

881201

14



ESCUELA DE ACTUARIA
UNIVERSIDAD ANAHUAC

Con Estudios Incorporados a la
U. N. A. M.

**"Pronósticos Ponderados: Un Nuevo Método
para Completar Series de Tiempo".**

TESIS PROFESIONAL

Que para Optar por el Título de:

ACTUARIO

Presenta el Alumno:

Luis Miguel Rodríguez Penagos

TESIS CON
FALLA DE CR.GEN

MEXICO, D. F.

1988.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

INDICE.

PROLOGO

I. INTRODUCCION AL PROBLEMA Y A SUS SOLUCIONES PREVIAS.....	1
1.1 INTRODUCCION.....	2
1.1.1 BREVE INTRODUCCION AL ESTUDIO DE LAS SERIES DE TIEMPO.....	2
1.1.2 PLANTEAMIENTO DEL PROBLEMA.....	7
1.2 ALGUNOS TRATAMIENTOS ANTERIORES.....	9
1.2.1 ESTIMACION DE SERIES INCOMPLETAS.....	9
1.2.2 ESTIMACION DE OBSERVACIONES FALTANTES.....	12
II. DESCRIPCION DEL METODO DE "PRONOSTICOS PONDERADOS".....	15
II.1 PLANTEAMIENTO.....	16
II.2 DEDUCCION DEL METODO.....	16
II.3 PROPIEDADES DE Y OBSERVACIONES SOBRE EL ESTIMADOR.....	18
II.4 ESTIMACION DE LOS PESOS OPTIMOS POR MAXIMA VEROSIMILITUD.....	19
II.5 ESTIMACION DE LA COVARIANZA PARA LOS PROCESOS GENERADORES.....	20
II.6 SUMARIO Y VERSION FINAL.....	21
II.7 GENERALIZACION DEL METODO Y FUTURAS EXTENSIONES.....	23

III. EJEMPLOS Y APLICACIONES DEL METODO.....	26
III.1 INTRODUCCION.....	27
III.2 SIMULADOR DE SERIES DE TIEMPO.....	27
III.2.1 GENERACION DEL RUIDO BLANCO.....	28
III.2.2 IMPLEMENTACION DE LA ESTRUCTURA DE RECURRENCIA.....	29
III.2.3 FUNCIONES PARA INTEGRAR SERIES.....	31
III.3 EJEMPLOS DE SERIES INCOMPLETAS Y USO DEL METODO EN ELLAS.....	32
III.3.1 SERIE 1.....	32
III.3.2 SERIE 2.....	34
III.3.3 SERIE 3 (INDICE DE LA BOLSA MEXICANA DE VALDRES	36
III.4 CONCLUSIONES.....	39
IV APENDICES, GRAFICAS Y BIBLIODRAFIA.....	43
IV.1 APENDICES	44
IV.2 GRAFICAS.....	56
IV.3 BIBLIODRAFIA.....	43

CAPITULO I

" INTRODUCCION AL PROBLEMA Y A SUS SOLUCIONES PREVIAS "

CAPITULO I

INTRODUCCION AL PROBLEMA Y A SUS SOLUCIONES PREVIAS.

I.1 INTRODUCCION.

En este primer capítulo se hablará del problema que representa la ausencia de algunas observaciones en una Serie de Tiempo Estadística. Se comenzará por dar algunos conceptos básicos del Análisis de Series de Tiempo. Al lector interesado en profundizar en esta área, se le sugiere referirse al texto de Box y Jenkins [1] y al texto de próxima publicación del Dr. Víctor M. Guerrero.

I.1.1 BREVE INTRODUCCION AL ESTUDIO DE LAS SERIES DE TIEMPO.

Una Serie de Tiempo es una colección de valores $\{Z_t\}$ medidos a lo largo del tiempo, generalmente a intervalos equidistantes. Cuando se supone que existe un elemento aleatorio, esto es, que no se puede hacer un análisis totalmente determinístico del fenómeno que ha generado la serie, se hablará de Series de Tiempo Estadísticas. A lo largo de esta tesis, al hablar de una serie, nos referiremos siempre al primer caso. El elemento aleatorio que se supone para el modelo es una sucesión de variables aleatorias independientes e idénticamente distribuidas usualmente llamada "Ruido Blanco" y denotada por $\{a(t)\}$. La distribución de cada una de ellas generalmente es una $N(0, \sigma^2)$.

Existen varios enfoques para el estudio de las Series de Tiempo. Uno de los que mayor atención ha recibido es el de Box

y Jenkins [1] desde la primera edición de su libro en 1970. Este es, fundamentalmente, el que se usará lo largo de este trabajo.

Segun estos autores, el Análisis de Series de Tiempo puede realizarse ajustando procesos estocásticos mixtos de tipo ARIMA a la serie en estudio a través de un proceso que se divide en 3 etapas: Identificación, Estimación y Verificación con 2 usos fundamentales : Control y Pronóstico.

La primera etapa consiste en identificar el proceso estocástico que ha generado la serie, la segunda en encontrar estimadores para el modelo identificado, la tercera en verificar si los supuestos de los modelos elegidos se cumplen para, finalmente, pronosticar los futuros valores de la serie con el modelo elegido o identificar la función de transferencia asociada al sistema dinámico en estudio.

La identificación se lleva a cabo a través de las funciones de autocorrelación (FAC) y de autocorrelación parcial (FACP) para las que existen patrones de comportamiento bien definidos. Sin embargo, quien quiera identificar un modelo adecuadamente debe tener detrás un buen bagaje de conocimientos y experiencia ya que los patrones presentan muchísimas variantes.

El metodo de Box-Jenkins es iterativo. La identificación que se lleva a cabo por primera vez difícilmente es la definitiva. Los residuales del modelo estimado (valores esperados menos valores observados) son a su vez analizados para determinar

si existe aún una estructura de dependencia entre ellos. De ser así, el modelo es incompleto y el número de parámetros necesita ser aumentado. Por otro lado, el número de parámetros involucrados puede ser excesivo. En este caso, se determina cuales son los parámetros que están de más para re-estimar el modelo sin ellos. En ambos casos, se regresa a la etapa de estimación y el proceso se repite hasta encontrar un modelo satisfactorio.

Los Procesos Estocásticos que forman parte de la familia de modelos ARIMA se presentan brevemente a continuación:

1. AUTOREGRESIVOS: Llamados así porque suponen un modelo de regresión de la variable en sí misma. Se denotan como AR(p), donde p es el número de observaciones anteriores significativas. Su forma funcional es:

$$Z_t = a - r_1 Z_{t-1} - r_2 Z_{t-2} - \dots - r_p Z_{t-p}, \quad a \text{ v.a.i.i.d.} \sim N(0, \sigma^2)$$

2. PROMEDIOS MOVILES: Son una combinación lineal del ruido blanco (a_t) q períodos atrás. El nombre de promedios móviles no es del todo correcto, ya que los pesos c_i en la forma funcional:

$$Z_t = a_t - c_1 a_{t-1} + c_2 a_{t-2} + \dots + c_q a_{t-q}, \quad a_t \text{ i.i.d.} \sim N(0, \sigma^2)$$

no suman uno. Se denotan por MA(q)

3. MODELOS MIXTOS (ARMA y ARIMA): Los modelos mixtos son una combinación de los dos anteriores. Suponen que el valor de la serie al tiempo t depende de p valores anteriores de la misma serie y del ruido blanco q momentos atrás. La diferencia entre ellos es que el modelo ARIMA supone que a la serie se ha aplicado el operador $(1-B)^d Z_t = Z_t - Z_{t-1}$ d veces. Los 2 tipos de modelos mixtos, Autoregresivos de Promedios Móviles y Autoregresivos Integrados de Promedios móviles, se denotan respectivamente por ARMA(p,q) y ARIMA(p,d,q) con forma funcional:

$$W_t = r_1 W_{t-1} - r_2 W_{t-2} - \dots - r_p W_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q},$$

a_t v.a.i.i.d. $\sim N(0, S^2)$

En el caso de los modelos ARIMA:

$$W_t = (1-B)^d Z_t$$

Asociado a cada modelo, existe un "Polinomio Característico" en función del operador de retraso ($B Z_t = Z_{t-1}$); cuyas propiedades son de gran utilidad para evaluar características importantes de la serie en estudio, tales como estacionaridad e invertibilidad. Esto se verá más adelante.

