más conveniente será obtener las primeras diferencias estacionales y las primeras diferencias no estacionales.

```
SMPL 49.02 60.12
GENR DIF1 = LOGSER - LOGSER(-1)
SMPL 50.02 60.12
GENR DIF2 = DIF1 - DIF1(-12)
```

Así se obtuvieron las primeras diferencias no estacionales $y_t - y_{t-1}$ y después las primeras diferencias estacionales $y_t - y_{t-12}$ ajustando con SMPL para cada uno de

los casos evitando con esto los errores por valores faltantes.

IDENT(20) DIF2

Obteniendo las funciones de autocorrelación y autocorrelación parcial nuevamente para una diferencia estacional y una diferencia no estacional.

SMPL 1950.02 - 1960.12

111 Observations

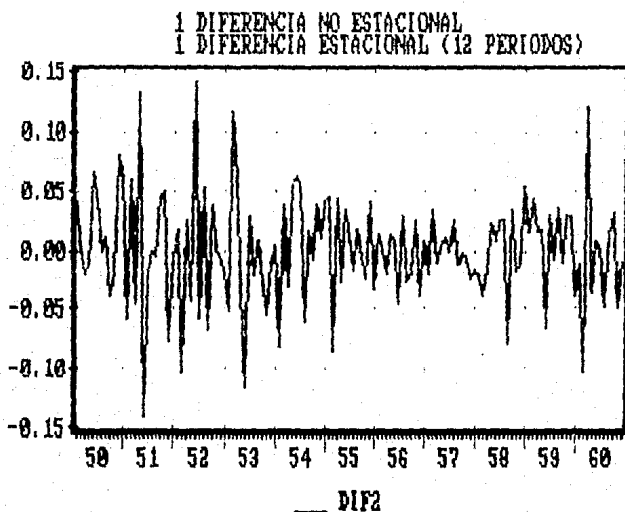
IDENT DIF2

Autocorrelations		Partial Autocorrelations		AC	PAC	
***		***		1	-0.2411	-0.2411
**		**		2	0.1650	-0.0135
*		*		3	-0.2021	-0.1311
				4	0.0214	-0.1250
				5	0.0557	-0.0211
				6	0.0305	0.0347
				7	-0.0552	-0.0601
				8	-0.0002	-0.0502
				9	0.1724	0.2290
				10	-0.0024	0.0471
				11	0.0214	0.0688
				12	0.2011	-0.1387
				13	0.1511	-0.1640
				14	-0.0571	-0.0711
				15	0.1495	-0.0112
				16	-0.1380	-0.1095
				17	0.0307	0.0207
				18	0.0152	0.1148
				19	-0.0102	-0.0132
				20	-0.1127	-0.1274
S.E. of Correlations		0.07204		Q-Stat (20 lags) 21.1027		

Para este caso se obtiene la menor derivación estándar, en comparación de las demás combinaciones entre diferencias estacionales y no estacionales (es decir, 1 dif. estacional y 0 no estacional, y viceversa). La función de autocorrelación de la serie con estas diferencias tiene picos significantes en los periodos uno y doce, lo que sugiere que no es aleatorio y es necesario modelarlo.

Para ver ahora la gráfica de la serie con diferencias.

PLOT(A) DIF2



La gráfica muestra que la estacionariedad ha sido obtenida. Observando ambas funciones (autocorrelación y autocorrelación parcial) se sugiere un modelo con un parámetro de promedios móviles y uno de promedios móviles estacional. Usualmente, claro, no es tan fácil distinguir el modelo a utilizar.

Habiendo terminado con la identificación del modelo se procederá con la estimación:

LS DIF2 MA(1) SMA(12)
IDENT(20)RESID

El modelo contendrá un término de promedios móviles, uno de promedios móviles estacionario y no contendrá constante. El reporte se obtiene de la siguiente forma:

```

SMPL 1950.01 - 1960.12
131 Observations
LG // Dependent Variable is DIFT
Convergence achieved after 10 iterations
=====
VARIABLE      COEFFICIENT      STD. ERROR      T-STAT.      D-FAIL SIG.
=====
MA(1)         -0.377496        0.0762111       -4.9557107    0.000
SMA(12)       -0.2500194      0.0997541       -2.5075400    0.000
=====
R-squared          0.364715      Mean of dependent var    1.000211
Adjusted R-squared 0.359215      S.E. of dependent var    0.349908
S.E. of regression 0.036672      Sum of squared resid     1.170537
Durbin-Watson stat 1.949644      Log likelihood            248.1547
=====

```

Finalmente se obtiene observando las funciones de autocorrelación y autocorrelación parcial de los residuos que TSP guarda en la variable RESID de la última estimación:

RMSE 1950.03 - 1950.17

121 Observations

IDENT RESID

Autocorrelations	Partial Autocorrelations	ac	pac	
		1	0.1201	0.1201
		2	0.0110	0.0100
*	*	3	-0.1230	-0.1235
**	**	4	-0.1484	-0.1457
**	**	5	0.0536	0.0624
**	**	6	0.0669	0.0577
*	*	7	-0.0767	-0.1312
*	*	8	-0.0697	-0.0612
*	*	9	0.0970	0.1413
*	*	10	-0.0750	-0.0587
**	**	11	0.0509	-0.0098
*	*	12	0.0226	0.0356
*	*	13	0.0477	0.0301
*	*	14	0.0352	-0.0174
*	*	15	0.0457	0.0440
**	**	16	-0.1502	-0.1146
*	*	17	0.0244	0.0157
*	*	18	0.0070	0.0073
*	*	19	-0.1087	-0.1295
**	**	20	-0.1069	-0.1647

S.E. of Correlations .0873704 G-Stat. (20 lags) 16.40616

Revisando estas gráficas (especialmente la de autocorrelación) puede verse que el modelo es adecuado, ya que cada punto es relativamente cercano a cero. Así el modelo final queda:

$$(dif^1)(DIF_{12}^1)(\log Z_t) = (1-3*B)(1-B^{12})a_t$$

Donde dif^m = diferencia de grado a no estacional.
 DIF_{nm}^m = diferencia de grado m estacional con n periodos.

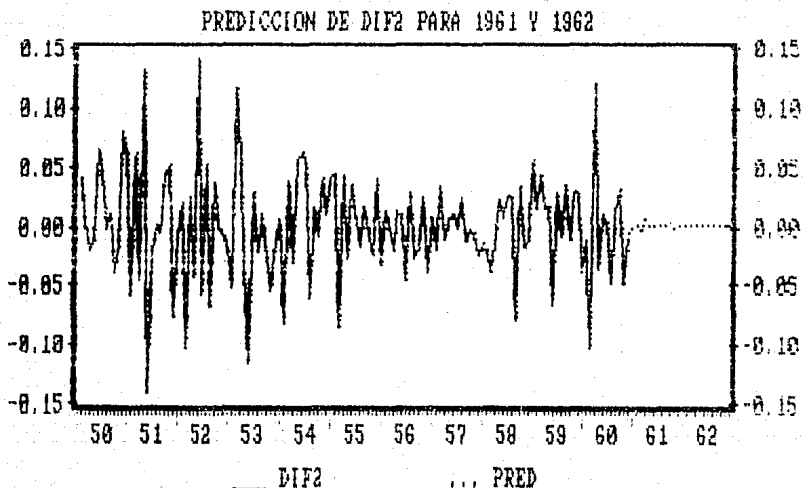
Una vez habiendo efectuado la estimación se va a predecir, para ello también se efectuó un programa BATCH:

EXPAND 61.01 62.12
 SMPL 61.01 62.12
 FORCST PRED
 SMPL 50.01 62.12

Con ello se pudo expandir y seleccionar el rango de periodos para los cuales se desea predecir (en este caso de enero de 1961 a diciembre de 1962), los valores de la predicción quedaron en la variable PRED y se fijo el rango para luego gráficar la serie.

PLOT(X) DIF2 PRED

Obteniéndose con ello la siguiente gráfica:



Los valores de la variable PRED pueden importarse a algún archivo ASCII para ser utilizados posteriormente. Finalmente para salir del paquete se utiliza la instrucción EXIT.

NOTA: Las gráficas mostradas, fueron modificadas con la opción de TYPE (T) que se puede utilizar al desplegar la gráfica en pantalla, permitiendo agregar leyendas acerca de los datos graficados. Se escogió dentro de las opciones (S, SET OPTIONS) la de B SOLID LINES & BOXES cuando se graficó una serie, y la de C SOLID & DASHED LINES cuando se graficaron dos series simultáneamente.

SYSTAT.

De este paquete se cuenta con la versión 3.0 en diskettes y con información de la 1.3 a través del libro "Paquetes Estadísticos para la familia IBM PC y Compatibles" [4]. Según este libro, la versión 1.3 de SYSTAT contiene sólo Regresión lineal y múltiple, por fortuna desde la versión 2.0 cuenta ya con modelos ARIMA, según puede comprobarse en el manual de dicha versión.

- Documentación del Usuario.

El manual de SYSTAT [19] que se revisó se encuentra bien ordenado, cuenta al principio con una tabla de contenidos, en el primer capítulo se relatan algunos de los procedimientos con los que cuentan y una serie de ejemplos guiados de como funcionan esos procedimientos, y aunque esta parte es bastante clara, ya que a cada paso muestra los posibles problemas y errores que pueden presentarse y la forma de remediarlos, supone que el usuario ya ha utilizado las versiones anteriores de SYSTAT y se refiere con frecuencia sólo a los cambios y mejoras con respecto a éstas. Para cada capítulo se cuenta con referencias bibliográficas acerca de la teoría estadística y acerca también de los algoritmos utilizados. Se cuenta además con cuatro apéndices: El primero es un sumario de la sintaxis de los comandos, divididos por módulos. El segundo contiene ejemplos elementales para introducir datos y de como correr algunos procedimientos estadísticos (no modelos ARIMA). El tercero, la estructura de los archivos de SYSTAT para que el usuario pueda escribir rutinas en FORTRAN que puedan ser accedidas fuera de SYSTAT. Y por último, explica el contenido del archivo DATA.DEF que es leído por el editor de SYSTAT para fijar ciertos atributos de este editor. Finalmente cuenta con un índice alfabético general.

En el manual no se menciona nada sobre la instalación del paquete, sin embargo en uno de los diskettes existe un programa (INSTALL.BAT) que dirige cada paso de la instalación para cuando se tiene un drive (A) y un disco duro (C). No se habla tampoco de los requerimientos mínimos de hardware, mencionándose sólo en un archivo (README) algunas limitaciones para los módulos, cuando sólo se tienen 256k de memoria y dando algunos suplementos de SYSTAT con los que pueden manejarse algunos de estos problemas. No se dice nada acerca de problemas de exactitud y precisión.

Se especifica además que no se cuenta con mensajes de error en el manual, pero los mensajes dentro del paquete son lo suficientemente claros y además en la sección de ejemplos guiados (cap. 1) se presentan algunos de ellos y el como resolverlos.

Para la sección de modelado de Series de Tiempo con modelos ARIMA (11, SERIES, del paquete) se tiene una introducción breve indicándose que la técnica utilizada es la de Box y Jenkins, muestra con un ejemplo (Precisamente la

serie G utilizada como ejemplo de este trabajo, del libro de Box y Jenkins) los pasos que se siguen para el análisis, sin ser demasiado explícito para la teoría estadística, muestra el cómo son desplegadas las salidas en la pantalla por SYSTAT, muestra también el funcionamiento de todos los posibles comandos de este módulo (los comandos comunes a todos los módulos son mostrados en otros capítulos del manual).

SYSTAT al efectuar el análisis (ARIMA) guarda en un archivo (RESIDUAL) el valor de los residuales, pudiendo guardarlos para posteriormente exportarlos. También si se utiliza la opción de FORECAST guarda en el mismo archivo las predicciones especificadas.

En este paquete se contempla la posibilidad de efectuar predicciones hacia atrás para tener una mejor aproximación en la función (BACKCAST), tal como lo proponen Box y Jenkins en su libro.

Especifica que es lo que se hace cuando existen valores faltantes, al copiar la variable o grupo de variables al área de trabajo los desaparece y al copiarlos nuevamente del área de trabajo a algún archivo de SYSTAT vuelve a colocar los valores faltantes en su lugar. Para este módulo los valores faltantes se ignoran. Sin embargo, pueden verse al graficar la serie.

Al final del capítulo se especifican las referencias bibliográficas de los algoritmos utilizados para residuales, sumas de cuadrados, etc. La estimación se hace con doble precisión y la predicción es dada en precisión sencilla. Además se cuenta con una lista de valores de default para cada uno de los parámetros de este módulo.

En cualquier momento se puede tener ayuda de parte del programa con sólo teclear HELP, o bien, HELP y el nombre del comando deseado, está última desplegará la sintaxis del comando y una breve explicación de lo que hace.

- Lenguaje de Control.

En cuanto al lenguaje de control, se tiene al principio del programa un menú general donde se encuentran el nombre de los módulos con un número asignado. Al entrar a alguno de los módulos (ya sea tecleando el nombre del módulo, o el número asignado) el menú desaparece, pero si se quiere saber los comandos utilizados en dicho módulo con teclear HELP se desplegarán en la pantalla. Aparece además al finalizar el módulo (QUIT) una lista de los comandos utilizados dentro de dicho módulo, ayudando con esto a seguir la pista de los resultados dudosos.

Las ordenes normales, como SAVE, USE o HELP, no tienen efecto hasta que se da una orden "fuerte" (HDT) como PLOT, RUN o ARIMA.

Pueden asignarse nombres a las variables, que a su vez pertenecen a un grupo que al utilizarse (USE GRUPO) se copiarán al área de trabajo.

La gran parte de los comandos describen su función, por lo que con un poco de experiencia pueden aprenderse realizando el análisis más rápidamente sin recurrir demasiado al manual.

La transformación de variables se hace mediante el módulo de datos (DATA), siendo fácil su manejo si se conoce el lenguaje BASIC, de lo contrario será necesario aprenderlo dentro de los primeros capítulos del manual.

Ciertos errores ocasionan la salida del programa al sistema DOS, sin embargo esto no es fatal ya que las variables utilizadas, pueden salvarse, así como las transformaciones realizadas durante la sesión.

- Estructura de Datos.

En cuanto a la estructura de datos, dentro del programa el concepto de archivo es algo diferente. Los datos pueden recolectarse de distintas fuentes (archivos ASCII, especificando que DBASE II o DBASE III pueden convertirlos a archivos ASCII (ver anexo B), directamente del teclado, o de Lotus cambiando la extensión .PRN por .DAT) ya que un módulo (DATA) los toma y los une en un archivo de SYSTAT. Sin embargo no puede listarse o editarse directamente por otros editores, a menos que sean exportados nuevamente a archivos ASCII.

Por desgracia, no es posible asignar una periodicidad al crear una serie, para SYSTAT todas las series contienen cierto número de observaciones sin saber si son datos anuales, mensuales, trimestrales, etc.

- Impresión de Resultados.

SYSTAT para la impresión de resultados no es muy adecuado, ya que cuando se desea imprimir algún resultado o gráfica se tiene que dar el comando OUTPUT @ antes de crear la gráfica o la tabla de resultados. Otro inconveniente es que ya no se podrán ver los resultados en pantalla, y para regresar el despliegue se tecleará la opción OUTPUT * con lo que ya no los mandará a impresora. Existe la posibilidad para algunas máquinas de utilizar CTRL P al mismo tiempo que permitirá ver los resultados en pantalla y en impresora, teclando nuevamente CTRL P dejará de imprimir.

Aunque la calidad de las gráficas y resultados no es muy buena, estos son claros y precisos. Por ejemplo, para las gráficas de los datos (PLOT), autocorrelaciones y autocorrelaciones parciales (ACF, PACF) se muestran el nombre de la serie, el número de casos, la media y desviación estándar de la serie, así como el tema de la gráfica. Sin embargo, para el reporte de la estimación de coeficientes sólo se muestran el número de iteraciones, con la suma de cuadrados y los valores de cada parámetro para cada iteración antes de la convergencia. Además una tabla que contiene para cada parámetro el nombre (AR, MA, SMA, SAR), la

estimación y el error estándar, sin especificar ni el nombre de la serie, ni el grado de la diferenciación.

- *Gráficas.*

Dentro de las gráficas las escalas pueden darse automáticamente o darse mediante un parámetro, así como puede decirse el número de periodos que se desea graficar (default 15). Por desgracia no pueden añadirse títulos o comentarios para mejorar su legibilidad.

Los gráficos también pueden guardarse para utilizarse posteriormente por medio de la previa instrucción SAVE.

- *Costo.*

SYSTAT cuenta con cuatro diskettes, a esto hay que añadir la necesidad de tener disco duro y para ciertos procesos una memoria mayor a 256k lo que hace que este paquete sea relativamente "costoso".

Este paquete da al usuario la oportunidad a cada paso de decidir si se quiere o no continuar, pudiendo salir del módulo en cualquier momento (QUIT).

A continuación se muestran los comandos más importantes con los cuales puede llevarse a cabo el análisis de Series de Tiempo. Para cada uno de ellos se muestra el uso y respuesta que se obtiene dentro del paquete.

SYSTAT

Con esta instrucción se entra al paquete donde se pide al usuario que pulse nuevamente RETURN para entrar al menú principal.

Ya en el menú principal, bastará con dar ya sea el nombre del módulo que se quiere, o bien, el número que lo acompaña. Para efectos de utilizar al análisis de Series de Tiempo sólo serán necesarios dos módulos, el de manejo de datos (DATA) que permite la importación, exportación y transformación de variables; y el del análisis (SERIES), que permite hacer diferencias, obtener logaritmos, graficar la serie, la función de autocorrelación y autocorrelación parcial, estimar y predecir.

Primero para la importación de datos:

DATA

Introduce al módulo de manejo de datos.

Para poder ver los posibles comandos utilizados dentro de cada módulo, estando dentro del que se desea, puede teclarse HELP, el cual desplegará los comandos permitidos en dicho módulo, pudiendo teclear HELP <comando> para que SYSTAT muestre el uso, la sintaxis y ejemplos del comando solicitado.

SYSTAT tiene diversos comandos llamados "débiles" que lo que hacen es quedarse en memoria hasta que se da un comando "fuerte" y se procesan en conjunto. Dentro del módulo de DATA existen ciertos comandos "débiles" necesarios para la importación, despliegue y transformación de los datos. Entre otros están:

GET <nombre del arch>

Permite leer los valores de las variables especificadas con el comando INPUT, el archivo ASCII deberá contener la extensión .DAT (este comando será ejecutado después de la instrucción RUN).

INPUT <vari> ...

Especifica el nombre o nombres internos que se les dará a las variables que podrán ser leídas del teclado o de un archivo ASCII. Puede especificarse un formato. (Necesita el comando RUN).

SAVE <archivo>

Salva los datos en un archivo de SYSTAT. Debe utilizarse antes del comando RUN si se desea crear el archivo.

USE <var1> ...

Lee las variables de un archivo de SYSTAT y es necesario siempre que se deseen utilizar las variables contenidas dentro de dicho archivo. Debe darse a la entrada de cada módulo, da como respuesta el nombre de las variables contenidas en el archivo, este comando no necesita de otro para ejecutarse.

LIST <var1> ...

Despliega la variable especificada, después de dar el comando RUN.

LET <var = expresión>

Permite transformar variables. Para lo cual pueden utilizarse los operadores +, -, *, / o INT, SQR, LOG, EXP, ABS, SIN, COS, TAN, etc. Es también necesario el comando RUN.

RUN

Es el comando "fuerte" que permitirá que se efectúen los comandos "débiles" anteriormente citados, no es necesario darlo después de cada comando, sino puede darse después de una combinación de ellos. Como resultado se obtendrán los respuestas de SYSTAT para cada uno de los comandos que se ejecuten.

QUIT

Sale al menú principal listando todos los comandos que fueron efectuados en este módulo.

Por desgracia, como ya se había mencionado, en ningún momento puede especificarse la periodicidad de la serie.

Para entrar ahora a efectuar el análisis, es necesario entrar al módulo de correspondiente:

SERIES

Introduce al módulo de Series de Tiempo.

Las instrucciones principales de este módulo son las a continuación mencionadas.

USE <var> ...

Puede usarse también en este módulo, leyendo así el archivo donde se encuentra la serie a analizar.

Para la identificación del modelo se utilizan:

PLOT <var>/LAG=p,STANDARDIZE,MIN=m,MAX=n

Grafica la variable var, pudiendo utilizar las opciones que permitan graficarla para p periodos (default 15), o estandarizarla, o dar el mínimo y máximo valor para poner la escala manualmente.

ACF <var>/LAG=p

Grafica la función de autocorrelación de la serie seleccionada para p periodos (default 15).

PACF <var>/LAG=p

Grafica la función de autocorrelación parcial de la serie seleccionada para p periodos (default 15).

Los comandos PLOT, ACF, PACF son "fuertes" por lo que desplegarán la gráfica de inmediato.

DIFFERENCE <var>/LAG=p

Obtiene las p diferencias estacionales de la serie, o bien, las primeras no estacionales si no se especifica la opción LAG=p. Ajusta la serie dejando los valores faltantes correspondientes. Es un comando "débil" necesitando de alguno "fuerte" para ejecutarse, PLOT por ejemplo. La respuesta es la leyenda de que la serie fue transformada.

LOG <var>

Obtiene el logaritmo de la serie sin necesidad de trasladarse al módulo de DATA, como respuesta se obtiene la confirmación de que la serie ha sido transformada.

Para la estimación y la predicción del modelo se tiene una sola instrucción:

ARIMA<var>/P=#,Q=#,PS=#,QS=#,SEASON=#,CONSTANT,ITER=#,TOL=#,BACKCAST=#,FORECAST=#

Con este comando se efectúa la estimación de los parámetros:

P para el grado del parámetro autoregresivo, Q el grado del de promedios móviles, PS el del parámetro autoregresivo estacional, QS el de promedios móviles estacionales, SEASON muestra la periodicidad de los parámetros estacionales, CONSTANT si se desea que el modelo contenga una constante. ITER permite especificar el número de iteraciones que se desean, TOL la tolerancia, BACKCAST permite calcular una mejor estimación de los parámetros ya que extiende la serie hacia atrás (predice hacia atrás) aproximando mejor la función. Finalmente la opción de FORECAST predice a la serie con el modelo estimado. Este es un comando fuerte.

Los comandos PLOT, LOG, ACF, PACF, DIFFERENCE y ARIMA no necesitan de otro comando para ser utilizados,

SYSTAT guarda después de cada estimación los residuales en la variable RESID, y después de cada predicción los de la variable son guardados en la variable FORECAST.

OUTPUT<SALIDA>

Direcciona la salida a monitor(), impresora(@) o a un archivo(nombre del archivo).*

QUIT

Sale al menú principal listando todos los comandos que fueron efectuados en este módulo.

QUIT

Al teclearlo por segunda vez saldrá al sistema operativo.

A continuación, utilizando el mismo conjunto de datos que para los demás paquetes analizados (serie G del libro de Box y Jenkins [1]) se procederá a efectuar un ejemplo.

```
SYSTAT
(Return)
1 o DATA
```

Con estos pasos se llegó al módulo de manejo de datos, en el cual se importa la serie de datos.