Si la media de la serie no depende del tiempo, esto es, la serie tiende a equilibrarse en algún momento dado y solo varía en adelante alrededor de un mismo valor, se dirá que es estacionaria de primer orden.

Si además la varianza es constante en el tiempo y las autocovarianzas (C_k) solo dependen del retraso con el que son me-

didas, se hablará de estacionaridad de orden 2. En general, si sus primeros k momentos son constantes, se hablará de estacionaridad de orden k.

Una de las formas típicas de no estacionaridad es el fenómeno conocido como "Estacionalidad". Si la serie presenta un patrón constante para intervalos de tiempo equidistantes, se hablará de una serie con un factor estacional. Este fenómeno puede modelarse suponiendo que el elemento aleatorio de la serie en lugar de ruido blanco es, a su vez, un modelo ARIMA. Un ejemplo de estacionalidad es el incremento de ventas los meses de Diciembre y Enero cada año.

Si el polinomio asociado con los modelos MA es invertible, el proceso también lo es. Esto se puede ver a través de sus raíces. Si todas caen fuera del círculo unitario, el polinomio es invertible. En el caso de los modelos AR, siempre estará presente esta característica. La no-invertibilidad implica la posible duplicidad de modelos.

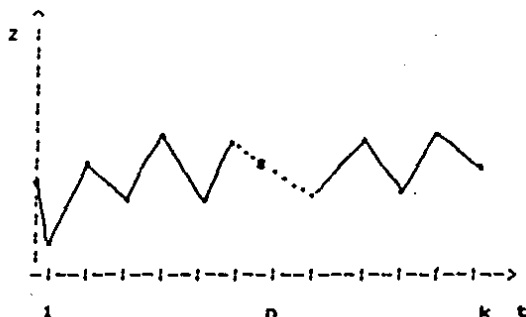
Todos los procesos MA son estacionarios; esto es, las raíces de su polinomio característico caen fuera del círculo unitario, y al aplicar el operador diferencia $-(1-B)$ al que ya nos hemos referido a los modelos AR lo que se pretende es obtener estacionaridad de primer orden. Para tener varianza constante, se aplica a la serie la transformación de Bartlett:

$$T(Z) = \begin{matrix} \text{---} & & i \\ | & & Z, & & i=0 \\ -| & & \ln(Z), & & i=0 \\ | & & & & \\ \text{---} & & & & \end{matrix}$$

Donde i se determina a través de los coeficientes de variación de la serie. Se escogerá aquella i para la que $CV(Z^i)$ sea mínimo. Cuando $i=0$, se aplica la transformación logaritmo. La serie, una vez presentando comportamiento estacionario de al menos orden 2, se modela conforme alguno de los procesos mencionados anteriormente.

I.1.2 PLANTEAMIENTO DEL PROBLEMA:

Supóngase que se han observado $Z_1, \dots, Z_{n-1}, Z_{n+1}, \dots, Z_k$ valores de una Serie de Tiempo. En el momento n , por alguna razón, no es posible obtener la observación Z_n correspondiente, como en la siguiente gráfica:



A este valor Z_n le llamaremos "observación faltante" de la serie $Z = \{Z_i : i \in T\}$ donde T es un conjunto indexado en el tiempo. Z_n es el valor a estimar.

Este no es de ninguna manera un problema poco común. Piénsese a modo de ejemplo en lo que ocurrió en Septiembre de 1982 cuando, debido a la nacionalización de la Banca, la Bolsa Mexicana de Valores dejó de cotizar durante 20 días consecutivos.

Así, surgen las siguientes interrogantes:

1. ¿Qué procedimiento y qué criterios deben usarse para estimar de la mejor manera Z_n ? En general, dados m valores de un total posible de k , ¿qué deberemos hacer para, a partir de las m observaciones conocidas, atacar el problema que representa la falta de información sobre las restantes $k-m$ observaciones?

2. ¿Qué características deseamos que tengan estos estimadores?

3. Si se decide ponderar 2 pronósticos, ¿qué pesos se les debe dar y de que manera se determinan?

4. ¿Cómo podemos estimar la covarianza entre estos estimadores y sus varianzas respectivas?

5. ¿Cómo comparar esta metodología con otras? ¿Es la Varianza de los estimadores un buen criterio?

Estas preguntas se contestarán en el siguiente capítulo.

1.2 ALGUNOS TRATAMIENTOS ANTERIORES.

En esta sección se muestran los resultados de la investigación bibliográfica sobre el tratamiento de observaciones faltantes en Series de Tiempo. Podemos dividir los artículos revisados en 2 grandes tipos :

1. Aquéllos cuyo objetivo es el identificar el proceso y estimar sus parámetros cuando la serie esta incompleta.

2. Aquéllos en donde el punto a tratar es la estimación de las observaciones faltante en sí.

El objetivo de esta tesis es el segundo, estimar estas observaciones y analizar las propiedades del estimador. Para lograrlo, sin embargo, se deberán identificar y estimar la series incompletas. Es por eso que se han incluido ambos tipos.

1.2.1 ESTIMACION DE SERIES INCOMPLETAS.

Se comenzará entonces por una revisión sumaria de la literatura del primer tipo:

Zellner [14] en 1966 trabaja la estimación de modelos AR(1) del tipo:

$$Y_t = a Y_{t-1} + b X_t + u_t \quad u_t \sim N(0, \sigma^2)$$

y, suponiendo datos trimestrales para X_t y anuales para Y_t , u_t

utiliza un enfoque de máxima verosimilitud para encontrar ecuaciones en \hat{a} , \hat{b} , y $\hat{\sigma}^2$. Este enfoque, sin embargo, está restringido a un solo tipo de modelo y por lo tanto carece de generalidad. Además, requiere de una serie correlacionada con la serie en estudio para poder ajustar un modelo adecuado.

P. M. Robinson [11] en 1977 trabaja con un modelo AR(p) :

$$y_t = \sum_{j=1}^p a_j y_{t-j} + w_t, \quad w_t \sim N(0, \sigma^2 I_T)$$

con 2 casos:

$$f_t = \sum_{i=1}^q b_i z_{it} \quad \text{y} \quad f_t = \sum_{i=1}^q b_i (z_{it} - \sum_{j=1}^p a_j z_{i,t-j})$$

El artículo trata con observaciones censadas y considera varias estrategias para estimar los parámetros del modelo. Dos enfoques son utilizados para encontrar estimadores M.V. : El método de Newton-Raphson y el algoritmo E-M. El primero converge rápidamente y si es iniciado con un estimador consistente, produce otro asintóticamente eficiente en un solo paso. El algoritmo E-M, a pesar de su convergencia a estimadores M.V., no es apropiado para el caso de las series de tiempo ya que las esperanzas condicionales del paso E suelen complicarse y corresponden a momentos de segundo orden. Además, solamente se puede llevar a cabo si el proceso es AR(p), lo cual le resta generalidad.

Ferreiro [5] , en su examen preliminar de Doctorado en 1980, al efectuar una investigación sobre la contaminación del Río Milwaukee se enfrentó al problema de una Serie de Tiempo incompleta.

Resolvió el problema desarrollando una modificación al paso E del algoritmo E-M, de modo que se calculan las esperanzas condicionales de las observaciones faltantes en vez de hacerlo con las esperanzas condicionales de las estadísticas suficientes. Da a este método el nombre de "PEM" (Pseudo E-M).

Jones [7] en 1980 utiliza una formulación de estado espacial para el modelo ARIMA. Con esta representación markoviana, sugiere calcular la función de verosimilitud recursivamente usando un filtro de Kalman. Se utilizan programas de optimización no lineal para maximizar la función de verosimilitud.

En 1986 Kohn y Ansley [10] utilizan un enfoque similar, generalizando resultados para cualquier patrón de observaciones faltantes. Suponen una representación de estado espacial para los modelos a trabajar con condiciones iniciales difusas. Así, usan una modificación del Filtro de Kalman para obtener la verosimilitud y una modificación del algoritmo de suavizamiento de punto fijo para interpolar las observaciones faltantes. Este artículo es el último escrito a la fecha (Septiembre de 1986) y también el más completo y avanzado. Presenta tanto un método de estimación de la serie incompleta como un algoritmo para completarla, que son casos generales de los métodos presentados por

Box y Jenkins [1] y Jones [7] respectivamente.

1.2.2 ESTIMACION DE OBSERVACIONES FALTANTES.

En cuanto a estimación de valores faltantes, Chow y Lin [2] trabajan en 1976 con un modelo de regresión:

$$Y = Xb + u \quad u: \text{vector aleatorio de media cero y matriz de covarianza } V.$$

En el artículo se trata el caso de valores trimestrales de Y y valores mensuales en X (una matriz con p variables). El objetivo es llevar a cabo una desagregación de Y. La estimación se obtiene de minimizar la traza de la matriz de covarianzas del error de predicción bajo las condiciones de sesgo cero.

Wilson y Brubacher [13] en 1977 hacen una estimación de las observaciones faltantes por medio de mínimos cuadrados suponiendo un modelo ARMA.

Escribiendo el modelo como:

$$P(B)Z_t = Q(B)a_t \quad E(Z_t) = 0, \quad t=1,2,\dots$$

Se supone invertibilidad

$$a_t = (P(B)/Q(B)) Z_t = R(B) Z_t = Z_t - r_1 Z_{t-1} - r_2 Z_{t-2} - \dots$$

y la suma de los cuadrados de los errores

$$SS = \sum_{t=L}^L a_t^2 = \sum_{t=L}^L \sum_{i=0}^{M_0} (r_i)^2 \quad L \rightarrow \infty$$

es minimizada , resultando ecuaciones para los estimadores.

De Alba y Mendoza [4] en 1985 tratan el problema usando un enfoque bayesiano para la desagregación de series, llegando a un resultado equivalente al de Chow y Lin [2].

Damsleth [3] en 1980 presenta un tratamiento similar al que se dará en esta tesis, pero con algunos supuestos diferentes. Plantea el problema de modo que el error medio cuadrático sea minimizado. Las diferencias con el método que se presenta en esta tesis son:

1. Damsleth supone conocer perfectamente todos los parámetros del proceso generador. En el caso que nos concierne , éstos deberán ser estimados.

2. Damsleth supone procesos de ruido blanco distintos en cada intervalo de la serie entre observaciones faltantes. Aquí trabajaremos bajo el supuesto de que el proceso estocástico de ruido blanco es el mismo para cada trozo de la serie. Esto facilitará algunos cálculos - como el de la covarianza de

los pronósticos - y además parece muy natural, ya que se trata de la misma serie generada por el mismo proceso. En términos matemáticos, este supuesto implica que la covarianza de las variables aleatorias sea cero, independientemente del hecho de que correspondan a intervalos distintos de la serie.

3. Damsleth plantea una solución en el ambiente de series de tiempo y minimiza el error de pronóstico. En esta tesis se plantea un problema de optimización de varianzas cuya solución es extrapolada al caso específico de las series de tiempo.

4. Damsleth considera el caso de series no estacionarias. Este problema no se presenta al usar "Pronósticos Ponderados" ya que en cada intervalo de la serie se aplican, de ser necesario, las transformaciones adecuadas antes de la modelación.

Finalmente, existen soluciones "heurísticas". Entre ellas está el interpolar por medio de una línea recta las observaciones posterior y anterior a la faltante o, sencillamente, ignorarla. A pesar de que presentan la ventaja de ser muy sencillas, dejan mucho que desear. En el caso de la interpolación, el resultado obtenido es muy pobre ya que no toma en cuenta ninguna característica del proceso. El ignorarla es peor aún, ya que se altera toda la estructura de correlación. De ahí que sea necesario un método efectivo y bien fundamentado, pero que a la vez permita una implementación sencilla. Ese será el objetivo de esta tesis.

CAPITULO II.

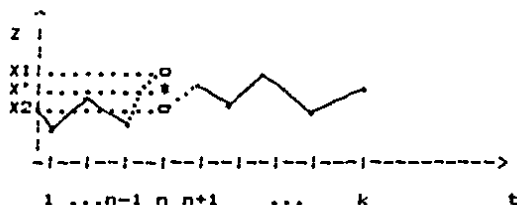
" DESCRIPCION DEL METODO DE PRONOSTICOS PONDERADOS ".

CAPITULO II:

DESCRIPCION DEL METODO DE "PRONOSTICOS PONDERADOS".

II.1 PLANTEAMIENTO

Como se recordará, el problema consiste en encontrar estimadores de una observación faltante Z_n en una Serie estacionaria e invertible. Veamos de nuevo la grafica hipotética de dicha Serie:



II.2 DEDUCCION DEL METODO.

i) Dividamos la serie original en 2 series w e y , tales que:

$$y = (Z_i \mid i=1, \dots, n-1) \quad \text{y} \quad w = (Z_j \mid j=n+1, \dots, k)$$

ii) Modelamos ambas y realizamos los siguientes pronósticos:

X_1 = Estimador de Z_n pronosticado a futuro por y .

X_2 = Estimador de Z_n pronosticado a pasado por w .

Es claro que así obtenemos 2 estimadores, no necesariamente iguales, para la misma observación Z_n . Así, procedemos a encontrar un solo estimador X' a partir de X_1 y X_2 .

iii) Supondremos que X' es de la forma $K \cdot X_1 + K' \cdot X_2$ (i.e. Una ponderación de X_1 y X_2).

iv) Deseamos que X' tenga las siguientes características :

a) $k' + k'' = 1$.

b) $MSE(X')$ debe ser mínimo.

Así, resolvemos el siguiente problema de minimización en términos de K' y K'' . Vale la pena hacer notar que, dado que X_1 y X_2 son estimadores obtenidos a través del Ajuste a modelos ARIMA, cada uno de ellos es insesgado respecto al pasado (X_1) o al futuro (X_2), pero no lo son si se toman en cuenta pasado y futuro de manera simultánea. Hacerlo así relaja muchos de los supuestos teóricos. Además, podemos utilizar la varianza del estimador en lugar de su error medio cuadrático.

$$\text{Min } (MSE(X')) = \text{Min} (V (K'X_1 + K''X_2))$$

$$\text{Sujeto a: } K' + K'' = 1 \quad \text{y} \quad K', K'' \geq 0$$

Realizando transformaciones en la función objetivo, el problema se replantea como:

$$\text{Min} (K'^2 V(X_1) + K''^2 V(X_2) + 2K'K'' \text{COV}(X_1, X_2))$$

$$\text{Sujeto a: } K' + K'' = 1, \quad K', K'' \geq 0$$

Para efectos de notación, sean:

$$V(X_1) = V_1, \quad V(X_2) = V_2 \quad \text{y} \quad \text{COV}(X_1, X_2) = C.$$

$$\text{Sea } L (K', K'', \lambda) = K'^2 V_1 + K''^2 V_2 + 2K'K''C - \lambda (K' + K'' - 1)$$

$$\Rightarrow \frac{\partial(L)}{\partial(K')} = 2K'V_1 + 2K''C - \lambda = 0 \quad (1)$$

$$\hat{\quad} \frac{\partial(L)}{\partial(K'')} = 2K''V_2 + 2K'C - \lambda = 0 \quad (2)$$

$$\hat{\quad} \frac{\partial(L)}{\partial(\lambda)} = K' + K'' - 1 = 0 \quad (3)$$

Despejando A de (1) y (2) e igualando se obtiene:

$$2K'V_1 + 2K''C = 2K''V_2 + 2K'C$$

$$\Rightarrow K'V_1 - K'C = K''V_2 - K''C$$

$$\Rightarrow K'(V_1 - C) = K''(V_2 - C)$$

Aplicando (3) y sustituyendo K'' por 1-K' se obtiene:

$$K'(V_1 - C) = (1 - K')(V_2 - C)$$

$$\Rightarrow K'(V_1 - C + V_2 - C) = V_2 - C$$

$$\Rightarrow K' = \frac{V_2 - C}{V_1 + V_2 - 2C} \quad \text{Sustituyendo el denominador por } V(X_1 - X_2):$$

$$K' = \frac{V(X_2) - \text{COV}(X_1, X_2)}{V(X_1 - X_2)} \quad \text{y, análogamente, } K'' = \frac{V(X_1) - \text{COV}(X_1, X_2)}{V(X_1 - X_2)}$$

son los pesos óptimos para X'.