Introduciendo la serie que se va a utilizar:

```
GET A:SERIEG
INPUT SER
SAVE SERIES
RUN
```

Con esta serie de comandos se introdujo el archivo de datos SERIEG.DAT en la variable SER (para uso interno del paquete) y se salvo en un archivo de SYSTAT llamado SERIES, en el cual para este caso, sólo contiene la variable SER. Notese que los comandos se realizan hasta que se da la instrucción RUN.

Para ver si la serie se copio correctamente, puede ser desplegada en pantalla.

```
USE SERIES
LIST o LIST SER
RUN
```

Se llamó primeramente al área de trabajo el grupo de series (en este caso sólo SER), se dijo a SYSTAT que la desplegará y se dio el comando RUN para realizar los comandos anteriores.

```
QUIT
```

Pudiendo entonces salir del módulo de manejo de datos para iniciar el análisis.

Al regresar al menú principal se escogió la opción de SERIES, en la que se encuentra el análisis de Series de Tiempo con modelos ARIMA.

```
11 o SERIES
```

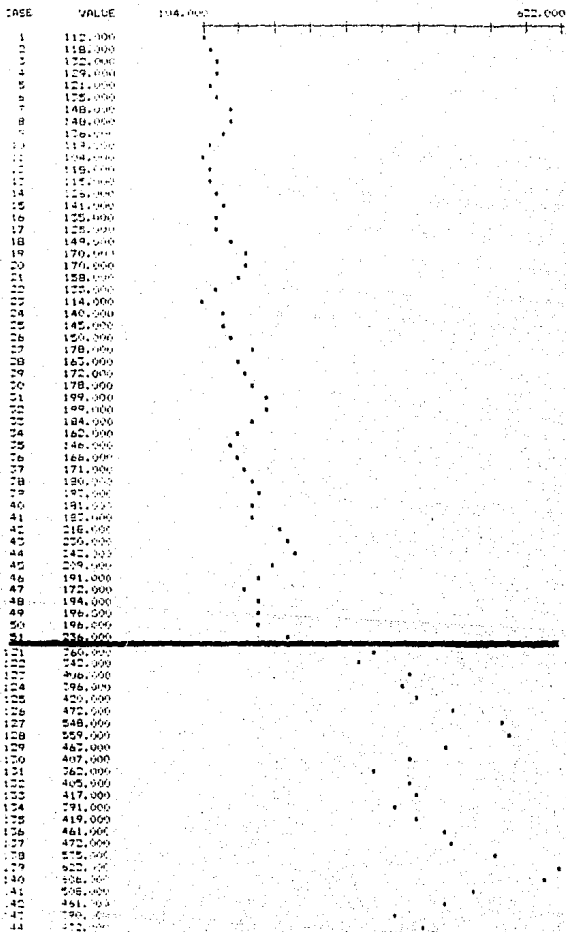
Ya en el módulo SERIES, se empieza el análisis:

```
USE SERIES
PLOT SER/LAG = 144
```

Primero se lee el archivo SERIES, de donde se grafica la variable que contiene la serie (SER) para las 144 observaciones, obteniéndose una como la que a continuación se muestra:

PLOT OF SER
 NUMBER OF CASES = 144
 MEAN OF SERIES = 200.299
 STANDARD DEVIATION OF SERIES = 119.549

SEQUENCE PLOT OF SERIES



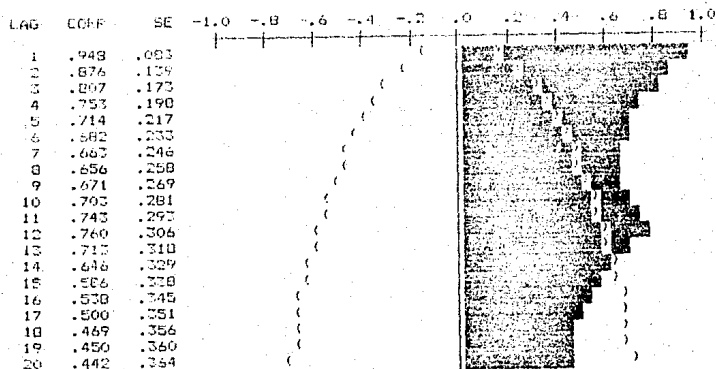
Como en la gráfica se observa una tendencia creciente y una componente periódica, además de una amplitud en la serie para los años más recientes, se deberá tomar en cuenta para la identificación del modelo.

Para obtener entonces la función de autocorrelación para 20 periodos de la serie.

ACF SER/LAG = 20

PLOT OF SER
 NUMBER OF CASES = 144
 MEAN OF SERIES = 280.277
 STANDARD DEVIATION OF SERIES = 119.549

PLOT OF AUTOCORRELATIONS



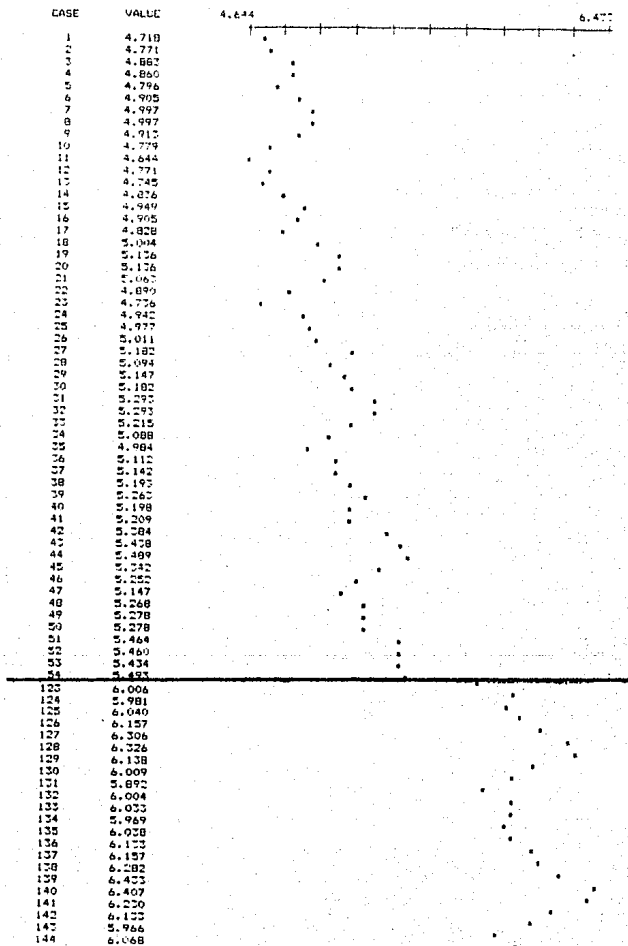
Se ve entonces que es necesario obtener logaritmos de la serie con el fin de remover la amplitud mostrada en la gráfica. Además, como la función de autocorrelación va decayendo y en el periodo 12 vuelve a crecer, sugiere una periodicidad de doce.

LOG SER
 PLOT SER/LAG=144

Habiendo obtenido el logaritmo de la serie se gráfico:

PLOT OF LOGSER
 NUMBER OF CASES = 144
 MEAN OF SERIES = 5.542
 STANDARD DEVIATION OF SERIES = 0.440

SEQUENCE PLOT OF SERIES



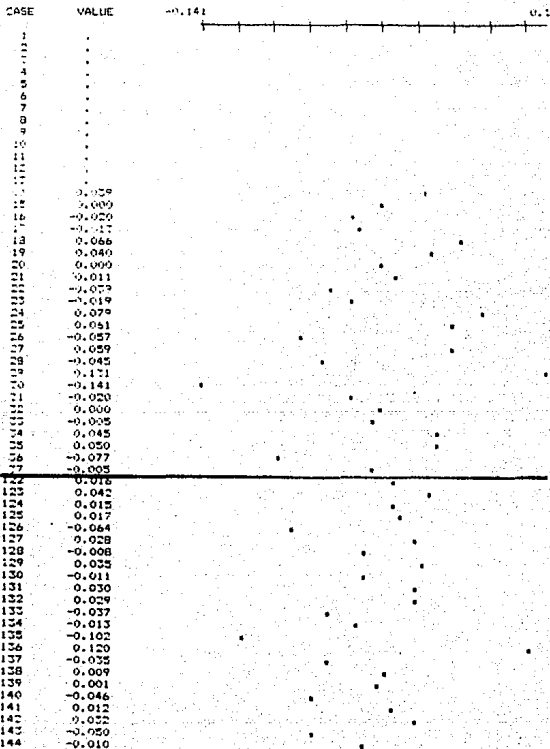
Ahora la amplitud de la serie es constante, sin embargo la serie sigue mostrando picos en forma anual y creciente, por lo cual, lo más conveniente es obtener primeras diferencias estacionales y no estacionales.

DIFFERENCE SER
DIFFERENCE SER/LAG=12
PLOT SER/LAG=144

Después de efectuar la primera diferencia no estacional y la primera estacional (de 12 periodos) se graficó para ver si se tenía o no estacionaridad.

PLOT OF LOSSER
NUMBER OF CASES = 144
MEAN OF SERIES = 0.000
STANDARD DEVIATION OF SERIES = 0.046

SEQUENCE PLOT OF SERIES

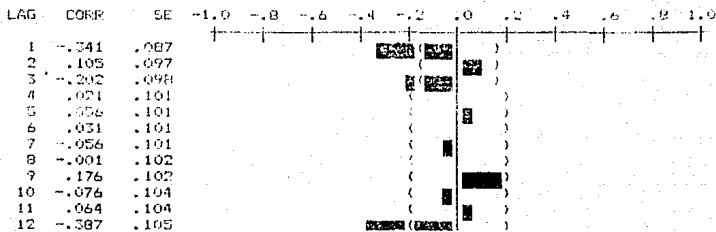


Mostrandose ahora que la estacionaridad ha sido obtenida se querrán ver las funciones de autocorrelación y autocorrelación parcial (se dan a continuación para 20 periodos).

ACF SER/LAG=20
PACF SER/LAG=20

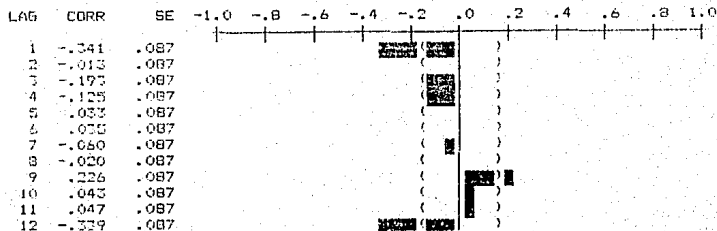
PLOT OF LOGSER
NUMBER OF CASES = 131
MEAN OF SERIES = 0.000
STANDARD DEVIATION OF SERIES = 0.046

PLOT OF AUTOCORRELATIONS



PLOT OF LOGSER
NUMBER OF CASES = 131
MEAN OF SERIES = 0.000
STANDARD DEVIATION OF SERIES = 0.046

PLOT OF PARTIAL AUTOCORRELATIONS



Observando ambas funciones se sugiere un modelo con un parámetro de promedios móviles y uno de promedios móviles estacional. Procediendo entonces a estimar los parámetros:

ARIMA SER/Q=1, QS=1, SEASON=12, BACKCAST=13

Que indica que se desea estimar a la serie SER con un modelo que contenga un parámetro de promedios móviles, uno de promedios móviles estacional y que el periodo será de 12. Además que se utilice la predicción hacia atrás para una mejor aproximación.

ITERATION	SUM OF SQUARES	PARAMETER VALUES	
0	.2392764D+00	.100	.100
1	.1835532D+00	.345	.407
2	.1764962D+00	.449	.603
3	.1759952D+00	.415	.592
4	.1758742D+00	.409	.613
5	.1758463D+00	.392	.614
6	.1758443D+00	.396	.613
7	.1758443D+00	.396	.613
8	.1758443D+00	.396	.613
9	.1758443D+00	.396	.613

ESTIMATED COEFFICIENTS

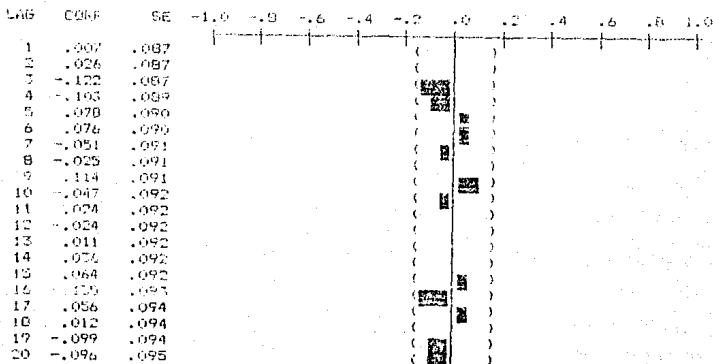
INDEX	TYPE	ESTIMATE	STANDARD ERROR
1	MA	0.396	0.100
2	SMA	0.613	0.073

Revisando ahora las funciones de autocorrelación y autocorrelación parcial para los residuales (SYSTAT los guarda automáticamente en la variable RESIDUAL):

ACF RESIDUAL/LAG=20
 PACF RESIDUAL/LAG=20

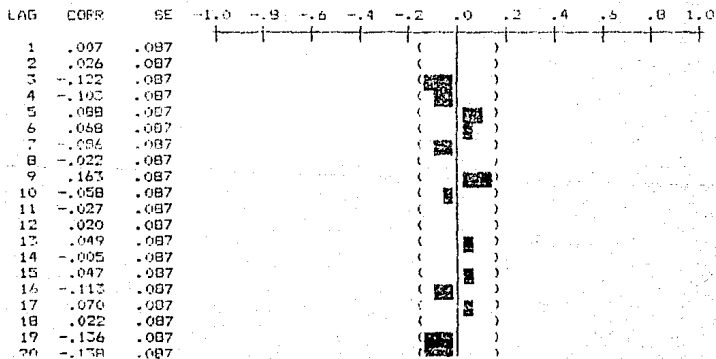
PLOT OF RESIDUAL
 NUMBER OF CASES = 131
 MEAN OF SERIES = 0.002
 STANDARD DEVIATION OF SERIES = 0.037

PLOT OF AUTOCORRELATIONS



PLOT OF RESIDUAL
 NUMBER OF CASES = 131
 MEAN OF SERIES = 0.002
 STANDARD DEVIATION OF SERIES = 0.037

PLOT OF PARTIAL AUTOCORRELATIONS



Revisando estas funciones puede verse que el modelo es adecuado ya que todos los puntos son relativamente cercanos a cero y no salen de las bandas.

Finalmente para predecir dos años, se tiene:

ARIMA SER/Q=1, QS=1, SEASON=12, BACKCAST=13, FORECAST=24

ITERATION	SUM OF SQUARES	PARAMETER VALUES	
0	.2392764D+00	.100	.100
1	.1835532D+00	.345	.433
2	.1764962D+00	.449	.633
3	.1759952D+00	.415	.592
4	.1758742D+00	.409	.613
5	.1758463D+00	.392	.614
6	.1758443D+00	.396	.613
7	.1758443D+00	.396	.613
8	.1758443D+00	.396	.613
9	.1758443D+00	.396	.613

ESTIMATED COEFFICIENTS

INDEX	TYPE	ESTIMATE	STANDARD ERROR
1	MA	0.396	0.100
2	SMA	0.613	0.073

FORECAST VALUES

PERIOD	LOWER95	FORECAST	UPPER95
145	418.853	450.287	484.079
146	391.989	426.571	464.205
147	438.336	482.107	530.249
148	447.306	493.257	546.613
149	453.912	508.358	569.455
150	516.602	583.416	658.872
151	587.758	668.282	750.486
152	581.067	666.011	763.373
153	484.235	559.829	644.914
154	427.675	495.729	577.010
155	368.094	430.257	502.918
156	406.841	478.177	563.301
157	415.840	496.469	592.720
158	389.995	470.321	567.193
159	436.573	521.553	647.197
160	441.716	542.744	666.978
161	452.195	560.404	694.729
162	514.626	643.253	804.028
163	584.737	726.923	928.465
164	572.313	734.318	942.868
165	451.504	616.144	788.434
166	424.874	547.686	705.977
167	365.394	474.385	615.887
168	403.280	527.222	689.255

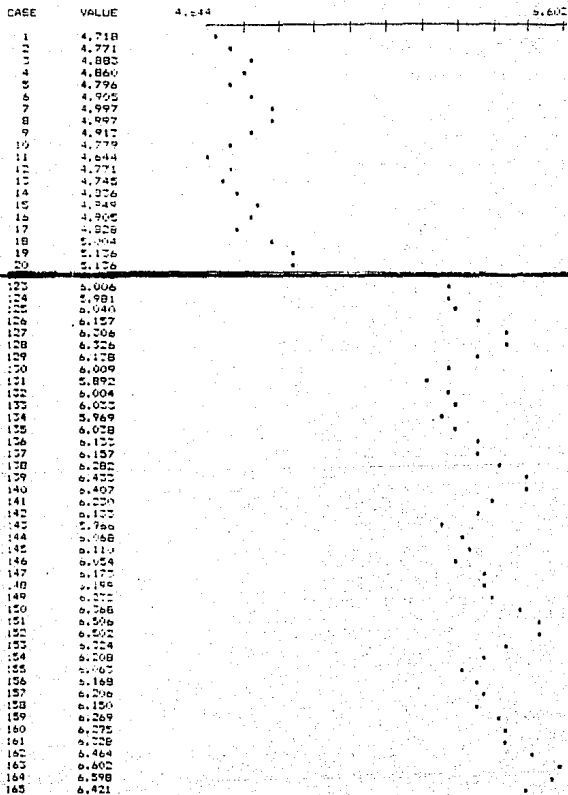
Con lo cual vuelve a estimar el modelo pero dejando ahora las predicciones en la variable FORCAST.

PLOT FORECAST/LAG = 165

Graficando las predicciones junto con la serie.

PLOT OF FORECAST
NUMBER OF CASES = 168
MEAN OF SERIES = 5.340
STANDARD DEVIATION OF SERIES = 0.489

SEQUENCE PLOT OF SERIES



QUIT Para salir del módulo.
QUIT Para salir del programa.

Notando que las predicciones son dadas para los logaritmos de la función y no para las diferencias de esos logaritmos, lo que da una gran ventaja, ya que para obtener los valores de la predicción no se tiene que transformar las diferencias y el logaritmo, sino sólo el logaritmo.

NOTA: Para obtener las gráficas por impresora es necesario utilizar los comandos de OUTPUT @ y OUTPUT * para activar y desactivar la salida, volviendo a repetir la instrucción del resultado que se desea imprimir.

STATGRAPHICS.

El siguiente paquete a analizar es el de STATGRAPHICS para el cuál se cuenta con un manual bastante extenso en comparación al de los paquetes anteriormente utilizados [16].

- Documentación del Usuario.

Este manual se encuentra distribuido de acuerdo a una tabla general, la cual se muestra en las páginas iniciales, y es precisamente la tercera pantalla al entrar al paquete, después de las dos primeras de presentación.

El manual se encuentra dividido en 10 partes: la primera de ellas contiene un tutorial que muestra la introducción a ciertos procedimientos de STATGRAPHICS. La segunda describe el cómo utilizar el paquete, así como los requerimientos de hardware, uso del teclado, etc. La tercera parte, describe los conceptos del manejo de datos y el ambiente del sistema en detalle. De la cuarta a la octava parte se dan los procedimientos gráficos y de análisis de datos. En la novena se muestra la utilización de STATGRAPHICS con el sistema APL*PLUS/PC (versión 4.0 o posterior) útil para escribir rutinas extras por el usuario. Finalmente en la última parte se encuentran los apéndices en los que se incluye un glosario, una lista con algunos mensajes de error con sugerencias para su corrección, una lista de las funciones de distribución utilizadas, otra con los archivos contenidos en los diskettes de STATGRAPHICS, referencias bibliográficas y un índice alfabético.

El procedimiento de ejemplos guiados muestra principalmente como introducir datos al paquete, aunque por desgracia entre los procedimientos que ejemplifica no se encuentra el de Análisis de Series de Tiempo con modelos ARIMA.

Dentro del manual, se tienen listados los requerimientos mínimos de hardware entre los que se tienen primero una computadora XT o AT compatible con IBM, 384 k de memoria, un teclado, dos drives o bien, un drive y un disco duro, un adaptador de gráficas, un monitor de color o monocromático y la versión 2.0 (o mayor) del sistema operativo MS-DOS. Opcionalmente se puede tener una impresora, un segundo monitor que permita desplegar en forma separada resultados y gráficas, y una mayor memoria que permitirá manejar archivos de mayor tamaño.

En el manual, se indica paso a paso el como se efectúa la instalación del paquete en caso de tener disco duro y el como pueden respaldarse los discos en caso de tener dos drives.

En la primera pantalla de presentación del paquete se dice que la versión de la que se trata es la 1.0.

Pueden añadirse procedimientos extras siempre y cuando el usuario se encuentre familiarizado con el lenguaje APL,

en el que se encuentra escrito STATGRAPHICS, y además cuente con la versión apropiada del lenguaje.

En cuanto a limitaciones de tamaño sólo se menciona que entre más memoria RAM se tenga se podrán tener archivos más grandes. Igualmente, acerca de tiempos aproximados en las corridas se menciona que dependiendo del procesador con el que se cuente se tendrá mayor o menor velocidad en los cálculos (sin especificar cuánto). No se dice nada acerca de la precisión.

Existe un procedimiento de ayuda para cualquier momento en la ejecución del programa, además es bastante interactivo debido a su manejo a base de menús y mensajes de requerimiento de datos.

En cuanto al procedimiento de modelos ARIMA (pag 23-26 del manual), se muestra muy brevemente el modelo utilizado con sus parámetros y se dice todo lo que puede realizarse con este módulo. Primero se muestran los requerimientos de entrada y la respuesta que deben darsele, además de la respuesta que se obtiene de parte del paquete, en este caso sólo es necesario el nombre de la variable que se va a utilizar. Después, por ser el análisis manejado por medio de una tabla, se muestra una explicación detallada de lo que hace cada parámetro dentro de dicha tabla, además el funcionamiento de las teclas de función (F1...F10) que van de acuerdo al panel de ayuda mostrado en la parte inferior de la pantalla.

Una gran ventaja que se tiene, es que sin tener que salir de este módulo puede efectuarse la identificación, estimación y predicción del modelo (teniendo por supuesto la serie ya introducida en STATGRAPHICS) pudiendo para ello graficar las series con o sin diferencias, estimar las funciones de autocorrelación y autocorrelación parcial con diferencias y sin ellas y teniendo además estas funciones para los residuales, posteriormente a haber efectuado la estimación de los parámetros y pudiendo con esto predecir los nuevos valores de las variables. Todo lo anterior por medio de las teclas de función, y llenando para la estimación del modelo un panel que contiene todos los parámetros posibles, así como criterios de paro, número máximo de iteraciones, número de periodos para las funciones de autocorrelación y autocorrelación parcial, etc.