II.3 PROPIEDADES Y OBSERVACIONES SOBRE EL ESTIMADOR.

1. VARIANZA: Como se recordará, el planteamiento del problema en términos matemáticos fue tal que la varianza resultara mínima. El resultado así obtenido proporciona también el valor de $V(X')$, solo debemos sustituir K' y K'' en la función objetivo.

2. SESGO: Supóngase que X_1 y X_2 son estimadores insesgados de Z_n . Entonces X' también lo será. Este resultado es muy sencillo de demostrar:

$$\text{P.D.: } E(K'X_1 + K''X_2) = Z_n$$

$$\text{Dem: } E(K'X_1 + K''X_2) = E(K'X_1) + E(K''X_2)$$

$$= K'E(X_1) + K''E(X_2) \dots \dots \dots (1)$$

Como X_1 y X_2 son insesgados, de (1) se sigue:

$$K' Zn + K'' Zn = Zn (K' + K'') = Zn (1) = Zn \quad ///$$

II.4 ESTIMACION DE LOS PESOS OPTIMOS POR MAXIMA VEROSIMILITUD.

Surge un nuevo problema: Una vez establecidos cuales deben ser los pesos óptimos, necesitamos obtenerlos. Observemos que tanto K' como K'' están en función de las variables aleatorias $V(X1)$, $V(X2)$ y $COV(X1, X2)$ y son, a su vez, variables aleatorias. Parece razonable pensar que si obtenemos los estimadores de MV (Máxima Verosimilitud) de $V(X1)$, $V(X2)$ y $COV(X1, X2)$, habremos resuelto el problema gracias a la propiedad de invarianza de estos estimadores. Conviene citar el siguiente Teorema [9]:

TEOREMA II.1 (PROPIEDAD DE INVARIANZA DE ESTIMADORES MV)

Sean:

- X un vector de parámetros $X=(X1, \dots, Xn)$

- X' un vector de estimadores MV $X'=(X1', \dots, Xn')$ tal que:

Xi' es MV de Xi para $i=1, \dots, n$. para la densidad:

$$f(x_1, \dots, x_n)$$

- $T(X) = (T(X)_1, \dots, T(X)_r)$, $1 \leq r \leq n$ una transformación

de espacio de parámetros.

- $M(T(X); x_1, \dots, x_n)$ la función de verosimilitud inducida por la transformación T en la muestra x_1, \dots, x_n .

ENTONCES:

Un estimador de Máxima Verosimilitud para $T(X)$ es $T(X')$

Dem.:

Sea X'' un estimador M.V. de X . Basta demostrar que:

$$M(T(X''); x_1, \dots, x_n) \geq M(T; x_1, \dots, x_n) \text{ para toda } T \text{ que se}$$

sigue de la desigualdad :

$$M(T, x_1, \dots, x_n) = \sup_{(X_i: T(X)=T)} L(X; x_1, \dots, x_n)$$

$$\leq \sup_{V X} L(X; x_1, \dots, x_n) = L(X^*; x_1, \dots, x_n) = \sup_{(X_i: T(X)=T(X^*))} L(X; x_1, \dots, x_n)$$

$$= M(T(X^*); x_1, \dots, x_n). \quad ///$$

Si somos capaces de encontrar estimadores de Máxima Verosimilitud para $V(X_1), V(X_2)$ y $COV(X_1, X_2)$, usando este Teorema habremos obtenido estimadores MV para k' y k'' y el problema estará resuelto. Nos interesa en particular cual debe ser la forma funcional de $COV(X_1, X_2)$ en función de las autocovarianzas del proceso, una vez identificado éste - nótese que se habla de un solo proceso generador a pesar de tener 2 series (w e y) -.

II.5 ESTIMACION DE LA COVARIANZA PARA LOS PROCESOS GENERADORES.

El problema consiste en encontrar algún estimador de la covarianza $COV(X_1, X_2)$. Recuérdese que son 2 series distintas - cada una de las mitades de la serie original -, pero generadas por exactamente el mismo proceso. Damsleth [3] propone un estimador general para la covarianza del pronóstico L períodos adelante y $L-M$ períodos hacia atrás. En este caso, estamos interesados únicamente en la covarianza 1 período adelante y un período atrás. El estimador que usaremos para este caso en particular es :

$$COV(X_1, X_2) = \frac{UU}{1112} B^2$$

Donde U_1 es el peso usado para obtener el pronóstico 1 período hacia adelante y U_2 es el peso usado para obtener el pronóstico 1 período hacia atrás.

11.6 SUMARIO Y VERSION FINAL.

El método de "Pronósticos Ponderados" para estimar observaciones faltantes en Series de Tiempo se dividirá en cuatro etapas:

1. Análisis de cada trozo de la serie usando la técnica Box-Jenkins. Esta etapa deberá cubrir tanto la identificación y estimación del proceso como la generación de pronósticos en cada caso.
2. Obtención de estimadores de Máxima Verosimilitud para $V(X_1)$, $V(X_2)$, $COV(X_1, X_2)$.
3. Ponderación de los pronósticos obtenidos en (1) de acuerdo a la siguiente expresión:

$$\hat{X} = \frac{(V(X_2) - COV(X_1, X_2)) X_1 + (V(X_1) - COV(X_1, X_2)) X_2}{V(X_1) + V(X_2) - 2 COV(X_1, X_2)}$$

4. Generación de un intervalo de confianza para el estimador. Esto se puede hacer de manera sencilla, ya que se conoce un estimador de su varianza - notese que los pesos óptimos son los valores minimizantes de la varianza - y, dado que se trata de un pronóstico ponderado, podemos usar la

metodología de construcción de intervalos de pronósticos.

Usando el supuesto de normalidad del ruido blanco, y siendo \hat{Z} y $\hat{V}(Z)$ los estimadores de la observación faltante y de su varianza respectivamente, un intervalo de confianza para Z estaría dado por :

$$(\hat{Z} - U(p) s(Z), \hat{Z} + U(p) s(Z))$$

donde $s(Z)$ es la raíz cuadrada de $\hat{V}(Z)$ y $U(p)$ es el p -ésimo percentil de una distribución normal estandarizada.

II.7 GENERALIZACION DEL METODO Y FUTURAS EXTENSIONES.

II.7.1 PLANTEAMIENTO.

En la sección anterior, se desarrolló el método de "Pronósticos Ponderados" para una observación faltante aislada. En esta sección, por otro lado, se desarrollará esta misma idea suponiendo que existen k observaciones faltantes consecutivas. La idea es esencialmente la misma y los pesos óptimos no cambian al alterar el supuesto de aislamiento. Solamente tendremos que investigar si los estimadores de la covarianza son los mismos.

Sean $X_1(k)$ y $X_2(k)$ los pronósticos a futuro y a pasado para la k -ésima observación faltante consecutiva X_{t+k} .