Puede tenerse la opción de "Back Forecasting" (predicción hacia atrás) sugerida por Box y Jenkins en su libro (ver capítulo III, "Estimación hacia atrás Backforecast", de este trabajo), sin embargo no dice como se hace, ni los periodos para los que se hace.

Permite además salvar los valores de las funciones de autocorrelación y autocorrelación parcial, los valores predichos y los residuales en variables que pueden ser exportadas a diversos archivos posteriormente.

Muestra finalmente, con un ejemplo detallado el como se efectúa un análisis, mostrando listados de como se obtienen las gráficas y resultados desplegados por STATGRAPHICS.

Sin embargo, es importante hacer notar que aunque el manual dice que pueden modelarse parámetros de promedios móviles estacionales y autoregresivos estacionales, el paquete sólo permite hacerlo con autoregresivos (SAR) y para el caso de promedios móviles (SMA) despliega un mensaje de error aritmético y no muestra la estimación, además de regresar al usuario hasta el menú principal. Esto es una seria restricción, ya que será posible utilizar STATGRAPHICS sólo si se tiene la seguridad de que el modelo no contendrá este parámetro (siendo esto imposible antes de analizar la serie). En caso contrario, al menos STATGRAPHICS dará una forma rápida de identificar el modelo.

En cuanto a los datos faltantes STATGRAPHICS al leer en el archivo de datos un campo en blanco, asigna internamente el valor de -32768 con el fin de que al exportar las variables al final de la sesión se sustituyan nuevamente por blancos. En muchos procedimientos los casos en los que se tienen datos faltantes son excluidos, y en el caso de las gráficas, éstos no aparecen.

- Lenguaje de Control.

El manejo del paquete está llevado por medio de menús, así, sólo con poseer el cursor en la opción deseada y teclear RETURN puede llegarse hasta el módulo en el que se desea trabajar, ahí se efectúan preguntas referentes a las necesidades de ese módulo (nombre de la variable, grados de libertad, transformación deseada, etc.), además existe un panel en la parte inferior de la pantalla con los comandos que se refieren al módulo en el que se está situado, ayudando en gran manera a recordar el procedimiento.

En cuanto a las variables, puede asignarseles nombre, tipo, rango y longitud dentro del módulo de "Manejo de datos" (Data Management) en donde se encuentra contenido un directorio con todas las variables presentes en esa sesión de STATGRAPHICS.

Es importante hacer notar que el paquete distingue entre letras mayúsculas y minúsculas, es decir, se puede nombrar a una variable como "VAR" y otra como "var" y STATGRAPHICS las guardará como dos variables distintas.

Existe también un módulo donde puede efectuarse la transformación de variables teniendo un menú con ciertas transformaciones comunes, como raíz cuadrada, logaritmo, seno, coseno, etc., pudiendo también si se desea transformarlas algebraicamente.

- Estructura de Datos.

STATGRAPHICS permite exportar e importar datos de archivos ASCII, LOTUS, DIF (ver anexo B) o pueden teclearse directamente en el editor del paquete. Aunque para el caso de Series de Tiempo no permite especificar el período de los datos (anual, mensual, diario, etc.)

- *Impresión de Resultados.*

En cuanto a la impresión de resultados, esta es muy completa y de buena presentación, cuando el resultado es desplegado en pantalla este puede ser impreso mediante el uso simultáneo de las teclas SHIFT y PRN (o solamente PRN, con lo que la calidad de impresión es menor), esto convierte en opcional la impresión de cualquier resultado, permitiendo además el uso de graficadores.

STATGRAPHICS permite también cambiar el color del despliegue de resultados y gráficas a gusto del usuario, en caso de tener un monitor a color.

- *Gráficas.*

Para las gráficas se da una leyenda de su contenido sin dar el nombre de la serie que lo produjo. Para los resultados (estimación de parámetros, por ejemplo), se da para cada uno de los parámetros, el nombre, la estimación, el error estándar, el valor de la estadística T, además de algunas otras estadísticas, pero sin incluir el nombre de la variable, el procedimiento del cual se trata ni la fecha de proceso.

En las gráficas además de una gran calidad, se tienen etiquetas que pueden cambiarse, modificarse o incluso hasta añadirse, con el fin de facilitar su lectura. Automáticamente pone la escala, los residuales son mostrados en la misma escala que las observaciones. Se permite salvar las gráficas tal cual, y además permite guardar los datos que las produjeron. Permite cambiar su tamaño a lo largo y ancho para su salida a impresora.

- *Costo.*

STATGRAPHICS cuenta con cuatro diskettes, entre los cuales se tiene uno que es totalmente de ayuda.

Dentro de cualquier módulo de STATGRAPHICS puede "escaparse" al menú principal y dejar cualquier procedimiento por medio de la tecla F10.

En este paquete puede efectuarse con gran facilidad tanto la introducción de datos como el Análisis de Series de Tiempo con modelos ARIMA, sin embargo, de requerir un modelo con uno o más parámetros de promedios móviles estacionales, este paquete no será útil debido al error contenido en esta sección, esperando por supuesto que para una nueva versión del paquete ya se haya corregido este problema.

A continuación se hará una revisión de los módulos necesarios para efectuar el análisis de una serie.

Para iniciar la primera sesión de STATGRAPHICS es necesario correr el archivo STATGRAF.EXE que viene contenido en el primer diskette del paquete (SET UP).

STATGRAF *Como respuesta se obtiene el despliegue de la pantalla de inicio mostrando el siguiente mensaje:*

INICIALIZING. PLEASE BE PATIENT.
THIS WILL TAKE A FEW MOMENTS.

Después de algunos segundos preguntará por el número de monitores disponibles (1 o 2), por el tipo de monitor (color o monocromático), por los drives que se utilizarán para el sistema y para los datos. Despliega posteriormente una pantalla con todos los adaptadores que puede utilizar, pidiendo se elija la letra correspondiente al que se está utilizando. Finalmente STATGRAPHICS pregunta si se desean salvar estos datos para que la próxima vez que se entre al paquete se pueda dar la opción de configuración automática. Posteriormente se despliega una pantalla de inicio que pide se presione RETURN para continuar y con lo cual se obtiene una serie de mensajes informativos sobre el área de memoria disponible, en que drive se encuentra el sistema y en cual los datos, pidiendo nuevamente que se presione RETURN para llegar al menú principal.

Para entrar a cualquier módulo de STATGRAPHICS es necesario poseer el cursor sobre el módulo deseado y dar RETURN, o bien, teclear la letra correspondiente al menú (o la letra y el número del módulo) y presionar RETURN.

Revisando primero el módulo DATA MANAGEMENT que permite la importación de los datos y su manejo .

A. DATA MANAGEMENT *Al entrar a este módulo se obtendrá un submenú con las opciones que se tienen para el manejo de datos y las posibilidades de importación y exportación que se tienen.*

Para analizar la importación de archivos ASCII como se ha hecho con los paquetes anteriores se puede ver que en el submenú existe la opción de IMPORT DATA FROM ASCII DATA FILE la cual permite convertir archivos con o sin formato de tipo ASCII en archivos de STATGRAPHICS.

5. IMPORT DATA FROM ASCII DATA FILE

Se preguntará primero si se desea leer un archivo con o sin formato (FORMATED O UNFORMATED F/U).

Si se escoge un archivo formateado (F) se preguntará por el nombre del archivo y por la máxima longitud de los datos que contenga el archivo, pidiendo además un nombre para llamar a la variable internamente en el paquete. Si se escoge un archivo no formateado (U) preguntará por el nombre del archivo que deberá llevar la extensión .PRN, preguntará si el archivo contiene datos numéricos o no, pedirá el nombre de la variable en la cual se desea dejar los datos.

Se crea así una variable de STATGRAPHICS y regresa al submenú de DATA MANAGEMENT. Para llegar al menú inicial basta con presionar F10.

Para realizar el análisis de BOX y JENKINS existe una opción dentro del módulo TIME SERIES ANALYSIS llamada:

O. TIME SERIES ANALISYS.

14. BOX-JENKINS ARIMA MODELING

Al entrar lo primero que pregunta es el nombre de la variable que se va a analizar a lo que se contestará el nombre de la variable que contenga la serie.

Al contestar a esta pregunta aparece en pantalla un panel como el que a continuación se muestra:

- OUTPUT TIME SERIES *Contiene el nombre de la variable que se va analizar o alguna transformación de ella.*
- ORDER OF NONSEASONAL DIFF. *Contiene el orden de diferencia que se aplicará en la estimación del modelo (0,1,2).*
- CONSTANT CONTAINED IN MODEL *Si se quiere incluir en el modelo se dará YES de lo contrario se dará NO.*
- ORDER OF NONSEASONAL AR FACTOR *Contiene el orden del término autoregresivo (0 a 6). Especifica el número de parámetros estimados.*
- ORDER OF NONSEASONAL MA FACTOR *Contiene el orden del término de promedios móviles (0 a 6). Especifica el número de parámetros estimados.*
- ORDER OF SEASONAL DIFF. *Contiene el orden de diferencia que se aplicará en la estimación del modelo (0,1,2).*
- LENGTH OF SEASONALITY *Contiene la longitud de estacionalidad (0 si es no estacional, o mayor o igual a 2).*
- ORDER OF SEASONAL AR FACTOR *Contiene el orden del término autoregresivo estacional (0 a 6). Especifica el número de parámetros estimados.*

ORDER OF SEASONAL MA FACTOR

Contiene el orden del término autoregresivo estacional (0 a 6). Especifica el número de parámetros estimados.

MAXIMUN LAG FOR ACF PLOTS

Contiene el máximo de períodos que se desean para la función de autocorrelación.

LAGS FOR CHI-SQUARE TEST

Contiene el número de términos para la prueba de Chi-cuadrada en la autocorrelación de los residuales.

MAXIMUN LAG FOR PACF PLOTS

Contiene el máximo de períodos que se desean para la función de autocorrelación parcial.

NUMBER OF FORECAST DESIRED

Contiene el número de puntos que se desean predecir, una vez fijado el modelo.

BACKFORECASTING

Dar YES si se desea que se predigan hacia atrás valores adicionales en el desarrollo del modelo.

MAXIMUN ITERATIONS

Especifica el máximo número de iteraciones para la estimación del modelo.

STOPPING CRITERION 1

Se detendrá la estimación cuando el cuadrado de la suma de residuales entre dos iteraciones sea menor a lo indicado en este campo.

STOPPING CRITERION 2

Se detendrá la estimación cuando los parámetros estimados entre dos iteraciones sean menor a lo indicado en este campo.

Para actualizar el panel es necesario dar RETURN. Notese que en la parte inferior de la pantalla se muestran las funciones que se tienen para las teclas F1 a F10.

F1 (HELP)

Provee una pantalla de ayuda que explica el funcionamiento del módulo.

F2 (SERIES)

Grafica la serie, pregunta si se desea la serie original o con diferencias. Se pregunta el grado de diferencia estacional o no estacional que se desea. Al final se da RETURN para regresar al panel.

F3 (ACF)

Grafica la función de autocorrelación. Pregunta el grado de diferenciación (estacional y no estacional). Al terminar con la gráfica, y presionar RETURN se desplegará en la pantalla la estimación de los coeficientes y el error estándar. Al terminar la tabla y dar F10, el

F4 (PACF)

paquete preguntará el nombre de la variable en la que se desea guardar la estimación de las autocorrelaciones, si no se desean guardar, con presionar RETURN será suficiente.

Grafica la función de autocorrelación parcial. Pregunta el grado de diferenciación (estacional y no estacional). Al terminar con la gráfica, y presionar RETURN se desplegará en la pantalla la estimación de los coeficientes y el error estándar. Al terminar la tabla y dar F10, el paquete preguntará el nombre de la variable en la que se desea guardar la estimación de las autocorrelaciones parciales, si no se desean guardar, con presionar RETURN será suficiente.

F5 (ESTIM)

Procesa la estimación del modelo. Si se completa, STATGRAPHICS despliega una tabla con los parámetros del modelo estimado, errores estándar, estadística T, etc. Al terminar con teclar RETURN regresará al panel.

F6 (RACF)

Grafica la función de autocorrelación para los residuales, es necesario haber utilizado anteriormente F5 para fijar un modelo. Al terminar con la gráfica, y presionar RETURN se desplegará en la pantalla la estimación de los coeficientes y el error estándar. Al terminar la tabla y dar F10, el paquete preguntará el nombre de la variable en la que se desea guardar la estimación de las autocorrelaciones, si no se desean guardar, con presionar RETURN será suficiente.

F7 (RPACF)

Grafica la función de autocorrelación parcial para los residuales, es necesario haber utilizado anteriormente F5 para fijar un modelo. Al terminar con la gráfica, y presionar RETURN se desplegará en la pantalla la estimación de los coeficientes y el error estándar. Al terminar la tabla y dar F10, el paquete preguntará el nombre de la variable en la que se desea guardar la estimación de las autocorrelaciones parciales, si no se desean guardar,

con presionar RETURN será suficiente.

FB (INTPER)

Produce un periodograma integrado de los residuales del modelo. Es necesario haber fijado antes un modelo (con F5).

F9 (FORCST)

Grafica las predicciones usando el modelo estimado previamente con F5, además da los límites al 50% y 95% de confianza para dichas predicciones. Al terminar con la gráfica y dar RETURN se tiene la oportunidad de salvar los valores de las predicciones y sus límites en variables que pueden exportarse para usos posteriores.

F10

Sale del módulo al submenú de TIME SERIES ANALYSIS.

A continuación se mostrará el ejemplo utilizado con el procedimiento a seguir, mostrándose también el error encontrado, se utilizará la serie G del libro de Box y Jenkins, que contiene datos del total de pasajeros (en miles) que viajaron en una aerolínea internacional de enero de 1949 a diciembre de 1960. Los datos se encuentran listados al final de este trabajo en el anexo A.

AUTOEXEC

Llama al programa. Al preguntar si se desea configurar nuevamente los datos o están bien se contestó "Y" (ya está configurado), y con dos RETURN mostró la pantalla principal, después de dar información general.

Entrando a la opción para importar un archivo de datos ASCII, para introducir la serie al paquete se dio.

A. DATA MANAGEMENT 5. IMPORT DATA FROM ASCII DATA FILE

Al entrar al módulo se pregunta si el archivo que se desea importar tiene o no formato (en este caso no), se pide el nombre de la serie (.PRN), se pregunta si tiene o no datos no numéricos (en este caso no) y el nombre interno de la variable.

U
SERIEG
N
SER

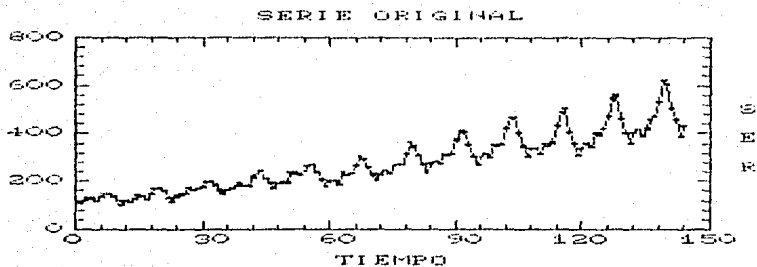
Después de esto se regresará al submenú de manejo de datos. Para regresar al menú principal se dio F10.
Empezando ahora el análisis:

D. TIME SERIES ANALYSIS 14. BOX-JENKINS ARIMA MODELING SER

Introduciéndose al módulo de análisis de Series de Tiempo por el método de Box y Jenkins, y dando el nombre de la variable en la que se encuentra la serie, se muestra el panel de control.

F2(SERIES)
0

Con esto puede verse una gráfica de los datos originales (sin diferencias), como la mostrada a continuación, para imprimirse se utilizó la tecla PRTSC.

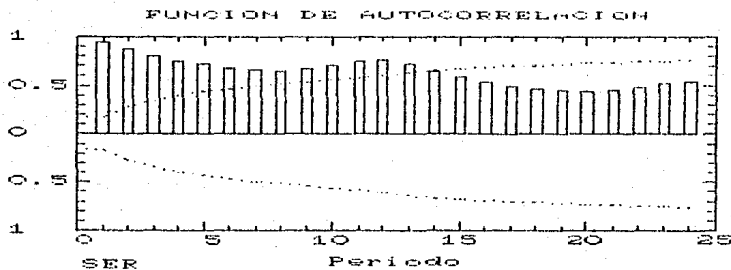


Como puede observarse en la gráfica, se tiene una tendencia creciente y un componente periódico que ocurre aproximadamente cada año. También existe una amplitud distinta para los años más recientes.

Observando la función de autocorrelación (de la serie original):

F3(ACF)

0



Lag	Estimate	Std.Error	Lag	Estimate	Std.Error
1	.94805	.05339	2	.87557	.11998
3	.80468	.17342	4	.75269	.19777
5	.71377	.21675	6	.69173	.23250
7	.66290	.24559	8	.65561	.25810
9	.67095	.26342	10	.70272	.28079
11	.74324	.29274	12	.76040	.30556
13	.71266	.21843	14	.64624	.32952
15	.58532	.33802	16	.53796	.34500
17	.49875	.35077	18	.46873	.35568
19	.44967	.35995	20	.44163	.36323
21	.45722	.36754	22	.46248	.37146
23	.51713	.37579	24	.53219	.38070

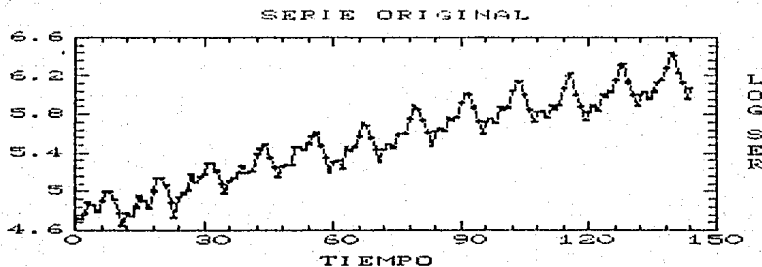
La función de autocorrelación muestra también la no estacionaridad de la serie, ya que para los primeros periodos la función es alta, además para el periodo doce la función vuelve a crecer, afirmando lo que se mostraba en la gráfica anterior.

La amplitud de la serie podría removerse si se obtiene el logaritmo de la serie. Para lo cual se modificó en el panel el nombre de SER por LOG SER y se pulsó RETURN para fijar el nuevo panel.

F2(SERIES)

D

Grificando ahora el logaritmo de la serie original.



En ella puede notarse que la amplitud que se mostraba anteriormente es ya uniforme. Sin embargo se sigue mostrando la no estacionaridad anual y de forma creciente, por lo que se sugiere obtener una diferencia estacional y una no estacional.

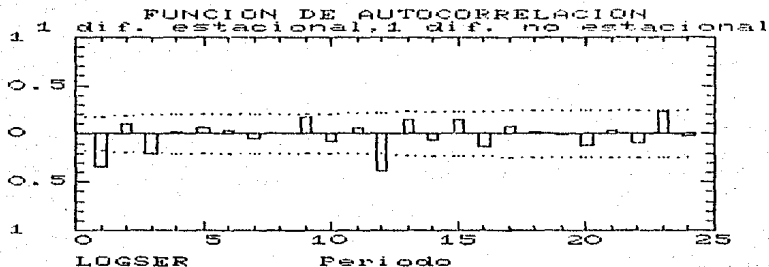
F3(ACF)

D

1

1

Muestra la función de autocorrelación del logaritmo de la serie con diferencias (una estacional y una no estacional).



Lag	Estimate	Std. Error	Lag	Estimate	Std. Error
1	-.34112	.08737	2	.10505	.09701
3	-.20214	.09787	4	.02136	.10101
5	.05555	.10104	6	.03080	.10128
7	-.08858	.10135	8	-.00076	.10158
9	.17637	.10168	10	-.07636	.10189
11	.06436	.10432	12	-.36661	.10462
13	.15160	.11501	14	-.05761	.11653
15	.14957	.11674	16	-.13994	.11820
17	.07048	.11944	18	-.01563	.11975
19	-.01001	.11977	20	-.11673	.11978
21	.03885	.12064	22	-.09136	.12074
23	.22327	.12126	24	-.01842	.12436

La función de autocorrelación muestra picos significantes en los periodos uno y doce lo que sugiere que es necesario modelarlo.

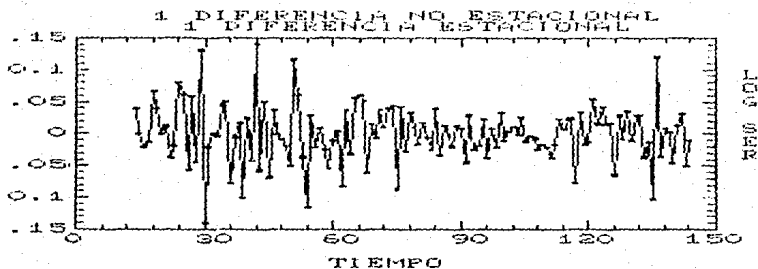
Es deseable ahora ver la gráfica de la serie con diferencias para tener una idea visual de la estacionariedad (una diferencia estacional y una no estacional).

F2(SERIES)

D

1

1



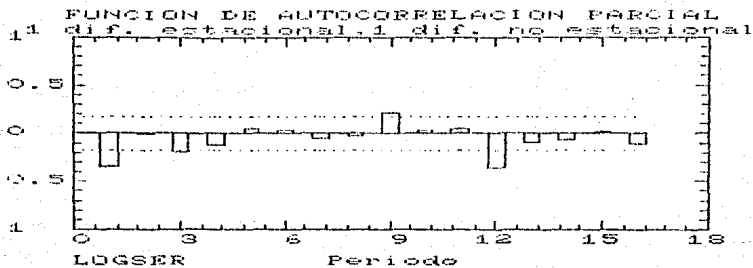
La gráfica muestra que la estacionariedad ha sido conseguida, vease ahora la función de autocorrelación parcial (una diferencia estacional y una no estacional).

F4(PACF)

D

1

1



Lag	Estimate	Std. Error	Lag	Estimate	Std. Error
1	-.34112	.08737	2	-.01201	.08737
3	-.19214	.08737	4	-.11622	.08737
5	.04198	.08737	6	.03897	.08737
7	-.05621	.08737	8	-.01737	.08737
9	.08207	.08737	10	.03006	.08737
11	.04294	.08737	12	-.35165	.08737
13	-.08798	.08737	14	-.05779	.08737
15	.21103	.08737	16	-.11298	.08737

Así, observando las funciones de autocorrelación y autocorrelación parcial sugieren un modelo que contenga un parámetro de promedios móviles y uno de promedios móviles estacional.

Modificando ahora el panel como se muestra a continuación para obtener el modelo con una diferencia estacional y una no estacional, con un parámetro de promedios móviles estacional y uno no estacional, sin constantes, estacionalidad anual (12 periodos) y los demás parámetros con sus valores de default :

 GENERAL UNIMODEL FROM MODEL FITTING

Output time series: LOG GEP

Order of nonseasonal diff.: 1	Order of seasonal diff.: 1
Constant contained in model: NO	Length of seasonality: 12
Order of nonseasonal AR factor: 0	Order of seasonal AR factor: 0
Order of nonseasonal MA factor: 1	Order of seasonal MA factor: 1
Maximum lag for ACF plots: 14	Maximum lag for PACF plots: 14
Lags for chi-square test: 20	Number of forecasts desired: 24
Backforecasting: NO	
Maximum iterations: 25	
Stopping criterion 1: .00010	
Stopping criterion 2: .00100	

PRESS ENTER TO UPDATE PANEL, THEN DECIDE BY KEY.