El error de pronóstico k períodos adelante (o atrás, si se invierte la serie) para un pronóstico $Z(k)$ de Z_{t+k} y su varianza σ^2 , según Box y Jenkins[1]:

Error de Pronósticos:

$$U_{t+k} = X_{t+k} - X_1(k) = \sum_{i=0}^{k-1} \psi_i a_{t+k-i} \quad (1)$$

con varianza:

$$\text{var } U_{t+k} = \sigma^2 \sum_{i=0}^{k-1} \psi_i^2 \quad (2)$$

Supongamos ahora que (Z_t) se observa hasta el tiempo n y de $n+1$ a $n+k-1$ faltan observaciones. El valor de Z_{n+i} , $n+1 < i < n+k-1$,

puede ser pronosticado a futuro de las observaciones anteriores y a pasado de las posteriores,obteniendo:

$Z_n(i)$: Estimador a futuro (Pronóstico) de Z_{n+i}

$Z_{n+k+1}(k+1-i)$: Estimador invirtiendo la serie de Z_{n+i}

Haciendo el mismo planteamiento de la sección anterior, y tratando de minimizar el error de pronóstico,obtenemos que los pesos óptimos son :

$$k' = (V(v_{n+1}) - COV(u_{n+1}, v_{n+1})) / V(v_{n+1} - u_{n+1})$$

$$k'' = (V(u_{n+1}) - COV(u_{n+1}, v_{n+1})) / V(v_{n+1} - u_{n+1})$$

Las varianzas de los errores vienen dadas por (2) y la covarianza por:

$$COV(u_{n+1}, v_{n+1}) = COV\left(\sum_{j=0}^{i-1} \psi_j a_{n+1-j}, \sum_{j=0}^{k-1} \psi'_j e_{n+1+j}\right)$$

Así:

$$COV(u_{n+1}, v_{n+1}) = \sum_{j=0}^{i-1} \sum_{l=0}^{k-1} \psi_j \psi'_l Cae(i+j)$$

donde Cae es la función de covarianza cruzada de los procesos de ruido blanco de las series (a y e). Damsleth (3) sugiere como obtenerlas a partir de un sistema de ecuaciones diferenciales. Sin embargo, si somos consistentes con el supuesto de que el proceso de ruido blanco es el mismo para los 2 proce -

2
sos, C_{ae} se convierte en S_a^2 , la varianza del ruido blanco, como en el caso de una sola observación faltante.

Entre las futuras extensiones del método, se han considerado las siguientes:

1. Implementación del método para el Módulo de Manejo de Series Incompletas en un Sistema Experto actualmente en desarrollo en el Centro Científico de IBM en Mexico.
2. Comparación teórica con otros métodos .
3. Obtención de Intervalos de Confianza para el estimador, haciendo Teoría de Distribución.

Estos puntos pueden ser usados como posibles temas de tesis para estudiantes de Actuaría, Estadística o Economía.

CAPITULO III.

" EJEMPLOS Y APLICACIONES DEL METODO ".

CAPITULO III.

"EJEMPLOS Y APLICACIONES DEL METODO ".

III.1 INTRODUCCION.

En este capítulo se mostrará la manera en que el método de "Pronósticos Ponderados" puede completarse en la práctica Series de Tiempo. Con el objetivo de estudiar el modo en que el estimador ajusta al valor faltante, se utilizarán 2 series que posteriormente se supondrá que están incompletas. La simulación de una de ellas se hará en Q'NIAL, lenguaje desarrollado conjuntamente por la Universidad de Queen's en Kingston, Canada y el Centro Científico de IBM en Cambridge, Ma. Se mostrarán tanto la codificación como las ideas usadas en su implementación y las pruebas de Bondad de Ajuste realizadas. Otra de las series se tomará del libro de Box y Jenkins [1].

Una vez mostrados los resultados obtenidos de la aplicación del método a los ejemplos, se tomará una serie económica incompleta real - el índice diario de cotización de la Bolsa Mexicana de Valores - y se estimarán sus valores faltantes.

III.2 SIMULADOR DE SERIES DE TIEMPO.

En esta sección, se muestran las funciones utilizadas en la generación de una Serie de Tiempo. Los modelos a simular son $ARIMA(p,d,q)$. Dividiremos las funciones en 3 clases: Las que

intervienen en la generación del ruido blanco, las que implementan la mecánica de cada modelo y las que integran series.

III.2.1 GENERACION DEL RUIDO BLANCO.

A) SIMULACION DE UNA DISTRIBUCION UNIFORME.

En D'NIAL, la función primitiva "random n" genera números pseudoaleatorios enteros entre 0 y n, otorgando a cada uno de ellos la misma probabilidad $1/n$. Para generar Normales, es necesario contar con un buen generador de números uniformes. Esa fue la primera meta: generar números distribuidos de acuerdo a una $U(0,1)$. Se logró con la función:

```
unif IS OPERATION n ( (Random (n Reshape 10000))/10000)
```

Toma como argumento un número entero - n - y devuelve una muestra simulada de tamaño n de una distribución $U(0,1)$, con números hasta la cuarta cifra decimal.

B) SIMULACION DE UNA DISTRIBUCION NORMAL $N(0,1)$

Para simular una Distribución Normal a partir de una Uniforme, existen varios métodos. Uno de ellos, el método polar de Box-Muller se implementó y su codificación se presenta a continuación:

```
B_Muller IS OP n (((-2*Ln(unif n)) Power .5)*Sin (2*Pi*unif n))
```

Se le da como argumento un número entero, regresando una

muestra simulada de tamaño n de una Distribución Normal $N(0,1)$.

C) SIMULACION DE UNA DISTRIBUCION NORMAL $N(0,S^2)$.

Una vez simulada una Distribución Normal Estandarizada, es muy fácil modificar su Varianza. Gracias a sus propiedades, lo único que se debe hacer es multiplicar toda la muestra por la raíz cuadrada de la varianza que se desea. Así, la siguiente función nos proporciona una muestra de una Distribución Normal con media cero y cualquier varianza con solo proporcionarle dicha varianza y el tamaño de muestra a simular que se desea.

```
rbianco IS OPERATION n vara ( (Sqrt vara)*B_Muller n )
```

Las muestras así generadas pasaron las pruebas de Lilliefors y ji cuadrada para bondad de ajuste de Normales. Dichas pruebas se presentan en el Apendice IV.7.

III.2.2 IMPLEMENTACION DE LA ESTRUCTURA DE RECURRENCIA.

A) PROCESOS AUTOREGRESIVOS.

Recordemos que la estructura general de un proceso autoregresivo de orden p es la siguiente:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a(t), \quad a(t) \sim N(0, S^2)$$

Así, necesitamos una función que, dados el tamaño de la Se-

rie, el vector de coeficientes autoregresivos y la varianza del ruido blanco,proporcione una lista de valores que representen la Serie de Tiempo generada por un proceso autoregresivo. Esta función es:

```

ar IS OPERATION varar ph nar (
p Gets Tally ph;
at Gets rblanco nar varar;
z Gets p Reshape 0;
For I With Count nar Do
  z Gets z Link Sum Sum ((p Takeright z)#voltea ph (I Take
  at));
  at Gets I Drop at;
endfor;
z Gets p Drop z)

```

B) PROCESOS DE PROMEDIOS MOVILES.

La relación de recurrencia de los procesos de promedios móviles es la siguiente :

$$z_t = \theta_1 a(t-1) + \theta_2 a(t-2) + \dots + \theta_p a(t-p) + a(t), \quad a(t) \sim N(0, S^2)$$

La función que, dados el tamaño de la serie, la varianza de el ruido blanco y el vector de parámetros de promedios móviles , regresa una colección de valores que representa la Serie de Tiempo generada por un proceso de promedios móviles es:

```

ma IS OPERATION varma th nma (
q Gets Tally th;
at Gets rblanco nma varma;
(Outer + Tell (nma-q) (Sortdown(Tell (q+1)) Choose at))
Inner[+,#] (I Link th))

```

D) MODELOS MIXTOS ARMA .

La función que genera Modelos Autoregresivos y de Promedios Móviles, cuya forma funcional es:

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

donde $a(t) \sim N(0, S)$, es la siguiente:

```

arma IS OPERATION vararma th ph narma (
p gets Tally ph;
at Gets ma vararma th narma;
z Gets p Reshape 0;
For I With Count narma do
  z Gets z Link Sum Sum ((p Takeright z) #voitea ph) (1
  Take at));
at Gets 1 Drop at;
Endfor;
z Gets p Drop z)
  
```

Nótese su gran parecido con la función generadora de procesos autoregresivos puros. Esto es porque la mecánica de ambas es la misma, difiriendo en que en este caso suponemos que el ruido blanco es, a su vez, un proceso de promedios móviles.

Hasta este punto, todas las funciones generan procesos estacionarios. En la siguiente sección presentamos funciones que sirven para agregar series.

III.2.3 FUNCIONES PARA INTEGRAR SERIES.

La idea es implementar el operador $S = (1-B)^{-1} = \sum_{i=1}^k Z_i$, $k=1, \dots, n$

Se hicieron 2 funciones, con estilos distintos de programa-

ción, que a continuación se muestran:

```
      integra is op serie (
EACH Sum Rows(((Tally serie) (Tally serie) Reshape serie) * (
Duter >= (Count Tally serie) (Count Tally serie))))
```

```
      integral is op serie (
z Gets 1 Take serie;
  For I With 2 + Tell (-1 + Tally serie) Do
    z Gets z Link Sum I Take serie;
  Endfor;
z)
```

La primera usa ideas del tipo APL, esto es, se trabaja con el arreglo - la serie - como un todo. En la segunda, por otro lado, se trabaja con los elementos uno a uno. Debido a que O'NIAL es un lenguaje nuevo cuya implementación en computadoras no es aún lo eficaz que puede desearse, al usar la primera función para series grandes - del orden de 100 observaciones - el espacio de trabajo de la computadora se llena rápidamente. La segunda función no presenta este problema.

III.3 EJEMPLOS DE SERIES INCOMPLETAS Y USO DEL METODO EN ELLAS.

III.3.1 SERIE1

La primera simulación se hizo con un proceso AR(1), usando los siguientes parámetros:

1. Parámetro Autoregresivo : 0.8
2. Valor de la Varianza de $a(t)$: 4.0
3. Número de Observaciones : 99.0
4. Parámetro de Media : 15.0

Resultando el modelo:

$$Z_t = 15 + 0.8 Z_{t-1} + a(t), \quad a(t) \sim N(0,4)$$

La serie obtenida se muestra a continuación:

1-8	13.87	12.48	12.80	14.85	13.81	13.90	10.70	11.14
9-16	11.32	9.86	12.78	9.22	10.71	11.38	8.67	10.63
17-24	15.76	14.64	17.27	18.94	21.86	18.08	14.85	15.42
25-32	13.62	15.59	18.41	16.43	16.00	14.55	13.43	11.89
33-40	13.37	13.03	14.72	13.50	14.87	15.44	15.13	14.04
41-48	12.45	13.56	13.03	7.22	(6.90)	4.86	7.92	12.19
49-56	12.67	11.02	12.50	13.26	16.54	16.64	16.80	21.33
57-64	20.19	20.91	21.44	20.37	19.57	19.58	20.10	20.36
65-72	20.51	18.68	19.29	17.11	14.38	14.27	16.17	17.96
73-80	16.87	17.96	15.24	15.20	18.09	18.40	17.64	14.04
81-88	14.42	13.47	12.98	15.00	13.03	10.89	13.42	14.08
89-96	17.02	18.88	17.16	17.44	17.18	13.28	11.24	14.21
97-99	14.71	12.73	9.5					

* Valor a estimar por el método.

La observación número 45 con un valor de 6.90 se escogió

para ser estimada por 2 razones: Primero, está localizada a la mitad de la serie y segundo, es un valor "patológico" ya que es uno de los mas pequeños. Esto debe ayudarnos a ver que, aún en situaciones extremas, el metodo ajusta razonablemente bien.

Se estimaron modelos para la primera mitad de la serie , la segunda mitad invertida y la serie completa. Los resultados se muestran en los apéndices IV.1, IV.2 y IV.3 respectivamente.

Los resultados para fines del método son los siguientes:

$$X_1 = 8.804485 \quad V(X_1) = 4.33171$$

$$X_2 = 5.978841 \quad V(X_2) = 3.82541$$

$$\text{COV}(X_1, X_2) = 2.5616$$

$$K' = 0.415518 \quad K'' = 0.584482$$

$$\hat{X} = 7.15 \quad \hat{V}(X) = .8086$$

INTERVALO DE CONFIANZA AL 95% :

(5.39036 , 8.91541101)

Notese como es mejor que cualquiera de los 2 pronósticos y que su promedio, aún en condiciones extremas como en este caso.

III.3.2 SERIE 2

La siguiente serie a completar no fué simulada, se tomó de libro de Box y Jenkins [1] donde aparece como la se -

rie F - " 70 Yields of Batch Data " -. La serie se presenta a continuación :

1 - 16	47	64	23	71	38	64	55	41	59	48	71	35	57	40	58	44
17 - 32	80	55	37	74	51	57	50	60	45	57	50	45	25	59	50	71
33 - 48	56	74	50	58	45	54	36(54)	48	55	45	57	50	62	44	64	
49 - 64	43	52	38	59	55	41	53	49	34	35	54	45	68	38	50	60
65 -70	39	59	40	57	54	23										

* - Valor a estimar.

Las estimaciones para la primera mitad de la serie, la segunda mitad invertida y la serie completa aparecen en los apéndices IV.4, IV.5 y IV.6 respectivamente. Para efectos del método, tenemos :

X1 = 58.246048 X2 = 50.6822662
V1 = 130.95915 V2 = 75.012101
COV (X1 , X2) = 13.9185565
K' : 0.342996567 K" : 0.657003433
ESTIMACION POR EL METODO : 53.2766174
VARIANZA DEL ESTIMADOR : 41.5168958
INTERVALO DE CONFIANZA AL 95%:
(40.6476306 , 65.9056042)

En este caso es interesante destacar la gran diferencia de magnitud entre las 2 varianzas. Una es casi el doble de la otra. Esta gran diferencia se hace patente en los pesos obtenidos para la ponderación, ya que uno de ellos es también casi el doble del otro. De nuevo hemos obtenido un estimador mejor que cualquiera de los 2 pronósticos e inclusive, que su promedio aritmético.

III.3.3 SERIE 3.

La última serie que se usará como ejemplo es la correspondiente al índice diario de cotización de la Bolsa Mexicana de Valores. El lector recordará que durante el último informe de gobierno del presidente López Portillo, tuvo lugar la nacionalización de la Banca. Este hecho repercutió, entre otras cosas, en la paralización temporal de actividades de la Bolsa Mexicana de Valores durante los primeros veinte días de septiembre de 1982. Como la mayoría de las series económicas de nuestro país, el análisis del índice diario de la Bolsa Mexicana de Valores implica el uso de técnicas avanzadas de modelación estadística, como el análisis de intervención. Las series escogidas para realizar los pronósticos a pasado y a futuro son las siguientes:

- Primer Trozo (Serie BMV1): 17 Mayo-31 Agosto de 1982
- Segundo Trozo (Serie BMV2): 20 Septiembre de 1982-21 de Enero de 1983.

Estos períodos se escogieron porque proporcionaban sufi-

ciente información y porque el hecho de haber tomado más ob - servaciones hubiera hecho necesario el uso de más interven - ciones.

Los listados de ambas series, sus gráficas y el resultado de sus modelos aparecen en el apéndice IV.5. Una vez removido el efecto de las intervenciones, ambas series se modelaron de acuerdo al siguiente modelo:

$$(1-B) \log(BMV)_t = (1+\theta B)a_t + \text{intervenciones.}$$

Los pronósticos para fines del método se presentan en la siguiente tabla:

Período	Pronóstico	Pronóstico	PESOS		Pronóstico Ponderado.
	BMV1	BMV2	K'	K''	
1	547.43	49.27	0.994	0.006	544.42
2	547.82	58.64	0.955	0.045	525.67
3	548.22	69.80	0.911	0.089	505.69
4	548.63	83.07	0.871	0.129	488.71
5	549.03	98.87	0.820	0.180	468.00
6	549.43	117.67	0.764	0.236	447.48
7	549.84	140.04	0.702	0.298	427.84
8	550.24	166.67	0.634	0.366	410.01
9	550.65	198.36	0.560	0.440	395.36
10	551.05	236.07	0.475	0.525	385.82
11	551.46	280.96	0.379	0.621	383.61
12	551.86	334.39	0.326	0.674	405.35
13	552.27	397.97	0.235	0.765	434.34
14	552.67	473.64	0.117	0.883	482.89

Es interesante destacar varios aspectos de los resultados de este análisis:

- El trozo faltante de la serie que se estima es una intervención, ya que fue causada por un evento exógeno. Esto hace particularmente interesante el análisis de este problema.