-HELP 2SERIES 3ACF 4FACF 5ESTIM 6SPACE 7EPCAF 8INTREP 9FORCAST 10QUIT
 INPUT SAT APR 9 1989 08:23:00 PM VERSION 1.0 904:0FF

Y es con esta forma (y por incluir el parámetro de promedios móviles estacionales) como despliega el siguiente error:

```
ERROR TRAPPED BY STATGRAPHICS ERROR HANDLING ROUTINES.  
TYPE OF ERROR: DOMAIN ERROR  
LOCATION: Line 15 of function @RESID  
PROBABLE CAUSE: Illegal mathematical operation.  
SUGGESTED ACTION: Check for division by zero or other illegal operation.
```

Press ENTER to return to main menu.

Hay que hacer notar que con los demás parámetros este módulo funciona bien.
NOTA: Las gráficas fueron corregidas con el fin de desplegar las leyendas en español facilitando su lectura.

RATS.

Aunque el manual [14] no especifica la versión de la que se trata el programa, indica a la entrada del paquete que se trata de la 1.2. Para esta versión, se cuenta con el análisis de Series de Tiempo por medio del procedimiento de Box y Jenkins.

- Documentación del Usuario.

Al inicio del manual, se tiene una tabla de contenidos que muestra los tópicos que se tienen disponibles y en una sección inicial de que capítulos son necesario revisar para realizar ciertos tipos de análisis, (en particular muestra los capítulos necesarios para efectuar un proceso ARIMA) incluyendo la entrada de datos. Se muestra además, en otra sección las convenciones de sintaxis para los comandos, esto es debido a que cada una de las secciones del manual está compuesta por los comandos necesarios para efectuar el análisis correspondiente a dicha sección.

Su manual no es muy claro, no se dice en ningún momento ni como instalarlo, ni el hardware necesario, ni tampoco el como entrar al paquete suponiendo que todo esto es conocido por el usuario.

Dentro de los archivos con los que cuenta el paquete se tienen el de RATS86.EXE y RATS8611.EXE con los cuales puede entrarse al paquete llamando a cualquiera de los dos sin existir aparentemente ninguna diferencia ni en la pantalla inicial, ni en su funcionamiento.

En el capítulo 1 del manual se muestran los comandos necesarios para introducir los datos, sin especificar claramente ni que tipo de archivos pueden importarse ni el procedimiento a seguir.

Cuenta al final de cada capítulo con un ejemplo que consta de una serie de instrucciones continuas (explicadas a lo largo del capítulo), sin explicar dentro del ejemplo detalladamente lo que se hace con cada instrucción o parámetro ni para todos existe la salida o respuesta que se obtiene del paquete. Para algunos casos se muestran gráficas o tablas de resultados obtenidos por el paquete.

Se tienen tres apéndices en los cuales se muestra el conjunto de caracteres manejados por RATS, nombres y variables reservados, una lista de los mensajes de error (divididos en seis tipos), sin decir como pueden evitarse o resolverse. Además se presenta una bibliografía estadística general y un índice de comandos con una breve descripción de éstos y la sección donde se explica su uso, así como donde se encuentra ejemplificado. Finalmente contiene un índice general ordenado alfabéticamente.

El manual esta dividido por temas en los que se explica cada comando necesario para poder desarrollar el proceso en cuestión.

En particular para la sección de modelos ARIMA se muestra en primer lugar, la ecuación utilizada y como

modelarla con comandos de RATS, no explica nada acerca del procedimiento que se debe seguir para modelar una serie de acuerdo a como lo plantean Box y Jenkins, aunque indica en que libro debe consultarse (Time Series Analysis, Forecasting and Control de Box y Jenkins o Forecasting Economic Time Series de Gringer y Newbold) para conocer la técnica.

Acerca de exactitud, tamaño, precisión y tiempo de ejecución no se menciona nada en el manual, ni tampoco se menciona nada sobre el procedimiento que se sigue de existir datos faltantes.

No se cuenta con ayuda dentro de las sesiones de RATS, sin embargo cuenta con un tutorial que muestra algunos de los procedimientos ofrecidos por el paquete, en ningún lado se indica como utilizarlo o llamarlo, siendo necesario teclear ya dentro del paquete SOURCE D:MENU.DAT (donde D es el drive donde se encuentre el diskette de demostración). Dentro del tutorial se muestra un menú bastante amigable (muy distinto al paquete), del cual puede escogerse la opción del procedimiento del que se desea la demostración, entre los que se tienen la introducción de datos al paquete y el modelado ARIMA, para este último se muestra el procedimiento desde la indentificación hasta la predicción de valores para la serie.

- Lenguaje de Control.

Tampoco se tiene un panel de opciones disponibles o comandos utilizables, lo que hace indispensable para el usuario inexperto tener a la mano el manual para cualquier duda, comando o interpretación de resultados.

Su manejo es através de comandos, para lo cual RATS muestra un símbolo "?" con el cual indica que esta listo para recibir el comando siguiente. Pueden asignarse etiquetas a las variables en el momento de su creación.

Para la transformación de variables se tiene el comando SET en la cual pueden utilizarse ciertos operadores similares a FORTRAN (indicados en el manual).

Para facilitar y agilizar los análisis pueden efectuarse procedimientos que contienen una serie de comandos de RATS.

- Estructura de Datos.

En cuanto a la importación y exportación de datos, se explica en una parte de la demostración (datos) que puede hacerse de archivos ASCII, LOTUS y DIF (ver anexo B), con una opción de la instrucción DATA.

- Impresión de Resultados.

Los resultados son mostrados en forma clara. Para la estimación, se muestra el número de ecuación que se utilizo, el nombre y periodo que tiene la variable, el número de

observaciones, y para cada parámetro el nombre, grado, valor estimado, error estándar y estadística T, además de algunas otras estadísticas interesantes.

La impresión de ellos se hace presionando (antes de dar el comando que produce el resultado que se quiere imprimir) al mismo tiempo las teclas de "CTRL" y "P" desplegándose así los resultados en pantalla e impresora, al finalizar la impresión es necesario dar las teclas anteriores para suprimir la impresión de resultados. Es necesario especificar que esto no se dice en el manual, ni en el tutorial, permitiendo solamente imprimir la serie de datos con la instrucción de COPY.

- Gráficas.

De las gráficas puede decirse que son claras aunque no tienen una excelente presentación, en los comandos de graficación se tienen varios parámetros que permiten efectuar cambios dentro de las gráficas, como por ejemplo, el máximo y mínimo valor para la gráfica, el símbolo con el que se quiere graficar, la fecha a la que pertenece cada observación, etc. permitiendo graficar hasta ocho series al mismo tiempo. Su impresión se hace de igual manera que con los resultados.

Al desplegar la gráfica (tanto datos como gráficas de autocorrelación y autocorrelación parcial) se dice la serie de la que se trata, el mayor y menor valor de la serie y la fecha a la cual pertenece cada observación. Las gráficas no pueden guardarse ni pueden ser modificadas.

Los datos de funciones de autocorrelación y autocorrelación parcial pueden guardarse y exportarse si así se desea.

- Costo.

El paquete cuenta con dos diskettes en los cuales se tienen diversos procedimientos para modelar Series de Tiempo pero por desgracia para el de modelos ARIMA no pueden utilizarse los parámetros estacionales de promedios móviles o autoregresivos (SMA o SAR), lo que implica una gran restricción.

A continuación se muestran los principales comandos con los que puede efectuarse el análisis, para ellos, se muestra su uso y la respuesta que se obtiene de ellos por parte del paquete.

RATS86 o RATS8611

Teclando cualquiera de estas dos formas se puede llamar al paquete, sin existir aparentemente ninguna diferencia. Mostrando autor, versión y dejando el "prompt" (?) para indicar que esta listo para recibir algún comando.

Para la introducción de los datos al paquete, teniendo un archivo ASCII puede efectuarse utilizando los siguientes comandos:

OPEN DATA <ARCHIVO>

Se utiliza para asociar a RATS unidades de I/O, en este caso será el nombre del archivo con extensión .DAT.

CALENDAR <AÑO> <PERIODO> <PER. I AÑO>

Define el tipo de serie que se utilizará dando el año en que comienza, el período de ese año (1er. trimestre, 3er mes o 1er año, por ejemplo) y el número de períodos por año.

El formato para éstos parámetros es el siguiente: Para el año pueden darse las últimas dos cifras si la serie está situada en el siglo XX, o bien, las cuatro cifras. El siguiente parámetro deberá ser el número del período correspondiente a la primera observación (sin dar el año). En período por año se especifica 1 si es anual, 4 trimestral, 12 mensual, etc.

ALLOCATE <N><FIN SERIE>

Define el tamaño de la serie, dando el período donde termina la serie, pudiendo "etiquetar" la serie con un número (n), pudiendo ir en cero.

El período final de la serie deberá darse por el año y el período en que finaliza la serie separados por una coma, por ejemplo, 1968,12 para decir que la última observación de la serie es la de diciembre de 1968.

DATA<P.INICIO><P.FIN><NOMBRE>

Lee la serie asociada en el comando OPEN DATA asignando los valores de la serie a partir del período <p.inicio> al período <p.fin> de acuerdo a lo que se definió en los parámetros CALENDAR y ALLOCATE.

Dando tambien un nombre interno a la serie.

El periodo inicial y final deberá darse con el año y el periodo correspondiente, separado por una coma.

PRINT(<OPCIONES>) <P. INICIO><P. FIN><NOMBRE>

Despliega la serie en pantalla, desde el periodo inicial hasta el final especificado (con el formato año, coma, periodo) de la serie <nombre>; pudiendo dar entre las opciones, la de DATES con la que se desplegará al lado de cada valor el periodo al que pertenece.

Para el análisis de la serie se tiene.

PLOT(<OPCIONES>) <N>
<NOMBRE><P. INICIO><P. FIN>
<SIMBOL0>

Produce en pantalla la gráfica de n series (máximo 8), debiendose asignar para cada una de ellas el nombre de la serie deseada, y los periodos inicial y final que se desean graficar. Pudiendo dar las opciones de DATES con la cual, al lado de la gráfica se dará el periodo correspondiente a cada observación, y la de CHARS con la que podrá escogerse un caracter específico para marcar cada punto en la gráfica (por "default" se grafican con "A", "B", "C"... las series 1,2,3 ... respectivamente).

A falta de alguna indicación en el manual para imprimir cualquier resultado es necesario antes de dar el comando teclear CTRL P.

CORRELATE<NOMBRE><P. INICIO><P. FIN><N><NOMBRE-2>

Permite desplegar los valores de la función de autocorrelación para la serie <nombre> desde el periodo inicial hasta el periodo final indicado, para n periodos, dejando los valores en la serie <nombre-2> para poderse graficar posteriormente si se desea.

PARTIAL<NOMBRE><P. INICIO><P. FIN><N><NOMBRE-2>

Permite desplegar los valores de la función de autocorrelación parcial para la serie <nombre> desde el periodo inicial hasta el periodo

final indicado, para n períodos, dejando los valores en la serie <nombre-2> para poderse graficar posteriormente si se desea.

SET <RESULTADO><P. INICIO><P. FIN>=<ECUACION>

Deja en la variable <resultado> del periodo inicial al final especificados el resultado de la ecuación.

DIFFERENCE <NOMBRE><P. INICIO><P. FIN><N><NOMBRE-2>

Obtiene las n diferencias no estacionales de la serie <nombre> en el periodo especificado dejando el resultado en la serie <nombre-2>. No permite las diferencias estacionales.

EQUATION <N><NOMBRE><AR><MA>

Define el modelo que se va a estimar da a la ecuación un número n con la cual se identificará, deberá darse la serie sobre la que se hará la estimación y el grado de parametros autoregresivos y de promedios móviles no estacionarios que se desean.

INITIAL <N>

Da las estimaciones iniciales de los parámetros definidos por la ecuación n especificada por el comando EQUATION.

ITERATE (<OPCIONES>) <N><P. INICIO><P. FIN><NOMBRE>

Calcula la estimación final de los parámetros mostrando en una tabla los valores de cada uno con sus errores estándar y estadísticas T , además de otros datos interesantes. Calcula la estimación de la ecuación n previamente definida sobre el periodo especificado y dejando los residuales en la variable <nombre>. Entre las opciones se tiene la de backforecasting que se obtiene poniendo la opción de STARTUP=(aa,pp) donde aa es el año en que se desea se comience a predecir valores iniciales para las iteraciones y pp es el periodo correspondiente.

Para la predicción se tiene el siguiente comando.

FORECAST <N><P. PRED><INICIO PRED>

Produce las predicciones utilizando la ecuación n ya estimada, para el número de periodos especificados por <p.pred> iniciando la predicción en el periodo <inicio pred>.

END

Permite la salida del programa.

NOTA: Todos los parámetros de <p.inicio>, <p.fin> e <inicio pred> tienen el formato de año y período correspondiente separados por una coma, por ejemplo 68,12 para indicar que se trata de diciembre del año 1968.

A continuación se muestra el ejemplo que se presenta en el libro de Box y Jenkins "Time Series Analysis, Forecasting and Control" (serie G) [1] que contiene datos del total de pasajeros (en miles) que viajaron en una aerolínea internacional de enero de 1949 a diciembre de 1960. Los datos se encuentran en el anexo A de este trabajo.

Para entrar al paquete se tecleó RATS86. Primero para introducir datos al paquete y teniéndolos en un archivo llamado SERIEG.DAT se hizo lo siguiente:

```
OPEN DATA A:SERIEG.DAT
CALENDAR 1949 1 12
ALLOCATE O 1960,12
DATA 1949,1 1960,12 SER
PRINT(DATES) 1949,1 1960,12 SER
```

Con lo anterior se especificó que la serie estaba contenida en un archivo llamado SERIEG.DAT que se encontraba en el drive A, que empezaría en el primer periodo del año de 1949 y que la serie contendría 12 periodos por cada año, además se le dijo a RATS que el último periodo de la serie es el de diciembre de 1960 (60,12), diciéndole que la leerá y la guardará con el nombre de SER para uso interno del paquete, finalmente se desplegó en pantalla para hacer una revisión.

Ya con la serie en RATS pudo efectuarse el análisis.

```
PLOT(DATES,CHARS) 1
# SER 1949,1 1960,12
# '*'
```

Con lo que se graficaron los datos como a continuación se muestra.

En la gráfica puede observarse una tendencia creciente y un componente periódico consistente en un patrón estacional que ocurre aproximadamente en el verano de cada año. También se muestra una amplitud en la serie para los años más recientes. Estos factores deberán tomarse en cuenta para identificar el modelo.

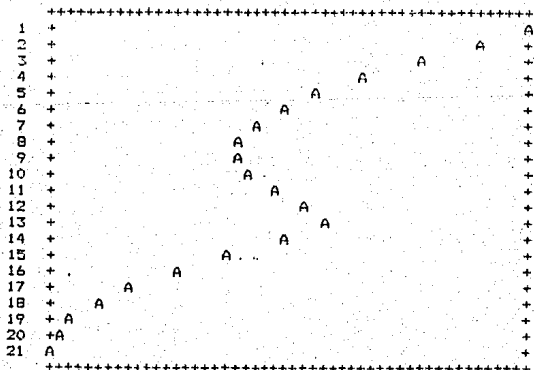
A continuación se muestra la función de autocorrelación.

```
CORRELATE SER 1949,1 1960,12 20 CORSER
PLOT(DATES,CHARS) 1
# CORSER 1949,1 1960,12
# 'A'
```

```
?CORRELATE SER 49,1 60,12 20 CORSER
AUTOCORRELATIONS OF SERIES 1 SER VARIANCE 14292.
NUMBER OF OBSERVATIONS 144
FROM 49- 1 UNTIL 60-12
1 : .9480 .8756 .8067 .7526 .7138 .6817
7 : .6629 .6556 .6709 .7027 .7432 .7604
13 : .7127 .6463 .5859 .5380 .4997 .4687
19 : .4499 .4416
```

```
?PLOT(1
?# CORSER
```

```
SYMBOL A: SERIES 2 CORSER
MIN VALUE .44163 MAX VALUE 1.0000 SPACING .11395E-01
```

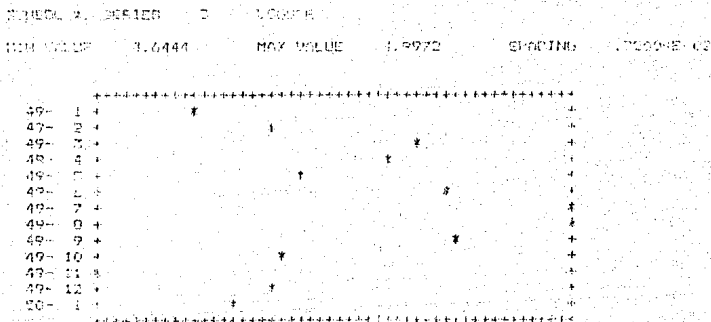


La no estacionaridad puede observarse también aquí, para los primeros períodos la función es alta, decreciendo conforme avanza. Se sugiere entonces una periodicidad de 12 debido a que para el período doce la función vuelve a aumentar.

La amplitud puede removerse al transformar la serie obteniendo el logaritmo de cada observación.

```
SET LOGSER 1949,1 1960,12 = LOG(SER(T))
PLOT(DATES,CHARS) 1
# LOGSER 1949,1 1960,12
# '*'
```

```
DIFFERENCE LOGSER 1949,2 1960,12 1 DIF1 1949,2
SET DIF2 1950,2 1960,12 = DIF1(T) - DIF1(T-12)
```



Habiendo transformado la serie, se graficó nuevamente y se pudo ver que la amplitud era constante. Sin embargo, la serie muestra todavía no estacionaridad en forma anual y en forma creciente. Para remover estos factores, se sugiere obtener las primeras diferencias estacionales y las primeras diferencias no estacionales.

```
DIFFERENCE LOGSER 1949,2 1960,12 1 DIF1 1949,2
SET DIF2 1950,2 1960,12 = DIF1(T) - DIF1(T-12)
```

Con esto se obtuvieron las primeras diferencias no estacionales y las primeras estacionales respectivamente. Como RATS no permite sacar por medio del comando DIFFERENCE diferencias estacionales, se utilizó el comando SET para obtenerlas.

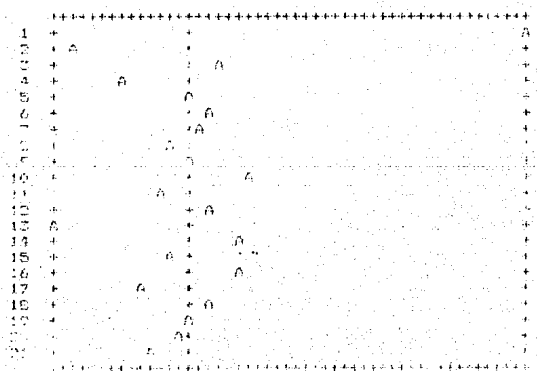
```
CORRELATE DIF2 1950,2 1960,12 20 CORDIF2
PLOT(DATEB,CHARS) 1
# CORDIF2 1950,2 1960,12
# '*'
```

Una vez obtenidas se calculó la función de autocorrelación.

```
*****
AUTOCORRELATION OF THE SERIES CORDIF2
NUMBER OF OBSERVATIONS 17
FROM TO C UNITL 1950,2 1960,12
1 : 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
2 : -.00000000 -.00000000 .00000000 .00000000 .00000000 .00000000
3 : -.00000000 -.00000000 .00000000 .00000000 .00000000 .00000000
4 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
5 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
6 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
7 : -.00000000 -.00000000 .00000000 .00000000 .00000000 .00000000
8 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
9 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
10 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
11 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
12 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
13 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
14 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
15 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
16 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
17 : .00000000 .00000000 .00000000 .00000000 .00000000 .00000000
```

```
PLOT 1
OF CORDIF2
```

```
SYMBOL A: SERIES 4 CORDIF2
MIN VALUE -1.000000 MAX VALUE 1.000000 SPACING 1.000000
```

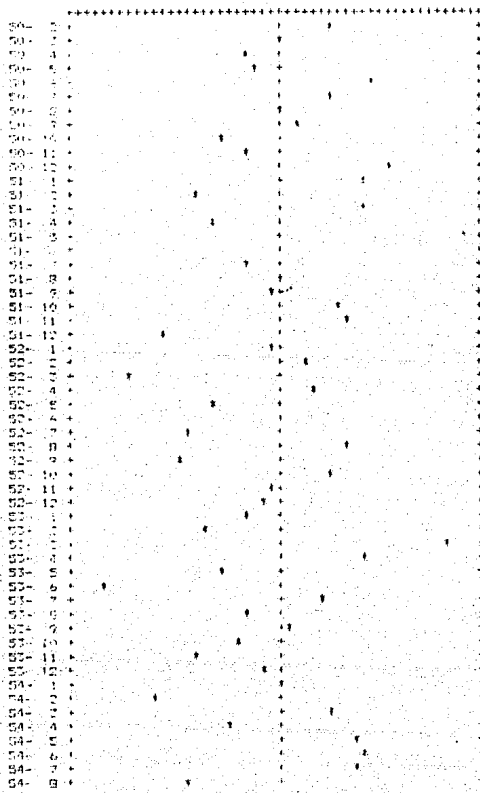


Para este caso se obtiene la menor desviación estándar, en comparación con las demás combinaciones entre diferencias estacionales y no estacionales. La función de autocorrelación de la serie tiene picos significantes en los periodos uno y doce, lo que se tendrá en cuenta para modelarlo.

PLOT(DATES,CHARS) 1
DIF2 1950,2 1960,12
'*'

Graticando nuevamente la serie se obtuvo:

CRUSON # SERIES 5 DIF2
MIN VALUE .14134 MAX VALUE .14977 SPACING .57564E-02



La gráfica muestra que la estacionaridad ha sido obtenida. Obteniendo ahora la función de autocorrelación parcial.

PARTIAL DIF2 1950,2 1960,12 20 PCORDIF2
 PLOT(DATES,CHARS) 1
 # PCORDIF2 1950,2 1960,12
 # '*'

PARTIAL DIF2 50,2 60,12 20 PCORDIF2
 PARTIAL AUTOCORRELATIONS OF SERIES 5 DIF2
 NUMBER OF OBSERVATIONS 131
 FROM 50- 2 UNTIL 60-12

1 :	-.3411	-.1281E-01	-.1927	-.1250	.3309E-01	.3468E-01
7 :	-.6019E-01	-.2022E-01	.2256	.4307E-01	.4659E-01	-.3387
13 :	-.1092	-.7684E-01	-.2175E-01	-.1395	.2589E-01	.1148
19 :	-.1316E-01	-.1674				

?PLOT 1
 ?# PCORDIF2

SYMBOL A: SERIES 7 PCORDIF2
 MIN VALUE -.34112 MAX VALUE 1.0000 SPACING .27370E-01

```

+++++
 1 + + A
 2 A + +
 3 + + A +
 4 + A + +
 5 + A + +
 6 + + A +
 7 + + A +
 8 + A + +
 9 + A + +
10 + + A +
11 + + A +
12 + + A +
13 A + + +
14 + A + +
15 + A + +
16 + A + +
17 + A + +
18 + +A + +
19 + + A +
20 + A + +
21 + A + +
+++++

```

Observando ambas funciones (autocorrelación y autocorrelación parcial) se sugiere un modelo con un parámetro de promedios móviles y uno de promedios móviles estacional, pero RATS no puede modelar este tipo de modelos (por el parámetro estacional) por lo que no se pudo efectuar la estimación ni la predicción en este paquete.

Con el comando END se sale del paquete.

FORECAST PLUS.

De este paquete sólo se pudo conseguir el conjunto de diskettes que lo forman, sin poder ver el manual que lo acompaña, así, el análisis de este paquete será basado únicamente en aquellas observaciones que pudieron hacerse al trabajar con él.

- Lenguaje de Control.

Este paquete es sumamente accesible, aunque poco flexible, debido ésto a su manejo por medio de menus y preguntas, las cuales sólo conducen a caminos previamente definidos. Basta dar el número correspondiente a la opción deseada para estar dentro del módulo escogido o haber obtenido el resultado deseado. Ciertos procedimientos (como el de estimación para el análisis de Box y Jenkins) necesitan algunos datos para correr. Estos son pedidos al usuario por medio de paneles, para los cuales existen ciertos mensajes de error, apareciendo solamente de no dar un valor dentro del rango permitido, avisando así el tipo de valor que espera para ese campo.

Al entrar al paquete, se muestra un menú inicial con cinco incisos:

```
Data Management
Exploratory Package
Forecasting Analysis
Batch Mode
End of Program.
```

- Estructura de los Datos.

Aparentemente FORECAST PLUS es muy flexible para la importación de datos, ya que, en el módulo de manejo de datos en la opción de introducción de datos (opción 1 de este módulo) pide el nombre del archivo (con su extensión) donde se encuentra la serie sin existir restricción en la extensión que se tiene (se probó con .DAT, .PRN, .TXT) pide además ciertos datos acerca del archivo como el número de variables, el número de observaciones, la periodicidad de la serie, el período en el que empiezan los datos de la serie y el nombre con el que se quiere etiquetar a cada variable. Sin embargo no se puede especificar nada en caso de que el archivo tenga cierto formato. Después de tener éstos datos FORECAST PLUS creará un archivo .LBL en el que se tendrán éstos datos para próximos análisis con esta serie. Es también posible introducir datos por medio de un editor propio de FORECAST PLUS.

Existe también una opción (Change Parameter Table) dentro del módulo de manejo de datos en la que se permite

definir ciertos procesos como son: El ancho de hoja (default 80 caracteres), los codigos de inicialización de impresora (de los cuales no se conoce su funcionamiento), el drive de default para los datos, el nombre del dispositivo de salida o nombre del archivo deseado (monitor, impresora o archivo tipo ASCII) y avance de página antes de imprimir, entre otros.

Para realizar el análisis de Box y Jenkins es necesario utilizar dos de los módulos, el de Exploratory Package (en la opción de Time Plot) para graficar la serie con alguna de sus transformaciones y para graficar la función de autocorrelación. La de Forecasting Analysis (en la opción de análisis de Box-Jenkins) para la identificación, estimación y predicción del modelo. Es necesario notar que la función de autocorrelación puede graficarse en ambos módulos con la única diferencia de que en el primero pregunta el número de periodos que se desean graficar y en el segundo siempre da 32 periodos. Para cada caso se muestran paneles que es necesario llenar y con los que se obtiene automáticamente el resultado pedido. Entre paso y paso los paneles son memorizados con los datos de la última identificación o estimación efectuada, sin que esto signifique que no puedan cambiarse en el último momento, además siempre es posible regresar al paso anterior. Además para graficar la serie cuando se ha hecho alguna diferencia (estacional o no estacional), es necesario primero transformarla en el módulo de manejo de datos (opcion 5, transformación de vars.) y luego graficarla en módulo de Exploratory Package (en la opción de Time Plot)

- Impresión de Resultados.

La impresión de resultados no es opcional, ya que cuando se decide direccionar la salida a impresora todos los resultados se van directo a ella y no pueden verse en pantalla o viceversa. Así si se desea graficar un resultado visto en pantalla se deberá redireccionar la salida y volver a efectuar el comando que produjo el resultado, además, cuando los resultados son desplegados en pantalla éstos no se detienen automáticamente para poder analizarlos sino que se despliegan en forma continua y al terminar vuelve de inmediato al menú del que se partió sin poder verse ni siquiera el final del resultado. Aunque no permite modificaciones sobre las gráficas directamente, puede direccionarse la salida a algún archivo de tipo ASCII donde posteriormente con ayuda de un procesador de palabra pueden hacerse las modificaciones deseadas.

- Gráficas.

Por otro lado, los resultados y gráficas son mostrados en forma muy clara y completa, dan información como el nombre de la variable (especificando si hubo alguna transformación), el módulo que los produjo (en este caso por

ejemplo el análisis de Box y Jenkins), para las funciones de autocorrelación y autocorrelación parcial, el número de observaciones, la media y desviación estándar del proceso de la serie, el grado de diferencias estacionales y no estacionales, y además para la estimación, el nombre de cada parámetro, el valor estimado, la estadística T y la matriz de correlación.

Por desgracia los valores de las funciones de autocorrelación y autocorrelación parcial, de los residuales etc. no pueden salvarse para ser exportados y utilizados posteriormente.

- *Costo.*

FORECAST PLUS consta de tres diskettes entre los cuales no existe ningún archivo con ejemplos guiados ni de lectura previa al uso. Se espera que toda la información necesaria referente al paquete se encuentre contenida en el manual.

A continuación se muestra el funcionamiento de los módulos necesarios para efectuar el análisis de Box y Jenkins.

Para introducir los datos, escogiendo la opción :

- 1) DATA MANAGEMENT *Se desplegará un submenú en el cual la primera opción es la de crear un archivo de datos.*
1. ENTER NEW DATA *Permite leer de un archivo los datos de una serie para lo cual realiza una serie de preguntas:*
- What data file would you like to add to?

A la que habrá que contestar el nombre del archivo con cualquier extensión.

Si ya se ha creado en alguna otra sesión se poseccionará en el último registro por si se desea añadir datos, si nunca antes se ha trabajado con ella preguntará si se desea o no crear. Si a esta pregunta se contesta afirmativamente se desplegará un panel con las siguientes preguntas:

- Number of variables *Pidiendo el número de variables que contiene el archivo.*
- Type of Labeling *La periodicidad (anual, mensual, trimestral, semanal).*
- Starting label *El periodo del año en el que empieza la serie.*
- Starting observation *El periodo en que comienzan las observaciones.*
- Number of observations between each label *Para cada dos periodos especificados arriba, el número de observaciones que habrá entre ellos.*
- Increment of the labels *No se conoce su funcionamiento.*
- Variable <n> Label: *Para cada variable arriba especificada preguntará la etiqueta que se le desea dar.*

Después de haberla creado se posecciona en el último registro del archivo por si se quiere agregar algún otro dato. Para salir del modo de edición deberá darse ESC con lo que lleva a un submenú que pregunta si se desean salvar los nuevos datos, regresar al modo de edición o regresar al menú de DATA MANAGEMENT sin salvar los datos nuevos (de escoger está última preguntará si se está seguro de no querer salvar los datos).

Para comprobar que los datos esten correctos:

4. PRINT DATA FILE *Despliega los valores de una serie, preguntando por la serie que se desea desplegar. Puede darse RETURN si es la serie con la que se está trabajando, o bien, el nombre de la serie deseada.*

Para direccionar la salida a impresora, monitor o archivo.

6. CHANGE PARAMETER TABLE *Permite el cambio de ciertos parámetros.*

4 Name of output device or file name

Para direccionar a algún archivo basta dar 4 y el nombre del archivo, para el monitor dar 4 y CON y para impresora 4 y PRN.

12 Exit and save changes

Después si se desean salvar los cambios bastará dar 12, con lo que se regresará al menú de DATA MANAGEMENT.

10. RETURN TO MAIN MENU *Regresa al menú principal.*

Para graficar la serie y la función de autocorrelación, será necesario utilizar el módulo:

2) EXPLORATORY PACKAGE *Desde el menú principal puede introducirse a este módulo en el que se preguntará la serie con la se va a trabajar. Puede darse RETURN si es la serie con la que se esta trabajando, o bien, el nombre de la serie deseada.*

1. TIME PLOT

Despliega la gráfica de la serie escogida, después de escoger alguna transformación.

Transformation to original data series:

(1 = Reciprocal, 2 = 1/sqrt, 3 = 1/4th root,

4 = Log, 5 = 4th root, 6 = Sqrt, 7 = None):

Pudiendo escoger entre estas, dando el número desedo (default 7).

6. AUTOCORRELATION FUNCTION

Produce la función de autocorrelación de la serie deseada. Solicitando la siguiente información:

Transformation to original data series:

(1 = Reciprocal, 2 = 1/sqrt, 3 = 1/4th root,

4 = Log, 5 = 4th root, 6 = Sqrt, 7 = None):

Pudiendo escoger entre estas transformaciones, dando el número desedo.

Degree of regular differencing:

Grado de diferencias no estacionales deseadas.

Degree of seasonal differencing:

Grado de diferencias estacionales deseadas.

Length of seasonality:

Longitud de la estacionaridad.

Number of lags to print: *Número de periodos de la función de autocorrelación que se desean graficar.*

Regresa después al menú de EXPLORATORY PACKAGE.

7. RETURN TO MAIN MENU *Regresa al menú principal.*

Ahora para efectuar el análisis es necesario introducirse al menú de FORECASTING ANALYSIS (opción 3) en donde se vuelve a preguntar la serie que será utilizada en este módulo, después, despliega otro submenú con diversos métodos de análisis entre los que se encuentra el de Box y Jenkins (opción 13).

13 BOX-JENKINS ANALYSIS *Ya dentro de este menú se despliegan tres opciones identificación, estimación/predicción y regreso al menú de FORECASTING ANALYSIS.*

1) IDENTIFICATION *Pueden desplegarse las funciones de autocorrelación y autocorrelación parcial, pidiendo la siguiente información.*

Transformation to original data series:

(1 = Reciprocal, 2 = 1/sqrt, 3 = 1/4th root,
4 = Log, 5 = 4th root, 6 = Sqrt, 7 = None):

Pudiendo escoger entre estas transformaciones, dando el número deseado.

Degree of regular differencing:

Grado de diferencias no estacionales deseadas.

Degree of seasonal differencing:

Grado de diferencias estacionales deseadas.

Length of seasonal period: *Longitud de la estacionariedad.*

Plot of autocorrelation function:

Plot of partial autocorrelation function:

A estas dos últimas preguntas es necesario contestar Y (Yes) o N (No).

Posteriormente a la gráfica se regresa automáticamente al menú de análisis de Box y Jenkins.

2) ESTIMATION/FORECASTING *En este módulo puede ajustarse el modelo deseado para lo cual es necesario llenar el panel siguiente.*

Transformation to original data series:

(1 = Reciprocal, 2 = 1/sqrt, 3 = 1/4th root,
4 = Log, 5 = 4th root, 6 = Sqrt, 7 = None):

Pudiendo escoger entre estas transformaciones, dando el número deseado.

Degree of regular differencing:

Grado de diferencias no estacionales deseadas.

Degree of seasonal differencing:

Grado de diferencias estacionales deseadas.

Length of seasonal period: *Longitud de la estacionariedad.*

Regular Autoregressive Terms:

Número de parámetros autoregresivos no estacionales del modelo.

Regular Moving Average Terms:

Número de parámetros de promedios móviles no estacionales del modelo.

Seasonal Autoregressive Terms:

Número de parámetros autoregresivos estacionales del modelo.

Seasonal Moving Average Terms:

Número de parámetros de promedios móviles estacionales del modelo.

Include Constant Term(Y/N/A):

Si el usuario desea que el modelo contenga un término constante dará Y, si no N, o bien, podrá dejar que el paquete lo decida A (automatico).

Number of Backforecasts:

En este campo puede darse la opción de predecir hacia atrás el número indicado de períodos.

Maximun Iterations for Parameter Estimates:

Permite escoger el número de iteraciones deseadas para la estimación.

Convergence Tolerance:

Permite también escoger el valor de convergencia del algoritmo.

Initial Estimates from Data (A) or User (U):

Pueden darse al paquete valores iniciales para los parámetros a estimar.

Plot of Residual Autocorrelation Function:

Puede obtenerse opcionalmente la función de autocorrelación de los residuales dejando Y (Yes) en este campo, o N (No) si no se desea.

Al finalizar la estimación el paquete pregunta si se desea hacer otra estimación, si se desea predecir o si se desea regresar al menú del análisis de Box y Jenkins. En el caso de predicción:

1. DO FORECASTING? *Predecirá con los valores de la última estimación hecha preguntando:*
- Number of forecasts: *Aquí se puede especificar el número de periodos que se desean predecir.*
- Origin for forecast: *En este campo se dice el periodo en el que se quiere comience la predicción.*
- Percent forecast confidence interval: *Puede especificarse el intervalo de confianza para la predicción.*
- Time plot of original data, forecast, and error: *Si se desean graficar los datos originales, las predicciones y el error se dará Y (Yes), de lo contrario N (No).*
- Table of original data, forecast, and error: *Si se desea una tabla con los valores de los datos originales, las predicciones y el error se dará Y (Yes), de lo contrario N (No).*

Al terminar de desplegar la gráfica y/o la tabla el paquete regresa al menú del análisis de Box y Jenkins, el cual tiene una opción (3) para regresar al submenú de FORECASTING ANALYSIS que a su vez tiene otra opción (14) para regresar al menú principal, donde para salir del paquete será necesario teclear 5 que corresponde a la opción de END OF PROGRAM.

Es necesario decir que al final de cada uno de los paneles de preguntas, FORECAST PLUS se detiene preguntando si los datos están correctos, teniendo la posibilidad de corregirlos. Para esto aparece la leyenda "Press ENTER to accept".

A continuación se presenta un ejemplo, mostrándose en forma sencilla el análisis de una serie tomada del libro de Box y Jenkins "Time Series Analysis, Forecasting and Control" que contiene datos del total de pasajeros (en miles) que viajaron en una aerolínea internacional de enero de 1949 a diciembre de 1960 (serie G). Los datos se encuentran listados al final de la tesis en el anexo A.

Para llamar al paquete se utiliza el comando FP. Donde el primer paso es el de introducir los datos:

```
1) DATA MANAGEMENT
1. ENTER NEW DATA
A:SERIEG.DAT
```

Con esto se introdujo al submenú de introducción de una variable dentro del menú de manejo de datos, introduciendo la serie que está en el archivo SERIEG.DAT creándose con una variable, mensual, iniciándose en enero, con una observación entre cada período y llamandola SER. Con la serie en el paquete puede entonces graficarse:

```
10 RETURN TO MAIN MENU
2) EXPLORATORY PACKAGE
1. TIME PLOT
```

Pidiendo la gráfica de los datos originales se obtuvo:

SER - EXPLORATORY DATA ANALYSIS

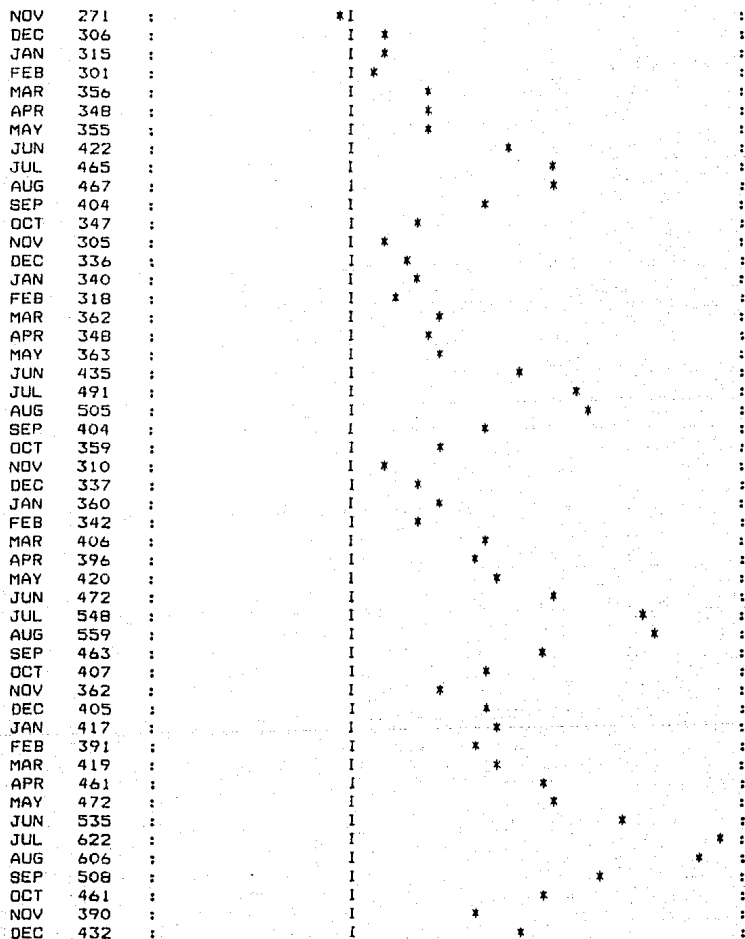
TIME PLOT

144 observations in the series

Mean of the series = 280.3

Standard deviation of the series = 119.966

Time	Value	104	280.3	622
JAN	112	*	1	:
FEB	118	*	1	:
MAR	132	*	1	:
APR	129	*	1	:
MAY	121	*	1	:
JUN	135	*	1	:
JUL	148	*	1	:
AUG	148	*	1	:
SEP	136	*	1	:
OCT	119	*	1	:
NOV	104	*	1	:
DEC	118	*	1	:
JAN	115	*	1	:
FEB	126	*	1	:
MAR	141	*	1	:
APR	135	*	1	:
MAY	125	*	1	:
JUN	149	*	1	:
JUL	170	*	1	:
AUG	170	*	1	:
SEP	158	*	1	:
OCT	133	*	1	:
NOV	114	*	1	:
DEC	140	*	1	:
JAN	145	*	1	:
FEB	150	*	1	:
MAR	178	*	1	:
APR	163	*	1	:
MAY	172	*	1	:
JUN	178	*	1	:
JUL	199	*	1	:
AUG	199	*	1	:
SEP	184	*	1	:
OCT	162	*	1	:
NOV	146	*	1	:
DEC	166	*	1	:
JAN	171	*	1	:
FEB	180	*	1	:
MAR	193	*	1	:
APR	181	*	1	:



 * = Observed Value

En la gráfica puede observarse una tendencia creciente, además, tiene un componente periódico consistente en un patrón estacional que ocurre aproximadamente en el verano de cada año. También se muestra una amplitud en la serie para los años más recientes. Estos factores deberán tomarse en cuenta para identificar el modelo.

6. AUTOCORRELATION FUNCION

A continuación se muestra la función de autocorrelación para veinte periodos

SER - EXPLORATORY DATA ANALYSIS

AUTOCORRELATION FUNCION

144 Observations in the Working Series
 Mean of the Working Series = 280.3
 Standard Deviation of the Working Series = 119.9663

Lag	Value	T-Value	-1.0	0.0	+1.0
1	0.95	11.44	:	[*****]
2	0.89	6.36	:	[*****]
3	0.83	4.76	:	[*****]
4	0.79	3.95	:	[*****]
5	0.78	3.53	:	[*****]
6	0.78	3.25	:	[*****]
7	0.78	3.04	:	[*****]
8	0.79	2.89	:	[*****]
9	0.82	2.85	:	[*****]
10	0.88	2.88	:	[*****]
11	0.94	2.94	:	[*****]
12	0.98	2.89	:	[*****]
13	0.94	2.62	:	[*****]
14	0.87	2.31	:	[*****]
15	0.80	2.07	:	[*****]
16	0.76	1.90	:	[*****]
17	0.75	1.82	:	[*****]
18	0.74	1.77	:	[*****]
19	0.74	1.72	:	[*****]
20	0.74	1.70	:	[*****]

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 20 Degrees of Freedom = 1989

La no estacionaridad de la serie puede observarse también en la función de autocorrelación, para los primeros periodos la función es alta, decreciendo conforme avanza. Es sugerida una periodicidad de doce debido a que la función crece nuevamente en el periodo 12.

La amplitud de la serie puede removerse al transformar la serie obteniendo el logaritmo de cada observación.

1) TIME PLOT

Graficandola ahora con logaritmos se obtiene

Log of SER - EXPLORATORY DATA ANALYSIS

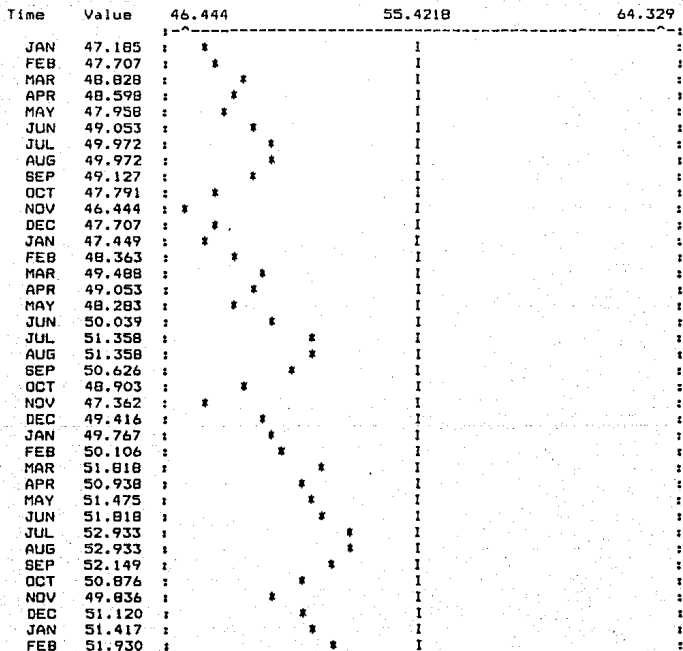
TIME PLOT

All Data Values are Scaled by $\times 10^1$

144 observations in the series

Mean of the series = 55.4218

Standard deviation of the series = .441608



BEP	58.721	:	I		*				:
OCT	57.236	:	I		*				:
NOV	56.021	:	I	*					:
DEC	57.236	:	I		*				:
JAN	57.526	:	I		*				:
FEB	57.071	:	I		*				:
MAR	58.749	:	I		*				:
APR	58.522	:	I		*				:
MAY	58.721	:	I		*				:
JUN	60.450	:	I		*				:
JUL	61.420	:	I		*				:
AUG	61.463	:	I		*				:
SEP	60.014	:	I		*				:
OCT	58.493	:	I		*				:
NOV	57.203	:	I	*					:
DEC	58.171	:	I		*				:
JAN	58.289	:	I		*				:
FEB	57.621	:	I	*					:
MAR	58.916	:	I		*				:
APR	58.522	:	I		*				:
MAY	58.944	:	I		*				:
JUN	60.753	:	I		*				:
JUL	61.964	:	I		*				:
AUG	62.246	:	I		*				:
SEP	60.014	:	I		*				:
OCT	58.833	:	I		*				:
NOV	57.366	:	I	*					:
DEC	58.201	:	I		*				:
JAN	58.861	:	I		*				:
FEB	58.348	:	I		*				:
MAR	60.064	:	I		*				:
APR	59.814	:	I		*				:
MAY	60.403	:	I		*				:
JUN	61.570	:	I		*				:
JUL	63.063	:	I		*				:
AUG	63.261	:	I		*				:
SEP	61.377	:	I		*				:
OCT	60.088	:	I		*				:
NOV	58.916	:	I	*					:
DEC	60.039	:	I		*				:
JAN	60.331	:	I		*				:
FEB	59.687	:	I		*				:
MAR	60.379	:	I		*				:
APR	61.334	:	I		*				:
MAY	61.570	:	I		*				:
JUN	62.823	:	I		*				:
JUL	64.329	:	I		*				:
AUG	64.069	:	I		*				:
SEP	62.305	:	I		*				:
OCT	61.334	:	I		*				:
NOV	59.661	:	I	*					:
DEC	60.684	:	I		*				:

* = Observed Value

Notese que la amplitud de la serie es ahora constante en el tiempo. Sin embargo, la serie transformada muestra no estacionaridad en forma anual y en forma creciente. Para remover éstos factores, lo más conveniente será obtener las primeras diferencias estacionales y las primeras diferencias no estacionales.