- La serie BMV2, que es la que se modela una vez reaunadas las actividades de la Bolsa presenta una caída muy fuerte en sus pronósticos (de 49.27 a 473.64). Esto se debe a la intervención número 7 y puede pensarse también a que la serie de alguna manera "conoce" del fenómeno y no es ilógico pensar que de haber abierto la Bolsa en esos días, hubiera sufrido una caída dramática (aunque quizá no a esos niveles). La serie BMV1, por otro lado, compensa esta caída tan brusca y lleva los pronósticos a los niveles anteriores.

- La curva de los pronósticos presenta una forma cóncava que es lo que se pudiera esperarse (una caída del índice de la Bolsa causada por pánico seguida de un repunte).

- Conviene aclarar que en este caso no se están estimando observaciones faltantes en un sentido estricto. Lo que se hace es "especular" sobre cuál hubiera sido el comportamiento del índice de haber existido actividad bursátil en esos días lo cual, para fines prácticos, es equivalente.

- Es interesante destacar como los pesos de pronóstico van cambiando de magnitud a medida que los pronósticos se alejan de un origen de pronóstico y se aproximan al otro.

III.4 CONCLUSIONES.

El método de "Pronósticos Ponderados" para completar Series de Tiempo funciona aceptablemente, aun en condiciones extremas. Si el valor faltante es un "outlier" - como en el caso de la serie 1 - , el método estima razonablemente bien . Si las varianzas de los pronósticos difieren considerablemente en magnitud - como en la Serie 2 -, el método pondera de acuerdo con esas diferencias. Inclusive, si la parte a estimar es una intervención cuyos efectos sean visibles en la información existente, "Pronósticos Ponderados" muestra sensibilidad respecto a dicha intervención. Además presenta las siguientes características :

1. El estimador tiene propiedades interesantes. Es insesgado, como se demuestra en el capítulo II y el planteamiento del problema es tal que su varianza se minimiza. En los 2 ejemplos trabajados, la varianza del pronóstico ponderado fue menor que la de cualquier pronóstico y el estimador fue mejor que ambos pronósticos e, incluso, que su promedio aritmético.

2. En casi todos los métodos revisados en el capítulo I - a excepción de los sugeridos por Damsleth [3] y Kohn y Craig [8] -, los valores faltantes son estimados a través de series correlacionadas con la que se quiere completar. Para encontrar una serie con estas condiciones, hace falta mucha experiencia por parte del investigador. El método de "Pronósticos Ponderados" lo hace a través de la serie misma, lo cual facilita su uso y lo

hace ideal para su implementación, por ejemplo, en un sistema experto donde se suponga que el usuario posee poca o ninguna experiencia en el Análisis de Series de Tiempo.

3. El estimador es una combinación lineal óptima en términos de la varianza de los pronósticos.

4. A pesar de su gran eficacia, el planteamiento en términos matemáticos del método es relativamente sencillo.

5. El Método es independiente de la forma funcional del modelo elegido, ya que depende únicamente de los pronósticos, sus varianzas y de los pesos de pronóstico. La importancia de este punto se hace patente cuando se tiene, por ejemplo, una serie estacional y el valor faltante es un "pico estacional". Si el proceso ha sido bien identificado, los pronósticos reflejarán este comportamiento de la serie y el estimador también. Esta característica puede generalizarse a cualquier tipo de comportamiento no estacionario y que haya sido corregido al modelar las series.

"Pronósticos Ponderados" tiene, sin embargo, deficiencias de orden teórico. La primera y más grave es el minimizar la varianza de la ponderación de los dos pronósticos, cuando lo correcto debería haber sido minimizar el MSE (Mean Squared Error o Error Medio Cuadrático) de esta ponderación. Esto equivale a suponer que los dos pronósticos sean insesgados, cosa que no necesariamente es cierta. En cambio, sí podemos suponer

es que X1 sea insesgado respecto al pasado únicamente y que X2 sea insesgado respecto al futuro únicamente. Esto, sin embargo, no basta en forma estricta para lograr la equivalencia de la varianza y del Error Medio Cuadrático.

La segunda deficiencia consiste en que, si bien "Pronósticos Ponderados" toma en cuenta las observaciones siguientes y anteriores a la(s) observacion(es) faltante(s), no lo hace de manera conjunta. Esto es, analiza el futuro y el pasado a fondo, pero de manera aislada. El estimar, por ejemplo, la función de verosimilitud de manera conjunta hubiera implicado un trabajo teórico de enorme dificultad.

La última deficiencia del método es que, finalmente, es una simplificación del método sugerido por Damsleth [3]. El darme cuenta de esto provocó en mí una mezcla de sentimientos: Por un lado, me di cuenta de que iba en la dirección correcta de investigación; pero, por otro lado, sentí que no estaba innovando. En última instancia, lo primero fue más importante, ya que me hizo tener confianza en mis ideas y me impulsó a seguir adelante con otros proyectos de trabajo en el área estadística.

Finalmente, quisiera comentar mi experiencia personal al desarrollar esta tesis. Cuando en un principio el Act. Carlos del Cueto - Director de esta Tesis- y yo comenzamos a discutir el enfoque que se debería dar a este problema, jamás nos imaginamos que resultara una experiencia tan enriquecedora y esti- en tantos aspectos:

1. El problema es de gran actualidad y hemos encontrado varias series donde el método es susceptible de ser aplicado, una de ellas es la que se presentó en este trabajo como serie 3.

2. Me dió la oportunidad de realizar un trabajo de investigación serio y una tesis profesional útil siendo todavía estudiante de Licenciatura.

3. Me permitió reforzar mis conocimientos en un área de la Estadística tan interesante y de tanta actualidad como lo es el Análisis de Series de Tiempo.

4. Permitted resolver un problema que existía desde hace tiempo: Completar satisfactoriamente el Índice de la Bolsa Mexicana de Valores durante Septiembre de 1982.

5. Finalmente, me produjo una gran satisfacción personal el trabajar tan de cerca con alguien como el Act. Carlos del Cueto, a quien debo gran parte de mi formación profesional y gracias a quien quedé convencido de que en México contamos con la capacidad de llevar a cabo proyectos de investigación a muy buenos niveles, algo que me llena de orgullo.

" APENDICES, GRAFICAS Y BIBLIOGRAFIA "

APENDICES Y BIBLIOGRAFIA.

IV.1 INTRODUCCION.

En esta sección final, se ofrece información adicional en forma de Apéndices y Bibliografía. Se presentan las pruebas de Bondad de Ajuste usadas para probar la Normalidad del pseudo-generador de ruido blanco usado en el simulador de Series de Tiempo y los resultados de la estimación de las Series completadas con el método. En la Bibliografía se encuentra toda la información a la que tuvo que recurrirse para poder presentar un panorama del problema y que fueron de gran ayuda en la solución del mismo.

APENDICE IV.1

ESTIMACION DE LA SERIE 1 (PRIMER TROZO).

1. MODELO ESTIMADO : AR(1).

2. ESTIMACION DE LOS PARAMETROS.

PARAMETRO	ESTIMADO	L. INFERIOR	L. SUPERIOR
AR - 1	0.7399	0.49794	0.98188
MEDIA	13.31206	10.81194	15.81219

3. MATRIZ DE CORRELACION DE LOS PARAMETROS.

	1	2
1	1	-0.2165
2	-0.2165	1

4. PRUEBA χ^2 DE LOS RESIDUALES : 11.12406 CON 24 G.L.

5. PRUEBA T DE LOS RESIDUALES : 0.00006.

APENDICE IV.2

ESTIMACION DE LA SERIE 1 (SEGUNDO TROZO INVERTIDO).