- 7. RETURN TO MAIN MENU.
- 3) FORECASTING ANALYSIS.
- 13 BOX-JENKINS ANALYSIS.
- 1. IDENTIFICATION.

Introduciendose hasta el menú donde se encuentra el panel de identificación del modelo para obtener las funciones de autocorrelación y autocorrelación parcial, se lleno el panel de la siguiente forma. 4 (logaritmo de la serie), 1 grado de diferencia estacional y 1 no estacional, 12 periodos de longitud estacional y Y para la obtención de las dos funciones.

Obteniendo las funciones de autocorrelación y autocorrelación parcial nuevamente para una diferencia estacional y una diferencia no estacional.

Log of SER - EXPLORATORY DATA ANALYSIS

AUTOCORRELATION FUNCTION

131 Observations in the Working Series
 Mean of the Working Series = .00029
 Standard Deviation of the Working Series = .0458483

Degree of Regular Differencing = 1
 Degree of Seasonal Differencing = 1

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.34	-3.89	:	****[***] :	:
2	0.11	1.08	:	[***] :	:
3	-0.20	-2.07	:	[****] :	:
4	0.02	0.21	:	[*] :	:
5	0.06	0.56	:	[*] :	:
6	0.03	0.31	:	[*] :	:
7	-0.06	-0.56	:	[*] :	:
8	-0.00	-0.00	:	[] :	:
9	0.19	1.83	:	[*****] :	:
10	-0.08	-0.79	:	[**] :	:
11	0.07	0.67	:	[**] :	:
12	-0.42	-4.05	:	*****[****] :	:
13	0.17	1.44	:	[****] :	:
14	-0.06	-0.54	:	[**] :	:
15	0.17	1.41	:	[****] :	:
16	-0.16	-1.34	:	[****] :	:
17	0.09	0.70	:	[**] :	:
18	0.02	0.16	:	[] :	:
19	-0.01	-0.10	:	[] :	:
20	-0.14	-1.17	:	[****] :	:

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 20 Degrees of Freedom = 67.8
 Probability = 0

Log of GER - BOX-JENKINS ANALYSIS: IDENTIFICATION

PARTIAL AUTOCORRELATION FUNCTION

131 Observations in the Working Series
 Mean of the Working Series = 0
 Standard Deviation of the Working Series = .0458

Degree of Regular Differencing = 1
 Degree of Seasonal Differencing = 1

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.34	-3.89		****[***]	
2	-0.01	-0.13		[]	
3	-0.19	-2.21		*[***]	
4	-0.12	-1.42		[***]	
5	0.04	0.40		[*]	
6	0.04	0.42		[*]	
7	-0.06	-0.69		[**]	
8	-0.02	-0.22		[]	
9	0.24	2.72		[[***]]*	
10	0.04	0.50		[*]	
11	0.05	0.58		[*]	
12	-0.38	-4.34		*****[***]	
13	-0.12	-1.40		[***]	
14	-0.08	-0.97		[**]	
15	-0.01	-0.15		[]	
16	-0.17	-1.91		[***]	
17	0.04	0.43		[*]	
18	0.15	1.72		[[***]]	
19	-0.00	-0.04		[]	
20	-0.22	-2.52		**[***]	
21	0.16	1.80		[[***]]	
22	-0.10	-1.14		[**]	
23	0.21	2.38		[[***]]*	
24	-0.10	-1.19		[***]	
25	-0.13	-1.49		[***]	
26	-0.01	-0.08		[]	
27	0.09	1.05		[**]	
28	-0.14	-1.65		[***]	
29	0.08	0.89		[**]	
30	-0.04	-0.40		[*]	
31	-0.13	-1.53		[***]	
32	0.01	0.06		[]	

[] = Estimated Two-Standard Error Limits

Para este caso se obtiene la menor desviación estándar, en comparación de las demás combinaciones entre diferencias estacionales y no estacionales (es decir, 1 dif. estacional y 0 no estacional, y viceversa). La función de autocorrelación de la serie con estas diferencias tiene picos significantes en los periodos uno y doce, lo que sugiere que no es aleatorio y es necesario modelarlo.

La gráfica muestra que la estacionariedad ha sido obtenida. Observando ambas funciones (autocorrelación y autocorrelación parcial) se sugiere un modelo con un parámetro de promedios móviles y uno de promedios móviles estacional. Usualmente, no es tan fácil distinguir el modelo a utilizar.

Habiendo terminado con la identificación del modelo se procederá con la estimación:

El modelo contendrá un término de promedios móviles, uno de promedios móviles estacionario y no contendrá constante. El reporte se obtiene de la siguiente forma:

2. ESTIMATION/FORECASTING.

Llenando el panel con la siguiente información. 4 (logaritmo de la serie), 1 grado de diferencia estacional y 1 no estacional, 12 periodos de longitud estacional, 0 para el término autoregresivo estacional y no estacional, 1 para el término de promedios móviles estacional y no estacional, N para no incluir constante, 13 predicciones hacia atrás y Y obteniéndose la función de autocorrelación de los residuales.

Log of SER - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

131 Observations Used For Parameter Estimates
 Number of Backforecasts = 13
 Degree of Regular Differencing = 1
 Degree of Seasonal Differencing = 1
 Seasonal Period = 12

Sum of Squared Errors = 0
 Residual Variance = .001

Parameter	Value	T-Statistic
-----	-----	-----
MA(1)	0.3901	12.02
SMA(1)	0.5628	17.39

Correlation Matrix

	MA(1)	SMA(1)
	-----	-----
MA(1)	1.00	
SMA(1)	-0.01	1.00

Log of SER - BOX-JENKINS ANALYSIS

RESIDUAL AUTOCORRELATION FUNCTION

130 Observations in the Residual Series
 Mean of the Residual Series = 0
 Standard Deviation of the Residual Series = .0363

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.02	-0.25	:	[#]	:
2	0.03	0.30	:	[#]	:
3	-0.13	-1.51	:	[###]	:
4	-0.14	-1.55	:	[###]	:
5	0.05	0.50	:	[#]	:
6	0.05	0.59	:	[#]	:
7	-0.06	-0.71	:	[##]	:
8	-0.02	-0.25	:	[#]	:
9	0.12	1.28	:	[####]	:
10	-0.10	-1.04	:	[##]	:
11	-0.02	-0.19	:	[#]	:
12	-0.04	-0.40	:	[#]	:
13	0.02	0.20	:	[#]	:
14	0.06	0.59	:	[#]	:
15	0.06	0.65	:	[##]	:
16	-0.19	-1.98	:	[#####]	:
17	0.03	0.34	:	[#]	:
18	-0.00	-0.02	:	[#]	:
19	-0.12	-1.24	:	[###]	:
20	-0.10	-1.06	:	[###]	:
21	-0.03	-0.34	:	[#]	:
22	-0.04	-0.43	:	[#]	:
23	0.25	2.51	:	[######]	:
24	0.01	0.10	:	[#]	:
25	-0.03	-0.27	:	[#]	:
26	0.09	0.83	:	[##]	:
27	-0.05	-0.46	:	[#]	:
28	-0.11	-1.01	:	[###]	:
29	-0.05	-0.49	:	[#]	:
30	-0.10	-0.96	:	[###]	:
31	-0.06	-0.54	:	[#]	:
32	0.20	1.82	:	[#####]	:
33	-0.20	-1.80	:	[#####]	:

[] = Estimated Two-Standard Error Limits

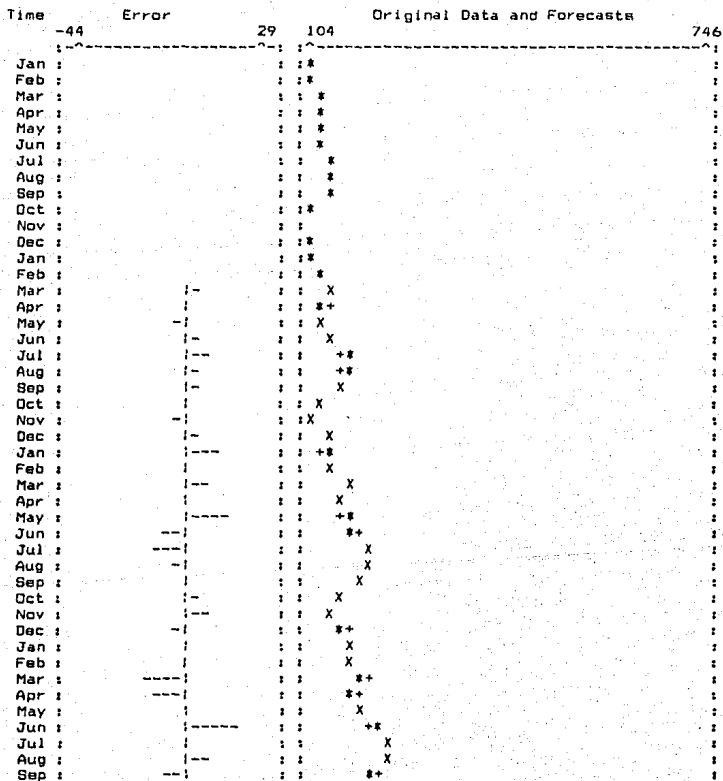
Box-Pierce Chi-Square Statistic with 31 Degrees of Freedom = 41.9
 Probability = .091

Revisando estas gráficas puede verse que el modelo es adecuado, ya que cada punto es relativamente cercano a cero. Para la predicción se tiene:

1 DO FORECASTING?

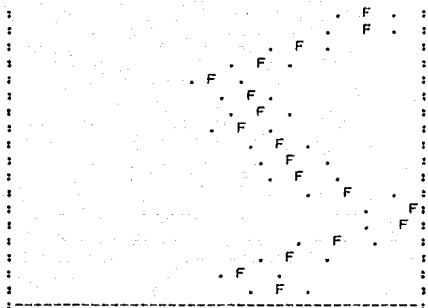
Se efectuó la predicción para dos años (24 periodos), empezando desde el periodo 144 que es el fin de la serie, 95% de confianza y pidiendo las gráficas y tablas de los datos originales, las predicciones y el error.

SER - BOX-JENKINS ANALYSIS
 TIME PLOT OF ORIGINAL DATA, FORECASTS, AND ERROR



Oct :	-	:	:	X	:
Nov :	--	:	:	+X	:
Dec :	-	:	:	X	:
Jan :	-	:	:	X+	:
Feb :	-----	:	:	X+	:
Mar :	-	:	:	+X	:
Apr :	-	:	:	+X	:
May :	-	:	:	X	:
Jun :	-----	:	:	X+	:
Jul :	-----	:	:	X+	:
Aug :	-	:	:	X+	:
Sep :	-	:	:	X+	:
Oct :	-	:	:	X+	:
Nov :	-	:	:	X+	:
Dec :	-----	:	:	X+	:
Jan :	-	:	:	X	:
Feb :	-----	:	:	X+	:
Mar :	-	:	:	X	:
Apr :	-	:	:	X	:
May :	-	:	:	+X	:
Jun :	-----	:	:	+X	:
Jul :	-	:	:	X	:
Aug :	-	:	:	X+	:
Sep :	-----	:	:	X+	:
Oct :	-	:	:	X+	:
Nov :	-	:	:	X	:
Dec :	-	:	:	X	:
Jan :	-	:	:	X	:
Feb :	-	:	:	X	:
Mar :	-----	:	:	X+	:
Apr :	-	:	:	X	:
May :	-	:	:	X	:
Jun :	-	:	:	+X	:
Jul :	-	:	:	+X	:
Aug :	-	:	:	X	:
Sep :	-	:	:	X	:
Oct :	-	:	:	X	:
Nov :	-	:	:	X	:
Dec :	-	:	:	+X	:
Jan :	-	:	:	X	:
Feb :	-	:	:	X	:
Mar :	-----	:	:	X+	:
Apr :	-	:	:	X	:
May :	-	:	:	X	:
Jun :	-	:	:	+X	:
Jul :	-	:	:	X	:
Aug :	-	:	:	X	:
Sep :	-----	:	:	X	:
Oct :	-----	:	:	X+	:
Nov :	-	:	:	X	:
Dec :	-	:	:	X+	:
Jan :	-	:	:	X	:
Feb :	-----	:	:	X+	:
Mar :	-	:	:	X	:

Jul	673.6	:
Aug	670.8	:
Sep	561.4	:
Oct	499.9	:
Nov	432.3	:
Dec	480.0	:
Jan	501.6	:
Feb	474.2	:
Mar	533.8	:
Apr	548.2	:
May	566.8	:
Jun	649.7	:
Jul	746.1	:
Aug	742.9	:
Sep	621.7	:
Oct	553.7	:
Nov	478.8	:
Dec	531.6	:



* = Observed Data Value
 + = 1-Step Ahead Forecast
 F = Forecast From Origin Period 144
 · = Approximate 95% Confidence Interval
 X = Overlap

SER - BOX-JENKINS ANALYSIS

TABLE OF ORIGINAL DATA, FORECASTS, AND ERROR

Time	Original Data	Forecast	95% Confidence Limits		Error
			Lower	Upper	
Jan	112				
Feb	118				
Mar	132				
Apr	129				
May	121				
Jun	135				
Jul	148				
Aug	148				
Sep	136				
Oct	119				
Nov	104				
Dec	118				
Jan	115				
Feb	126				
Mar	141	139.0	129.6	149.2	2.0
Apr	135	136.4	127.1	146.3	-1.4
May	125	128.7	120.0	138.1	-3.7
Jun	149	145.8	135.9	156.4	3.2
Jul	170	161.4	150.5	173.1	8.6
Aug	170	164.8	153.6	176.8	5.2
Sep	158	154.3	143.8	165.5	3.7
Oct	133	135.0	125.8	144.8	-2.0
Nov	114	118.9	110.8	127.5	-4.9
Dec	140	136.7	127.5	146.7	3.3
Jan	145	135.3	126.1	145.1	9.7
Feb	150	149.0	138.9	159.8	1.0
Mar	178	169.7	158.2	182.0	8.3
Apr	163	164.9	153.7	176.9	-1.9
May	172	156.5	145.9	167.9	15.5
Jun	178	184.3	171.8	197.7	-6.3
Jul	199	212.4	198.0	227.8	-13.4
Aug	199	204.5	190.6	219.3	-5.5
Sep	184	185.7	173.1	199.2	-1.7
Oct	162	157.9	147.2	169.3	4.1
Nov	146	137.8	128.5	147.8	8.2
Dec	166	168.1	156.7	180.3	-2.1
Jan	171	172.2	160.5	184.7	-1.2
Feb	180	179.8	167.6	192.9	0.2
Mar	193	206.7	192.7	221.7	-13.7
Apr	181	192.1	179.1	206.1	-11.1
May	183	183.1	170.7	196.4	-0.1
Jun	218	200.6	187.0	215.2	17.4
Jul	230	229.8	214.2	246.5	0.2

Aug	242	235.5	219.5	252.6	6.5
Sep	209	216.8	202.1	232.6	-7.8
Oct	191	188.6	175.8	202.3	2.4
Nov	172	164.3	153.2	176.3	7.7
Dec	194	196.1	182.8	210.3	-2.1
Jan	196	201.5	187.8	216.1	-5.5
Feb	196	209.7	195.5	224.9	-13.7
Mar	236	228.8	213.3	245.5	7.2
Apr	235	212.8	198.4	228.3	22.2
May	229	220.6	205.6	236.6	8.4
Jun	243	260.2	242.6	279.1	-17.2
Jul	264	280.5	261.5	300.9	-16.5
Aug	272	276.8	258.1	296.9	-4.8
Sep	237	245.0	228.4	262.8	-8.0
Oct	211	215.3	200.7	230.9	-4.3
Nov	180	189.1	176.3	202.9	-9.1
Dec	201	214.8	200.2	230.4	-13.8
Jan	204	212.0	197.6	227.4	-8.0
Feb	188	213.4	198.9	228.9	-25.4
Mar	235	237.7	221.6	255.0	-2.7
Apr	227	220.6	205.6	236.6	6.4
May	234	221.5	206.5	237.6	12.5
Jun	264	249.2	232.3	267.3	14.8
Jul	302	279.0	260.1	299.3	23.0
Aug	293	294.7	274.8	316.2	-1.7
Sep	259	265.9	247.9	285.2	-6.9
Oct	229	232.6	216.8	249.5	-3.6
Nov	203	201.7	188.1	216.4	1.3
Dec	229	228.7	213.2	245.3	0.3
Jan	242	233.0	217.2	249.9	9.0
Feb	233	230.3	214.7	247.1	2.7
Mar	267	279.0	260.1	299.3	-12.0
Apr	269	266.5	248.5	285.9	2.5
May	270	265.1	247.2	284.4	4.9
Jun	315	299.5	279.2	321.2	15.5
Jul	364	337.8	314.9	362.3	26.2
Aug	347	348.9	325.3	374.3	-1.9
Sep	312	313.6	292.4	336.4	-1.6
Oct	274	275.4	256.8	295.4	-1.4
Nov	237	242.3	225.9	259.9	-5.3
Dec	278	272.5	254.0	292.3	5.5
Jan	284	281.1	262.0	301.5	2.9
Feb	277	276.5	257.8	296.6	0.5
Mar	317	324.4	302.4	347.9	-7.4
Apr	313	317.6	296.1	340.7	-4.6
May	318	315.9	294.5	338.9	2.1
Jun	374	359.3	335.0	385.5	14.7
Jul	413	411.3	383.4	441.2	1.7
Aug	405	408.0	380.3	437.6	-3.0
Sep	355	361.7	337.2	388.0	-6.7
Oct	306	317.8	296.2	340.9	-11.8
Nov	271	273.7	255.1	293.5	-2.7
Dec	306	311.0	290.0	333.6	-5.0
Jan	315	318.9	297.3	342.1	-3.9

Feb	301	308.1	287.3	330.5	-7.1
Mar	356	355.1	331.0	380.9	0.9
Apr	348	347.8	324.2	373.0	0.2
May	355	352.4	328.5	378.0	2.6
Jun	422	407.0	379.5	436.6	15.0
Jul	465	460.6	429.4	494.1	4.4
Aug	467	456.9	425.9	490.1	10.1
Sep	404	405.8	378.3	435.3	-1.8
Oct	347	356.7	332.5	382.6	-9.7
Nov	305	310.2	289.1	332.7	-5.2
Dec	336	350.2	326.4	375.6	-14.2
Jan	340	356.1	331.9	381.9	-16.1
Feb	318	336.5	313.7	361.0	-18.5
Mar	362	385.1	359.0	413.1	-23.1
Apr	348	368.0	343.1	394.8	-20.0
May	363	363.7	339.1	390.1	-0.7
Jun	435	420.2	391.7	450.7	14.8
Jul	491	472.6	440.6	507.0	18.4
Aug	505	476.4	444.1	511.0	28.6
Sep	404	426.2	397.4	457.2	-22.2
Oct	359	370.7	345.5	397.6	-11.7
Nov	310	316.3	294.9	339.3	-6.3
Dec	337	353.4	329.5	379.1	-16.4
Jan	360	355.1	331.1	380.9	4.9
Feb	342	335.1	312.4	359.4	6.9
Mar	406	392.6	366.0	421.1	13.4
Apr	396	385.8	359.7	413.8	10.2
May	420	402.1	374.8	431.3	17.9
Jun	472	482.0	449.3	517.0	-10.0
Jul	548	543.2	506.4	582.7	4.8
Aug	559	542.1	505.4	581.5	16.9
Sep	463	462.8	431.5	496.5	0.2
Oct	407	408.2	380.6	437.9	-1.2
Nov	362	355.0	330.9	380.8	7.0
Dec	405	394.7	368.0	423.4	10.3
Jan	417	415.9	387.8	446.2	1.1
Feb	391	398.2	371.2	427.1	-7.2
Mar	419	463.1	431.8	496.8	-44.1
Apr	461	439.1	409.3	471.0	21.9
May	472	449.1	418.7	481.7	22.9
Jun	535	539.0	502.5	578.2	-4.0
Jul	622	614.9	573.2	659.6	7.1
Aug	606	621.8	579.7	667.0	-15.8
Sep	508	519.5	484.3	557.2	-11.5
Oct	461	449.2	418.7	481.8	11.8
Nov	390	396.2	369.3	425.0	-6.2
Dec	432	443.2	413.2	475.4	-11.2
Jan		452.9	422.3	485.9	
Feb		428.2	398.5	460.1	
Mar		482.0	447.8	518.7	
Apr		495.0	459.2	533.6	
May		511.8	474.1	552.6	
Jun		586.6	542.5	634.3	
Jul		673.6	622.1	729.5	

Aug	670.8	618.5	727.5
Sep	561.4	516.9	609.7
Oct	499.9	459.6	543.7
Nov	432.3	396.9	470.9
Dec	480.0	440.1	523.5
Jan	501.6	454.7	553.4
Feb	474.2	429.7	523.3
Mar	533.8	483.6	589.2
Apr	548.2	496.6	605.3
May	566.8	513.3	626.0
Jun	649.7	588.2	717.6
Jul	746.1	675.3	824.3
Aug	742.9	672.3	821.0
Sep	621.7	562.5	687.2
Oct	553.7	500.8	612.1
Nov	478.8	433.0	529.5
Dec	531.6	480.6	588.0

 Confidence Limits before Period 144 are 1-Step Ahead Limits

 Confidence

5.3 VENTAJAS Y FACILIDADES.

Hasta este punto se han analizado varios de los muchos paquetes estadísticos que contienen rutinas para realizar el análisis de Series de Tiempo propuesto por Box y Jenkins. Se sabe que existen muchos otros por conocer y analizar y a los cuales no se tuvo acceso, como el SAS versión para PC [15] y las rutinas estadísticas del IMSL [11].

De entre los paquetes que se analizaron y que tenían Series de Tiempo con modelos ARIMA pudo verse que se tiene una gran variedad de parámetros con los que pueden ser comparados, y dependerá principalmente de las necesidades del usuario el decidir cual de ellos cubre mejor sus requerimientos de trabajo.

En esta parte se presenta una comparación de los paquetes revisados basada en los parámetros observados (definidos en el capítulo IV de este trabajo), utilizando un método de análisis multivariado llamado "Electra I" (ver anexo C de este trabajo), para ello, se presentan a continuación las tablas que fueron consideradas.

Es necesario hacer notar que las calificaciones asignadas a los paquetes ESP y LIONHEART se obtuvieron de acuerdo a la evaluación de P. Seybold, L. O'keeffee y J. Klagge como lo exponen en su libro [4].

Para los criterios de los cuales no se obtuvo la información para ciertos paquetes, en vez de asignar calificación de cero, no se tomaron en cuenta dichos criterios para su comparación. Esto se hizo debido a que en los casos de FORECAST PLUS, LIONHEART y ESP no se cuenta con toda la información necesaria no porque no exista, si no porque no se tuvo acceso a ella, no siendo justo por ello bajar la puntuación de estos paquetes.

Las tablas mostradas son:

- 1) Los Criterios a considerar.
- 2) La Tabla Comparativa que contiene las características con las que cuenta cada paquete para cada uno de los criterios considerados.
- 3) La Tabla de Pesos en la que se encuentran los pesos asignados para cada criterio de cada paquete.
- 4) Las Funciones g (por criterio) donde se muestran las calificaciones que se asignaran a cada criterio dependiendo de las características del paquete.
- 5) El Cuadro de Valores Cuantitativos que tienen las calificaciones asignadas a cada paquete en cada en cada criterio.
- 6) Las Tablas de Concordancia y Discordancia arrojadas de las tablas anteriores.

CRITERIOS CONSIDERADOS

1) DOCUMENTACION DEL USUARIO.

- 1.1 ORGANIZACION
- 1.2 SECCION CON EJEMPLOS
- 1.3 SECCION DE INSTALACION Y REQUERIMIENTOS DE HARDWARE
- 1.4 SECCION CON SO Y SINTAXIS DE LOS COMANDOS
- 1.5 DESCRIPCION DE ERRORES
- 1.6 ESPECIFICACION DE TAMAÑO, PRECISION Y VALORES LE DEFAULT
- 1.7 AYUDA (DENTRO DEL MANUAL)
- 1.8 DESCRIPCION DEL PROCEDIMIENTO (ARIMA)
- 1.9 BIBLIOGRAFIA

2) LENGUAJE DE CONTROL.

- 2.1 PANEL DE COMANDOS O MENUS
- 2.2 ASIGNACION DE ETIQUETAS A LAS VARIABLES
- 2.3 TRANSFORMACION DE VARIABLES
- 2.4 AYUDA

3) ESTRUCTURA DE LOS DATOS.

- 3.1 IMPORTACION Y EXPORTACION DE DATOS

4) IMPRESION DE RESULTADOS.

- 4.1 IMPRESION COMPLETA, CLARA Y ESTANDAR
- 4.2 IMPRESION OPCIONAL

5) GRAFICAS.

- 5.1 EXISTENCIA DE ETIQUETAS DENTRO DE LAS GRAFICAS
- 5.2 PODER EFECTUAR CORRECCIONES A LA GRAFICA
- 5.3 PODER GUARDARSE PARA USOS FUTUROS

6) EFECTIVIDAD ESTADISTICA.

- 6.1 IDENTIFICACION
- 6.2 ESTIMACION
- 6.3 PREDICCION

TABLA COMPARATIVA

	SYD PARMATIC	1 2 3	3 A 1 1	1 2 3 A	FORREC-11 PLUS	1 2 3	1 2 3 A
1) DOCUMENTACION DEL USUARIO.							
1.1 ORGANIZACION	MUY BUENA	REGULAR	REGULAR	MUY BUENA		BUENA	MUY BUENA
1.2 EJEMPLOS CON LISTADO	MUY BUENA	MUY BUENA	REGULAR	BUENA		BUENA	REGULAR
1.3 INSTALACION Y REQUERIMIENTOS DE HARDWARE	MUY BUENA	MUY BUENA	NO EXISTE	REGULAR			MUY BUENA
1.4 USOS Y SINTAXIS DE COMANDOS	MUY BUENA	MUY BUENA	BUENA	MUY BUENA		BUENA	MUY BUENA
1.5 DESCRIPCION DE ERRORES	BUENA	MUY BUENA	REGULAR	BUENA DENTRO DEL PAQUETE		BUENA	MUY BUENA
1.6 ESPECIFICACION DE TAMAÑO, FRECUENCIA Y VALORES DE DEFAULT	REGULAR	REGULAR	NO EXISTE	REGULAR			BUENA
1.7 AYUDA EN EL MANUAL	EJEMPLOS QUIDADOS EN EL MANUAL	EJEMPLOS QUIDADOS EN EL MANUAL	EJEM. EN MANUAL	EJEMPLOS QUIDADOS EN EL MANUAL		EJEMPLOS QUIDADOS EN EL MANUAL	NO HAY
1.8 DESCRIPCION DEL PROCEDIMIENTO	FORMULAS ESTAD. Y USO	FORMULAS ESTAD. Y USO	FORMULAS ESTAD. Y USO	FORMULAS ESTAD. Y USO		FORMULAS ESTAD. Y USO	FORMULAS ESTAD. Y USO
1.9 BIBLIOGRAFIA	SOLO ESTADISTICA	SOLO ESTADISTICA	SOLO ESTADISTICA	SOLO ESTADISTICA			SOLO ESTADISTICA
2) LENGUAJE DE CONTROL.							
2.1 PANEL DE COMANDOS O MENUS	MANEJO A BASE DE MENUS	PANEL DE COMANDOS	MANEJO POR COMANDOS DIRECTOS	MANEJO POR MENUS Y COMANDOS	MANEJO POR MENUS Y PREGUNTAS	MANEJO POR COMANDOS DIRECTOS	MANEJO POR COMANDOS DIRECTOS
2.2 ASIGNACION DE ETIQUETAS A LAS VARIABLES	NOMBRE Y COMENTARIOS	NOMBRE Y COMENTARIOS	NOMBRE	NOMBRE		NOMBRE Y COMENTARIOS	NOMBRE Y COMENTARIOS
2.3 TRANSFORMACION DE VARIABLES	SE HACEN EN FORMAS	SE HACEN EN FORMAS	SE HACEN EN FORMAS	SE HACEN EN FORMAS	SE HACEN EN FORMAS	SE HACEN EN FORMAS	SE HACEN EN FORMAS
2.4 AYUDA (EN EL PAQUETE)	BUENA	BUENA	NO EXISTE	MUY BUENA	MUY BUENA	BUENA	BUENA
3) ESTRUCTURA DE LOS DATOS.							
3.1 IMPORTACION Y EXPORTACION DE LOS DATOS	LOTUS-DIF Y ASCII	LOTUS-DIF Y ASCII	LOTUS-DIF Y ASCII	LOTUS, BASE Y ASCII	ASCII Y LOTUS	LOTUS-DIF Y ASCII	LOTUS Y ASCII
4) IMPRESION DE RESULTADOS.							
4.1 IMPRESION COMPLETA, CLARA Y ESTANDAR	IMPRESION MUY BUENA	IMPRESION MUY BUENA	IMPRESION BUENA	IMPRESION BUENA	IMPRESION MUY BUENA	IMPRESION BUENA	IMPRESION BUENA
4.2 IMPRESION OPCIONAL	FOR GRAFICA O RESULTADO	FOR GRAFICA O RESULTADO	FOR GRAFICA O RESULTADO	FOR GRAFICA O RESULTADO	FOR PROCEDIMIENTO		
5) GRAFICAS.							
5.1 EXISTENCIA DE ETIQUETAS	CLARAS	CLARAS	REGULARMENTE CLARAS	REGULARMENTE CLARAS	CLARAS	CLARAS	
5.2 PODER EFECTUAR CORRECCIONES	SI PUEDEN EFECTUARSE	SI PUEDEN EFECTUARSE	SI PUEDEN EFECTUARSE	NO PUEDEN EFECTUARSE	NO PUEDEN EFECTUARSE		NO EXISTE
5.3 PODER GUARDARSE PARA USOS FUTUROS	SI	SI	NO	SI	NO		NO EXISTE
6) EFECTIVIDAD ESTADISTICA.							
6.1 IDENTIFICACION	BUENA Y FACIL	BUENA Y FACIL	BUENA Y NO MUY FACIL	BUENA Y NO MUY FACIL	BUENA Y FACIL	BUENA Y FACIL	BUENA Y FACIL
6.2 ESTIMACION	NO PUEDE COMPLETARSE	FACIL Y NO MUY BUENA	NO PUEDE COMPLETARSE	BUENA Y NO MUY FACIL	BUENA Y FACIL	BUENA Y FACIL	BUENA Y FACIL
6.3 PREDICCION	BUENA Y FACIL	BUENA Y FACIL	BUENA Y NO MUY FACIL	BUENA Y NO MUY FACIL	BUENA Y FACIL	BUENA Y FACIL	BUENA Y FACIL

P E S O S

1) DOCUMENTACION DEL USUARIO.		
1.1	ORGANIZACION	2
1.2	SECCION CON EJEMPLOS	3
1.3	SECCION DE INSTALACION Y REQUERIMIENTOS DE HARDWARE	2
1.4	SECCION CON SO Y SINTAXIS DE LOS COMANDOS	3
1.5	DESCRIPCION DE ENORES	3
1.6	ESPECIFICACION DE TAMAO, PRECISION Y VALORES DE DEFAULT	1
1.7	AYUDA (DENTRO DEL MANUAL)	2
1.8	DESCRIPCION DEL PROCEDIMIENTO (ARIMA)	4
1.9	BIBLIOGRAFIA	1
2) LENGUAJE DE CONTROL.		
2.1	PANEL DE COMANDOS O MENUS	3
2.2	ASIGNACION DE ETIQUETAS A LAS VARIABLES	1
2.3	TRANSFORMACION DE VARIABLES	3
2.4	AYUDA	6
3) ESTRUCTURA DE LOS DATOS.		
3.1	IMPORTACION Y EXPORTACION DE DATOS	8
4) IMPRESION DE RESULTADOS.		
4.1	IMPRESION COMPLETA, CLARA Y ESTANDAR	6
4.2	IMPRESION OPCIONAL	2
5) GRAFICAS.		
5.1	EXISTENCIA DE ETIQUETAS DENTRO DE LAS GRAFICAS	3
5.2	PODER EFECTUAR CORRECCIONES A LA GRAFICA	3
5.3	PODER GUARDARSE PARA USOS FUTUROS	2
6) EFECTIVIDAD ESTADISTICA.		
6.1	IDENTIFICACION	14
6.2	ESTIMACION	14
6.3	PREDICCION	14

FUNCIONES g (Por criterio)

1) DOCUMENTACION DEL USUARIO.

1.1 ORGANIZACION

MUY BUENA	10
BUENA	8
REGULAR	6
MALA	4

1.2 SECCION CON EJEMPLOS

MUY BUENA	10
BUENA	8
REGULAR	6
MALA	4
NO EXISTE	0

1.3 SECCION DE INSTALACION Y REQUERIMIENTOS DE HARDWARE

MUY BUENA	10
BUENA	8
REGULAR	6
MALA	4
NO EXISTE	0

1.4 SECCION CON USO Y SINTAXIS DE LOS COMANDOS

MUY BUENA	10
BUENA	8
REGULAR	6
MALA	4
NO EXISTE	0

1.5 DESCRIPCION DE ERRORES

MUY BUENA	10
BUENA	8
REGULAR	6
MALA	4
NO EXISTE	0

1.6 ESPECIFICACION DE TAMAÑO, PRECISION Y VALORES DE DEFAULT

MUY BUENA	10
BUENA	8
REGULAR	6
MALA	4
NO EXISTE	0

1.7 AYUDA (DENTRO DEL MANUAL)

EJEMPLOS GUIADOS - DISSETT O MANUAL - MOSTRANDO SALIDAS Y ENTRADAS DEL PAQUETE	10
EJEMPLOS GUIADOS SIN MOSTRAR PANTALLAS DEL PAQUETE	7
NO EXISTE	0

1.8 DESCRIPCION DEL PROCEDIMIENTO

FORMULAS UTILIZADAS - PROCES. NUMERICOS Y ESTADISTICOS - Y UTILIZACION DEL PAQUETE	10
FORMULAS UTILIZADAS - ESTADISTICAS - Y UTILIZACION	8
SOLO UTILIZACION DEL PAQUETE	6
NO EXISTE	0

1.9 BIBLIOGRAFIA

ESTADISTICA Y DE PROCESIMIENTOS NUMERICOS	10
ESTADISTICA SOLAMENTE	8
PROCEDIMIENTOS NUMERICOS SOLAMENTE	6
NINGUNA	0

2) LENGUAJE DE CONTROL

2.1 PANEL DE COMANDOS O MENUS

MANEJO A BASE DE MENUS O PREGUNTA-RESPUESTA	10
PANEL DE COMANDOS UNICAMENTE	8
MANEJO A BASE DE COMANDOS -SIN PANELES-	6

2.2 ASIGNACION DE ETIQUETAS A LAS VARIABLES

PUEDEN ASIGNARSE NOMBRES Y COMENTARIOS	10
PUEDEN ASIGNARSE SOLAMENTE	6
NO PUEDEN ASIGNARSE NADA	3

2.3 TRANSFORMACION DE VARIABLES

SE ESCOGEN DE UN MENU Y PUERE HACERSE EN FORMA MANUAL.	10
SON UN NUMERO CONSIDERABLE	9
SE ESCOGEN DE UN MENU Y SON UN NUMERO CONSIDERABLE	9
SE HACE EN FORMA MANUAL Y SON UN NUMERO CONSIDERABLE	8
SE ESCOGEN DE UN MENU PERO EXISTEN POCAS OPCIONES	6
SE HACE EN FORMA MANUAL Y HAY TODAS OPCIONES	5
NO PUEDEN EFECTUARSE TRANSFORMACIONES	0

2.4 AYUDA

MUY BUENA	10
BUENA	6
REGULAR	5
MALA	4
NO EXISTE	0

3) ESTRUCTURA DE LOS DATOS

3.1 IMPORTACION Y EXPORTACION DE DATOS

IMPORTA Y EXPORTA ARCHIVOS ASCII, LOTUS Y OTROS	10
IMPORTA Y EXPORTA SOLO ARCHIVOS LOTUS Y ASCII	9
SOLO IMPORTA O EXPORTA ARCHIVOS ASCII -UNA DE LAS DOS OPCIONES-	5
SOLO CUENTA CON EL EDITOR DEL PAQUETE	2

4) IMPRESION DE RESULTADOS

4.1 IMPRESION COMPLETA, CLARA Y ESTANDAR

IMPRESION MUY BUENA	10
IMPRESION BUENA	8
IMPRESION REGULAR	6
IMPRESION MALA	4
NO PUEDE IMPRIMIRSE	0

4.2 IMPRESION OPCIONAL

ES OPCIONAL -POR Grafica o TABLA DE RESULTADOS-	10
ES OPCIONAL -POR PROCEDIMIENTO CONJUNTO-	7
NO ES OPCIONAL	3

5) GRAFICAS

5.1 EXISTENCIA DE ETIQUETAS (CLARIDAD)

CLARAS	10
REGULARMENTE CLARAS	7
NO CLARAS	4

5.2 PODER EFECTUAR CORRECCIONES A LA GRAFICA

SE PUEDEN EFECTUAR	10
NO SE PUEDEN EFECTUAR	5

5.3 PODER SALVARSE PARA USOS FUTUROS

SI	10
NO	5

6) EFECTIVIDAD ESTADISTICA

6.1 IDENTIFICACION

BUENA Y FACIL	10
BUENA Y NO FACIL	7
MALA Y FACIL	7
MALA Y NO FACIL	5
NO PUERE HACERSE COMPLETA	3

6.2 ESTIMACION

BUENA : FACIL	10
BUENA : NO FACIL	1
MALA : FACIL	7
MALA : NO FACIL	1
NO PUEDE HACERSE COMPLET.	3

6.3 PREDICCION

BUENA : FACIL	10
BUENA : NO FACIL	9
MALA : FACIL	2
MALA : NO FACIL	1
NO PUEDE HACERSE COMPLET.	3

VALORES COMPARATIVOS

	STAT 89-2-173	T S I	S A T S	STAT	STAT 2.1	T S I	STAT
1) DOCUMENTACION DEL USUARIO.							
1.1 ORGANIZACION	10	4	4	10	--	2	10
1.2 EJEMPLOS CON LISTADO	10	10	6	6	--	3	6
1.3 INSTALACION Y REQUERIMIENTOS DE HARDWARE	10	10	8	6	--	--	10
1.4 USOS Y SINTAXIS DE COMANDOS	10	10	8	10	--	2	10
1.5 DESCRIPCION DE ERRORES	6	10	6	6	--	6	10
1.6 ESPECIFICACION DE TAMANO, FRASES Y VALORES DE DEFAULT	6	6	2	6	--	--	6
1.7 AYUDA EN EL MANUAL	10	10	10	10	--	10	0
1.8 DESCRIPCION DEL PROCEDIMIENTO	9	9	9	9	--	9	9
1.9 BIBLIOGRAFIA	8	8	8	6	--	--	8
2) LENGUAJE DE CONTROL.							
2.1 PANEL DE COMANDOS O MENUS	10	6	6	9	10	6	6
2.2 ASIGNACION DE ETIQUETAS A LAS VARIABLES	10	10	8	6	10	6	10
2.3 TRANSFORMACION DE VARIABLES	10	8	8	10	8	8	8
2.4 AYUDA (EN EL PAQUETE)	8	8	8	10	10	8	--
3) ESTRUCTURA DE LOS DATOS.							
3.1 IMPORTACION Y EXPORTACION DE LOS DATOS	10	10	10	10	9	10	9
4) IMPRESION DE RESULTADOS.							
4.1 IMPRESION COMPLETA, CLARA Y ESTANDAR	10	10	8	9	10	6	9
4.2 IMPRESION OPCIONAL	10	10	10	10	7	--	--
5) GRAFICAS.							
5.1 EXISTENCIA DE ETIQUETAS	10	10	7	8	10	10	8
5.2 PODER EFECTUAR CORRECCIONES	10	10	10	6	6	--	9
5.3 PODER GUARDARSE PARA USOS FUTUROS	10	10	6	10	6	--	8
6) EFECTIVIDAD ESTADISTICA.							
6.1 IDENTIFICACION	10	10	9	9	10	5	10
6.2 ESTIMACION	9	9	9	9	10	9	10
6.3 PREDICCION	10	10	9	9	10	9	10

INDICES DE CONCORDANCIA

	STAT GRAPHICS	T S P	BATS	SYSTAT	FORECAST PLUS	E S P	LION HEART
STAT GRAPHICS	---	.83	1	.90	.59	.75	.74
T S P	.86	---	1	.72	.56	.67	.75
BATS	.34	.25	---	.55	.18	.44	.38
SYSTAT	.51	.54	.97	---	.27	.89	.46
FORECAST PLUS	.61	.64	.67	.64	---	.64	.71
E S P	.40	.42	.81	.72	.24	---	.26
LION HEART	.49	.62	.74	.65	.54	.64	---

INDICES DE DISCORDANCIA

	STAT GRAPHICS	T S P	BATS	SYSTAT	FORECAST PLUS	E S P	LION HEART
STAT GRAPHICS	---	.5	0	.6	.7	.6	.7
T S P	.4	---	0	.5	.5	.3	.4
BATS	1	1	---	1	1	.8	1
SYSTAT	.4	.4	.4	---	.3	.2	.4
FORECAST PLUS	.4	.4	.4	.4	---	.1	0
E S P	.4	.4	.2	.3	.4	---	.2
LION HEART	1	1	.4	1	1	1	---

Con los datos anteriores pueden efectuarse el análisis de sensibilidad:

Primero se dará a p el valor de 0.9 y a q el de 0.1, para los cuales dentro de las matrices de concordancia y discordancia se obtiene que RATS es superado por STATGRAPHICS y TSP

```
STATGRAPHICS S RATS
TSP S RATS
```

de donde se obtienen los conjuntos,

$E = \{STATGRAPHICS, TSP, SYSTAT, FORACAST PLUS, ESP, LION HEART\}$

$E^c = \{RATS\}$.

Ahora, si se mantiene fija a q en un valor alto (0.5) y se varía el valor de p , con el fin de dar mayor importancia a la matriz de concordancia como criterio para seleccionar los paquetes sin dar tanta importancia a que uno de ellos mejor que otro en la mayoría de los criterios, puede estar muy bajo en uno de los criterios del paquete que aparentemente superó.

Para $p = 0.8$, $q = 0.5$

```
TSP S STATGRAPHICS
STATGRAPHICS S TSP
SYSTAT S RATS
ESP S RATS
SYSTAT S ESP
STATGRAPHICS S RATS
TSP S RATS
```

$E = \{STATGRAPHICS, TSP, FORACAST PLUS, ESP, SYSTAT, LION HEART\}$

$E^c = \{RATS\}$.

Para $p = 0.75$, $q = 0.5$

```
TSP S STATGRAPHICS
STATGRAPHICS S TSP
SYSTAT S RATS
ESP S RATS
SYSTAT S ESP
STATGRAPHICS S RATS
TSP S RATS
TSP S LION HEART
```

$E = \{STATGRAPHICS, TSP, FORACAST PLUS, ESP\}$

$E^c = \{RATS, SYSTAT, LION HEART\}$.

Para $p = 0.70$, $q = 0.5$

```
TSP S STATGRAPHICS
STATGRAPHICS S TSP
```

```

SYSTAT S RATS
ESP S RATS
SYSTAT S ESP
STATGRAPHICS S RATS
TSP S RATS
TSP S LION HEART
ESP S TSP
LION HEART S RATS
FORECAST PLUS S LION HEART
STATGHAPHICS S LION HEART
TSP S SYSTAT
ESP S SYSTAT

```

$E = (\text{STATGHAPHICS}, \text{TSP}, \text{FORACAST PLUS}, \text{ESP})$

$E^c = (\text{RATS}, \text{SYSTAT}, \text{LION HEART})$.

Para $p = 0.65$, $q = 0.5$

```

TSP S STATGRAPHICS
STATGRAPHICS S TSP
SYSTAT S RATS
ESP S RATS
SYSTAT S ESP
STATGRAPHICS S RATS
TSP S RATS
TSP S LION HEART
ESP S TSP
LION HEART S RATS
FORECAST PLUS S LION HEART
STATGHAPHICS S LION HEART
TSP S SYSTAT
ESP S SYSTAT
ESP S STATGRAPHICS
FORECAST PLUS S RATS
TSP S ESP
ESP S LION HEART

```

$E = (\text{TSP}, \text{FORACAST PLUS})$

$E^c = (\text{STATGHAPHICS}, \text{RATS}, \text{SYSTAT}, \text{LION HEART}, \text{ESP})$.

Si ahora se desea dar una mayor importancia a la diferencia máxima entre los valores de los criterios de cada paquete, se fija el valor de p en un valor bajo (0.5) y se varía el valor de q .

Para $p = 0.5$, $q = 0.2$

```

ESP S RATS
TSP S RATS
STATGRAPHICS S RATS
FORECAST PLUS S ESP
SYSTAT S ESP
FORECAST PLUS S LION HEART

```


E = (STATGRAPHICS, TSP, FORACAST PLUS, SYSTAT, LION HEART)
 E^o = (RATS, ESP).

Para $p = 0.5$, $q = 0.3$

```

ESP S RATS
TSP S RATS
STATGRAPHICS S RATS
FORECAST PLUS S ESP
SYSTAT S ESP
FORECAST PLUS S LION HEART
ESP S SYSTAT
TSP S ESP

```

E = (STATGRAPHICS, TSP, FORACAST PLUS, LION HEART)
 E^o = (RATS, ESP, SYSTAT).

Para $p = 0.5$, $q = 0.4$

```

ESP S RATS
TSP S RATS
STATGRAPHICS S RATS
FORECAST PLUS S ESP
SYSTAT S ESP
FORECAST PLUS S LION HEART
ESP S SYSTAT
TSP S ESP
TSP S STATGRAPHICS
SYSTAT S STATGRAPHICS
FORECAST PLUS S STATGRAPHICS
SYSTAT S TSP
FORECAST PLUS S TSP
SYSTAT S RATS
FORECAST PLUS S RATS
LION HEART S RATS
FORECAST PLUS S SYSTAT
TSP S LION HEART
SYSTAT S LION HEART

```

E = (FORACAST PLUS)
 E^o = (RATS, ESP, SYSTAT, STATGRAPHICS, TSP, LION HEART).

Puede verse que los resultados en los tres casos no coinciden, pero teniendo en cuenta después de haber trabajado con cada paquete, que aunque algunos eran buenos en muchos aspectos, no podía completarse el análisis de la serie satisfactoriamente, o viceversa, por lo que los parámetros p y q serán tomados de tal forma que se de más importancia a la diferencia máxima entre los valores de los criterios. Para ello se proponen los siguientes valores.

Para $p = .65$, $q = .4$

```

TSP $ STATGRAPHICS
STATGRAPHICS $ RATS
ESP $ RATS
TSP $ RATS
SYSTAT $ RATS
FORECAST PLUS $ RATS
LION HEART $ RATS
ESP $ SYSTAT
TSP $ ESP
TSP $ LION HEART
FORECAST PLUS $ LION HEART
SYSTAT $ ESP

```

E = (FORACAST PLUS, SYSTAT, TSP)
E = (RATS, ESP, STATGHAPHICS, LION HEART).

Así el núcleo queda compuesto de tres paquetes en los cuales puede efectuarse el análisis de Series de Tiempo con los modelos ARIMA.

VI CONCLUSIONES.

6.1 CONCLUSIONES ACERCA DEL SOFTWARE.

De los ocho paquetes que contenían rutinas de análisis de Series de Tiempo con modelos ARIMA, estuvieron dentro del núcleo los paquetes de TSP, SYSTAT y FORECASTPLUS (de ESP y LIONHEART no se conocen bien sus funciones) en los cuales puede efectuarse todo el procedimiento tal y como lo proponen Box y Jenkins, pudiendo identificar el modelo por medio de gráficas de los datos con o sin diferencias, las funciones de autocorrelación y autocorrelación parcial, pudiendo también manejar todos los parámetros para estimar un modelo, así como utilizar la predicción hacia atrás (excepto para el caso de TSP), para una mejor aproximación y poder predecir habiendo diagnosticado el modelo con ayuda de las funciones de autocorrelación de los residuales. Sin embargo, si se deja a un lado la versatilidad se obtendría que tomando en cuenta los demás atributos, sobresaldrían los paquetes de STATGRAPHICS y FORECAST PLUS debido a su excelente documentación (STAGRAPHICS) y a su buena interacción con el usuario por a su manejo a base de menús.

Por otro lado, el paquete superado por todos los demás fue el de RATS, ya que no ofrece ni la posibilidad de tener parámetros estacionales en el modelo, es poco interactivo con el usuario y tampoco ofrece una documentación clara que permita hacer un buen uso del paquete.

Realmente STATGRAPHICS de no ser por el error encontrado que se espera sea corregido en versiones posteriores (que ya se encuentran en el mercado y a las que no se tuvo acceso), es un buen paquete en todos sentidos, su documentación permite un uso completo de los procesos de manejo de datos, gráficas y modelos ARIMA que fué lo que se

observó principalmente, siendo además sumamente interactivo con el usuario.

TSP por ejemplo, cuenta con un tutorial que permite dar confianza al usuario para manejar el paquete, aunque su documentación no es excelente, siendo además de gran ayuda el panel mostrado en la parte inferior de la pantalla.

Para el caso de FORECAST PLUS una deficiencia son sus gráficos, ya que no es posible analizarlos en pantalla por su rápido despliegue, y sólo pueden verse en papel o en un archivo externo el que puede accederse estando fuera del paquete, por otro lado, puede desarrollarse perfectamente el análisis completo, sin tener que invertir mucho tiempo en aprender el funcionamiento del paquete, debido a que para obtener los resultados sólo es necesario llenar paneles y contestar preguntas haciendo el análisis todavía más rápido, ofreciendo además todos los recursos para para identificar, estimar y predecir las Series de Tiempo.

De SYSTAT, puede decirse que es muy potente, permitiendo efectuar todo el análisis, aunque en la presentación de gráficos y resultados no sea excelente ni tan interactivo con el usuario como otros paquetes, cuenta con una documentación que permite (aunque no en forma óptima) el manejo adecuado del programa.

Es importante hacer notar que los pesos asignados son subjetivos e indican la reacción personal que se tuvo ante los diversos paquetes y no es necesariamente la reacción que pudieran tener otras personas, sin embargo se espera sirva como una orientación para los posibles usuarios.

Box y Jenkins al analizar la serie que se utilizó para ejemplificar a cada paquete (Serie G de su libro, ver anexo A de este trabajo) obtuvieron ciertos resultados finales al efectuar su estimación, dichos resultados son mostrados en su libro [1], utilizando predicciones hacia atrás (back forecast), éstos resultados son.

Parámetro	Estimación
MA	0.396
SMA	0.614

Es decir, el modelo quedó:

$$(d\text{if}^1)(D\text{IF}_{12}^1)(1n Z_t) = (1-.396B)(1-.614B^2)a_t$$

Donde $d\text{if}^a$ = diferencia de grado a no estacional.
 $D\text{IF}_{n,m}$ = diferencia de grado m estacional con n periodos.

Los resultados para los paquetes analizados, en los que se pudo efectuar el análisis son los siguientes:

NOMBRE DEL PAQUETE	MA	SMA	BACK FORECAST
SPSS	0.39531	0.61406	SI
TSP	- 0.3776806	- 0.6523184	NO
FORECAST PLUS	0.3901	0.5628	SI
RATS	---	---	
STATGRAPHICS	---	---	
SYSTAT	0.396	0.613	SI

Los resultados de SPSS fueron tomados del libro "SPSS Update 7-9, New Procedures and Facilities for Release 7-9" de Hadlai y Norman [7] donde se muestra el procedimiento de como hace SPSS el análisis de Series de Tiempo con modelos ARIMA, utilizando la misma serie que se mostró para todos los demás paquetes y utilizando además, la versión para Burroughs y no para PC's. Como se utilizó una máquina grande puede verse que la aproximación es muy buena al menos para este caso en particular.

Por otro lado, TSP deja unos resultados "parecidos" en valor absoluto pero con el signo invertido, (es decir, multiplicados por -1) sólo para los parámetros de promedios móviles estacionales y no estacionales. Es de suponerse que como para el modelo propuesto por Box y Jenkins los signos son puestos por convención (ver sección 2.3.2 de este trabajo), los autores de TSP invirtieron el signo y consideraron así el modelo. Detalles como éstos es necesario conocer antes de utilizar el paquete, ya que pueden llevar al usuario a resultados erróneos. Esto podría evitarse si en el manual se mostrara la fórmula considerada para efectuar el análisis.

Habiendo observado lo anterior, se espera que el software estadístico para PC pueda mejorar poniendo atención en el diseño del programa en las siguientes áreas, entre otras:

- Mayor (y más flexible) uso de archivos externos.
- Mayor flexibilidad en la entrada de datos.
- Procedimientos de ejemplos guiados accesibles.
- Funciones de ayuda sensibles al contexto.
- Documentación basada en ejemplos que muestren la adecuada organización de datos y la sintaxis de los órdenes.
- Operaciones más robustas de modo que no se expulse del programa si hay un error o si solamente se desea cambiar de función.
- Explicaciones sobre los mensajes de error y sugerencias de como corregirlos.
- Capacidad para interrumpir o cancelar operaciones.
- Capacidad para almacenar secuencias de órdenes para realizar tipos comunes de análisis.

6.2 EL CUIDADO DE LOS RESULTADOS.

Hasta ahora se vió el cómo, por medio de diversos paquetes estadísticos, puede facilitarse el análisis de Series de Tiempo propuesto por Box y Jenkins. Es obvio que hacer todas las operaciones por medio de hojas de cálculo (o el utilizar calculadora, en el peor de los casos) multiplicaría el tiempo de ejecución de un análisis de este tipo, además, de que podría caerse en errores en los cálculos más fácilmente. Es por esto, que el contar con paquetes estadísticos que efectúen los cálculos, ayuda de gran manera simplificando este trabajo, reduciendo así el análisis sólo a tomar decisiones sobre los resultados obtenidos, con esto se quiere decir que es el usuario quien después de observar la serie de datos, las funciones de autocorrelación y autocorrelación parcial (calculadas y graficadas por el paquete) decidirá los términos autoregresivos y de promedios móviles, estacionales y no estacionales que deberá contener el modelo, así como si es necesario obtener diferencias en caso de considerar a la serie no estacionaria. Además, una vez estimados los parámetros del modelo (por el paquete) y analizando los residuales, decidirá si se queda con ese modelo o no.

En resumen, el usuario decidirá el modelo a utilizar para predecir los valores de su serie, consiguiendo con ayuda del software apropiado rapidez en los cálculos, una presentación de resultados claros y completos con formatos estandares y la eliminación de cierto tipo de errores.

Es por esto que el usuario debe conocer la técnica estadística para poder evaluar los resultados del paquete, es recomendable entonces, para los estudiantes de esta materia el revisar con cuidado los procedimientos utilizados, y de ser posible, antes de utilizar los paquetes el seguir de manera cuidadosa cada paso del proceso, ya sea efectuar los cálculos paso a paso, o bien, programando mediante lenguajes estadísticos que facilitan ciertas rutinas como sumatorias o cálculo de medias muestrales.

Predicción.

Cuando se ha obtenido una predicción de la serie es importante después de realizar el diagnóstico de residuales y estar de acuerdo con éste, tener en cuenta que al efectuar el análisis se están asumiendo diversos supuestos como el de estacionariedad de la serie, esto es, que un cambio brusco en la serie no podrá predecirse por el modelo estimado si nunca antes dentro de la serie de observaciones se había presentado.

Por esto es importante notar que se está utilizando un modelo que representa un patrón dejado por las observaciones disponibles de una serie y que se desea con esta base efectuar predicciones. Así, el modelo obtenido es

el que se cree aproxima a la serie, pudiendo haber tenido errores dentro del proceso al identificar o estimar los parámetros.

Entonces, aunque se efectuen predicciones suponiendo que se tiene el modelo correcto, estará claro que en la práctica dependerá en parte, de la adecuación del modelo propuesto y en parte, del cómo se efectuaron los pasos anteriores de identificación y estimación.

ANEXO A.

SERIE G PASAJEROS DE UNA AEROLINEA INTERNACIONAL:
TOTALES MENSUALES (MILES DE PASAJEROS) ENERO
1949 - DICIEMBRE 1960

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

ANEXO B.

TIPOS DE ARCHIVOS DE DATOS.

ASCII.

Este tipo de archivos no tienen una extensión única, se diferencian por contener sólo caracteres de tipo ASCII y no contener caracteres de control como son los que dejan ciertos procesadores de palabra u hojas de cálculo útiles para producir cierto tipo de formatos o funciones.

LOTUS O DBASE.

Los archivos producidos por LOTUS (hoja de cálculo) o DBASE (base de datos) guardan ciertas características de formato o función como caracteres de control. LOTUS deja sus archivos con extensión .PRN y DBASE con extensión .DBF.

DIF (DATA INTERCHANGE FORMAT).

Son utilizados para canjear datos entre ciertos programas como SuperCalc o VisiCalc por ejemplo. Un archivo DIF es autodocumentado, es decir, contiene información acerca de las series, número de observaciones y otros objetos (como títulos) y dentro del mismo archivo.

ANEXO C.

METODO ELECTRA I.

Los atributos deseables en un paquete estadístico se manifiestan mediante una lista de criterios, los cuales pueden ser cuantificables, no cuantificables (calidad de gráficos y resultados), dicotómicos (se tiene o no, por ejemplo, manual del usuario, impresión de resultados) y hasta criterios que pueden ser contradictorios y complejos. El decisor asigna a cada criterio un valor y de acuerdo a éste, de alguna forma guía su decisión.

En el presente anexo se expone el método llamado Electra I [20], cuya finalidad es apoyar la decisión en problemas de análisis multicriterio como es la selección de paquetes estadísticos.

Este método fue desarrollado por la Escuela Francesa de ayuda a la decisión y se diferencia de la mayoría de los modelos de decisión conocidos, por ésta doble particularidad:

- Permite tomar en consideración dentro del análisis diversos criterios de evaluación.
- Permite igualmente tomar en consideración la evaluación subjetiva.

El método Electra I se fundamenta en las comparaciones por pareja (dos a dos) de las variables o alternativas a considerar.

El método se vale de la construcción de una relación, no necesariamente transitiva, de "superación" que ayuda a comparar por pareja las alternativas a seleccionar.

A continuación se describe:

Objetivo.

Obtener una participación del conjunto de posibles soluciones A (finito) en dos subconjuntos E y E^c tales que:

- Toda acción de E^c es superada por al menos una acción de E .
- Las acciones de E no son comparables entre sí.

La relación de superación se basa en lo siguiente, a supera a b si:

- De entre los criterios adoptados se obtiene una mayoría suficiente para considerar que a no es "menos buena" que b .
- Ningún criterio en desacuerdo con ésta mayoría no revela una superioridad bastante fuerte de b con respecto a a .

Algunas Definiciones

Criterio.

Es un punto de vista que se considera para la comparación de los elementos del conjunto A de posibles soluciones.

Tomase la función g definida en el conjunto A la cual toma valores en los reales.

Se dice que a es mejor que b según un determinado criterio si y sólo si $g(a) > g(b)$.

Dominancia.

Dado un conjunto de criterios g_1, g_2, \dots, g_n , una acción a domina a otra b , si y sólo si

$$g_i(a) \geq g_i(b) \quad \text{para } i = 1, 2, \dots, n$$

siendo por lo menos alguna de las n desigualdades estricta.

Se denota como C_{ab} al conjunto de criterios para los cuales $g_i(a) \geq g_i(b)$.

Por otro lado D_{ab} será el conjunto de criterios para los cuales $g_i(a) < g_i(b)$.

El Método.

1) Seleccionar los elementos del conjunto de posibles soluciones (A).

Seleccionar los criterios para la comparación de los elementos de A .

2) Asignar ponderaciones (denotadas como p_i) a cada uno de los criterios adoptados. Mientras más grande sea el peso asignado a un criterio con relación a los demás, éste es más importante.

Un valor diferente de cero e igual en ciertos criterios será útil en el caso de que no se desee establecer alguna diferencia entre los mismos.

En este punto, el decisor manifiesta de acuerdo a sus intereses la inclinación por cada uno de los criterios.

3) Para cada criterio se define una función g_i que cuantificará las modalidades de los diferentes criterios.

Deberá considerarse que los valores de las funciones definidas sobre los criterios puedan compararse de criterio a criterio

Deberá considerarse también que los valores asignados a las diferentes modalidades de un criterio sean consistentes.

4) Calcular los índices de concordancia. Este será la base para medir la mayoría suficiente de criterios a favor de una determinada alternativa sobre otra.

El índice se calcula de la siguiente forma:

$$C(a, b) = \frac{\text{Suma de los pesos de los criterios en los cuales } a \text{ es preferido o indiferente a } b}{\text{Suma de todos los pesos de los criterios considerados en el problema.}}$$

para $a \neq b$

Puede observarse lo siguiente, acerca del valor del índice de concordancia.

- Varía entre cero y uno de manera no decreciente cuando el conjunto $C_{a,b}$ aumenta en cardinalidad.
- Es igual a 1 si y sólo si a domina a b .
- Conserva su significado y no conduce a ninguna incoherencia cuando un criterio se subdivide en varios criterios que lo reemplacen o cuando es agregado a otros para formar uno que lo considere. Desde luego se tendrá que tomar en cuenta la consistencia entre el valor de los pesos asignados a los nuevos criterios y el valor que tenían los criterios anteriores.

Para cada pareja de alternativas, se calculan dos índices de concordancia, obteniéndose así la matriz de concordancia.

Obsérvese que $C(a, b)$ no es necesariamente igual a $C(b, a)$.

1) Calcular los índices de discordancia. Este será la base para medir que tan importante es la superioridad de una alternativa con respecto a otra, considerando para esto él o los criterios que estén en desacuerdo con la mayoría y en los que b domina a a .

El índice se calcula de la siguiente forma:

$$D(b, a) = \begin{cases} 1 / \delta_i \in D_{a,b} \text{ MAX } |g_i(b) - g_i(a)| & \text{si } D_{a,b} = 0 \\ 0 & \text{si } D_{a,b} \neq 0 \end{cases}$$

donde δ es la máxima diferencia observada entre los valores extremos de g sobre un mismo criterio considerando el conjunto total de criterios.

Este índice presenta propiedades análogas a las del índice de concordancia.

Para cada pareja de alternativas se calculan los índices de discordancia obteniéndose la matriz de discordancia.

2) Se define la relación de superación S. Se dice que a supera a b, $a S b$, si y sólo si:

$$C(b,a) \geq p \quad (1)$$

$$D(b,a) \leq q \quad (2)$$

donde

- $C(b,a)$ y $D(b,a)$ se definieron los puntos 4 y 5 respectivamente
- p es un valor entre 0 y 1 preferentemente cercano a 1.
- q es un valor entre 0 y 1 preferentemente cercano a 0.

NOTA: A través de (1) se define la *mayoría suficiente* de puntos de vista o criterios a favor de una determinada alternativa. A través de (2), la importancia de la *superioridad* de una alternativa con respecto a otra considerando él o los criterios que están en desacuerdo con la mayoría.

1) Definir E y E^c . Para establecer los elementos de E y E^c tenemos lo siguiente:

- $\forall b \in E^c$, existe $a \in E$ tal que $a S b$
- $\forall a, b \in E$, $a S b$ y $b S a$

El conjunto E se denomina núcleo de la red construida en base a la relación S y en él se encuentra la selección.

OBSERVACIONES.

- La red construida en base a la relación S admite un núcleo único si ésta red no contiene ciclos.
- En la práctica se aconseja hacer variar los parámetros p y q, y observar la sensibilidad del resultado con respecto a estas variaciones.
- El decisor, al tener la posibilidad de variar los parámetros p y q tiene un tercer nivel en el cual puede manifestar sus inclinaciones.

ANEXO D.

BIBLIOGRAFIA.

- [1] G.E.P. Box, Madison; G.M. Jenkins, Lancaster
"Time Series Analysis: Forecasting and Control"
Holden-Day Inc.
1969, USA.

- [2] Charles R. Nelson
"Applied Time Series Analysis for Managerial Forecasting"
Holden-Day Inc.
1973, San Francisco, USA.

- [3] Bruce L. Bowerman; Richard T. O'Connell
"Forecasting and Time Series"
North Scituate, Mass.
Duxbury 1979.

- [4] Patricia B. Seybold; Linda D'Keefe; Jay Klagge
"Paquetes Estadísticos para la Familia IBM PC y Compatibles"
McGraw Hill
1987, España (Traducción al Español)

- [5] Ivor Francis; Richard M. Heiberger; Paul F. Vallerman
"Statistical Computing"
Artículo: "Criteria and Considerations in the Evaluation of Statistical Program Packages"
Vol. 29, No. 1, Pag 52-56
The American Statistician
1975, USA.

- [6] Kurt Enslein; Anthony Ralston; Herbert S. Wilf
"Statistical Methods for Digital Computers" Vol. 3
Artículo: David w. Bacon; Louis H. Broekhoven
"Time Series Forecasting"
Ontario, Canada; Queen's University
Wiley-Interscience (John Wiley & Sons, Inc.)
1977, USA.

- [7] C. Hadlai Hull; Norman H. Nie
"SPSS Update 7-9, New Procedures and Facilities for Release 7-9"
McGraw Hill
1981, USA.

- [8] Norman H. Nie
"Statistical Package for the Social Sciences"
McGraw Hill
1975, USA.

- [9] Burroughs Corporation
"BASIS Reference Manual"
1975, Detroit.

- [10] Tracy L. Gustafson, M.D.
EPISTAT, Manual (contenido en diskette)
1984, USA.

- [11] IMSL Inc.
IMSL Library, Reference Manual
Vol. 2, Cap. F
1980, Texas, USA.

- [12] Walter R. Harvey
Manual LSMLMW
Harvey
1987, USA.

- [13] Jerry L. Hintza
Manual NUMBER CRUNCHER STATISTICAL SYSTEM
IBM
1984, USA.

- [14] Thomas A. Doan; Robert B. Litterman
Manual RATS
VAR Econometrics
Minneapolis, USA.

- [15] SAS Corporation
SAS Language Guide for Personal Computers
USA.

- [16] STSC Company
Manual STATGRAPHICS
Primera versión, USA.

- [17] Quantitative Micro Software
Micro TSP User's Manual
McGraw Hill
Los Angeles.

- [18] Norman H. Nie, C. Hadlai Hull, Jean G. Jenkins, Karlo
Steinbrenner, Dale H. Bent.
SPSS Manual Reference
McGraw Hill
1975, USA.

- [19] Henry Simpson
"A Human-Factors Style Guide for Program Design"
BYTE Publication Inc.
1982, USA.

[20] García G. Javier
"La Selección de Equipos de Cómputo Utilizando un
método de ayuda a la decisión multicriterio"
Abril 1987, México.