1. MODELO ESTIMADO : AR(1).

2. ESTIMACION DE LOS PARAMETROS.

PARAMETRO	ESTIMADO	L. INFERIOR	L. SUPERIOR
AR ~ 1	.89132	.72523	1.05741
MEDIA	15.15454	10.05144	20.25763

3. MATRIZ DE CORRELACION DE LOS PARAMETROS.

	1	2
1	1	-0.2442
2	-0.2442	1

4. PRUEBA χ^2 DE LOS RESIDUALES : 10.06863 CON 22 G.L.

5. PRUEBA T DE LOS RESIDUALES : 0.00035.

APENDICE IV.3

ESTIMACION DE LA SERIE 1 (ORIGINAL).

1. MODELO ESTIMADO : AR(1).

2. ESTIMACION DE LOS PARAMETROS.

PARAMETRO	ESTIMADO	L. INFERIOR	L. SUPERIOR
AR - 1	.83708	.72108	.95307
MEDIA	14.55069	12.09981	17.00195

3. MATRIZ DE CORRELACION DE LOS PARAMETROS.

	1	2
1	1	-0.0742
2	-0.0742	1

4. PRUEBA χ^2 DE LOS RESIDUALES : 11.86653 CON 22 G.L.

5. PRUEBA T DE LOS RESIDUALES : 0.00001.

APENDICE IV.4

ESTIMACION DE LA SERIE 2 (PRIMER TROZO).

1. MODELO ESTIMADO : AR(2).

2. ESTIMACION DE LOS PARAMETROS.

PARAMETRO	ESTIMADO	L. INFERIOR	L. SUPERIOR
AR - 1	-0.31022	-0.64442	0.02398
AR - 2	0.19581	-0.15072	0.54234
MEDIA	53.24910	49.53516	56.96303

3. MATRIZ DE CORRELACION DE LOS PARAMETROS.

	1	2	3
1	1		
2	0.3679	1	
3	0.0016	-0.0218	1

4. PRUEBA χ^2 DE LOS RESIDUALES : 15.6719 CON 21 G.L.

5. PRUEBA T DE LOS RESIDUALES : 0.00288.

APENDICE IV.5

ESTIMACION DE LA SERIE 2 (SEGUNDO TROZO INVERTIDO).

1. MODELO ESTIMADO : AR(2).
2. ESTIMACION DE LOS PARAMETROS.

PARAMETRO	ESTIMADO	L. INFERIOR	L. SUPERIOR
AR - 1	-0.39678	-0.79352	-0.00003
AR - 2	0.11926	-0.30979	0.54832
MEDIA	49.61713	46.91013	52.32411

3. MATRIZ DE CORRELACION DE LOS PARAMETROS.

	1	2	3
1	1		
2	0.3982	1	
3	-0.0403	-0.0161	1

4. PRUEBA χ^2 DE LOS RESIDUALES : 4.33156 CON 21 G.L.

5. PRUEBA T DE LOS RESIDUALES : 0.00005.

APENDICE IV.6

ESTIMACION DE LA SERIE 2 (ORIGINAL).

1. MODELO ESTIMADO : AR(2).

2. ESTIMACION DE LOS PARAMETROS.

PARAMETRO	ESTIMADO	L. INFERIOR	L. SUPERIOR
AR - 1	-0.31529	-0.56842	-0.06216
AR - 2	0.19957	-0.06064	0.45977
MEDIA	51.39027	48.90884	53.87169

3. MATRIZ DE CORRELACION DE LOS PARAMETROS.

	1	2	3
1	1		
2	0.3270	1	
3	-0.0269	-0.0436	1

4. PRUEBA χ^2 DE LOS RESIDUALES : 11.76365 CON 21 G.L.

5. PRUEBA T DE LOS RESIDUALES : 0.00059

APENDICE IV.7

Se simuló una muestra de tamaño 100 de una $N(0,1)$. Para generarla, se usó la función B_Muller en Q^*NIAL de la que se habla en el Capítulo III.

Para probar su Bondad de Ajuste a una distribución Normal, se llevó a cabo una prueba de Lilliefors. Se hicieron también pruebas t de Student sobre su primer momento (Media). Se presentan a continuación los resultados de las pruebas realizadas, así como las estadísticas básicas de la muestra.

IV.7.1 ESTADISTICAS BASICAS:

1. MEDIA :	-0.0169598
2. DESVIACION ESTANDARD:	0.9899
3. SESGO :	-0.1857
4. COEF. DE VARIACION :	-5837.06
5. CURTOSIS :	0.20555
6. RANGO DE LA MUESTRA :	(-3.1 , 2.44)
7. MEDIANA :	0.06041
8. STEAM LEAF Y BOXPLOT (EN LA SIGUIENTE PAGINA).	

IV.7.2 PRUEBAS REALIZADAS:

1. MEDIA : PRUEBA T : -0.17046.

Al compararse con una t de Student al nivel .05 y 97 grados de libertad, no se halló evidencia suficiente para rechazar la hipótesis de que la media fuera cero.

2. BONDAD DE AJUSTE : PRUEBA D: .0552334.

Al comparar la estadística D con el valor crítico de Lilliefors (.08), no se halló evidencia suficiente para rechazar la Hipótesis de Normalidad.

.AS

UNIVARIATE

VARIABLE#X

MOMENTS

N	99	SUM WGT	99	100% MAX
MEAN	-0.0169998	SUM	-1.67007	75% Q3
STD DEV	0.949954	VARIANCE	0.90001	50% MFD
SKEWNESS	-0.165724	KURTOSIS	0.70554	25% Q1
USS	96.0A94	CSS	95.041	0% MIN
CV	-5.037.06	STD MEAN	0.0974947	
FIMFAN=0	-0.17046	PROB> T	0.064	RANGE
SGN RANK	3	PROB> S	0.993814	Q3-Q1
NUM = 0	99			MODE
NORMAL	0.0557334	PROB>D	>.15	

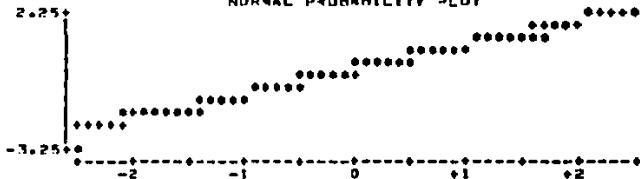
QUANTILES(DEF=4)

2.44529	33%	2.44529	LOWEST	HIGHEST
0.6742	95%	1.4694	-3.1	1.4694
0.0641	90%	1.2777	-1.999	1.7314
-0.6371	10%	-1.2714	-1.84277	1.8557
-3.1	5%	-1.7103	-1.7301	2.17
	1%	-3.1	-1.7103	2.445

EXTREMFS

STEM	LEAF	#	BOXPLOT
2	14	2	
1	54579	5	
1	0111344	7	
0	5455466677788899999	14	-----+-----
0	11111122222333444	14	-----+-----
-0	444332211111000	16	*
-0	999876666655555	14	-----+-----
-1	3322221110	10	
-1	677765	6	
-2	0	1	
-2	0	1	
-3	1	1	0

NORMAL PROBABILITY PLOT



APENDICE 4.8

RESUMEN DE LOS MODELOS PARA COMPLETAR LA SERIE DE LA
BOLSA MEXICANA DE VALORES.

I. SERIE BMV1

PERIODICIDAD : DIARIA. TRANSFORMACION(Z) = $\log\left(\frac{Z}{t}\right)$

PERIODO: 17 MAYO-31 AGOSTO, 1982.

UNA DIFERENCIA.

RESUMEN DE LA ESTIMACION DE LOS PARAMETROS:

TIPO	ORDEN	ESTIMADO	L. INF.	L. SUP.

MA(1)	1	-0.3143	-0.6185	-0.01015
INT(1)	***** (3-4) *****			
OMEGA1	0	-0.03591	-0.044	-0.02714
INT(2)	***** (35-69) *****			
OMEGA1	0	-0.02402	-0.03362	-0.01443
OMEGA2	3	-0.021529	-0.03267	-0.01038
INT(3)	***** (43-69) *****			
OMEGA1	0	0.07998	0.0631	0.09282
OMEGA2	1	-0.03984	-0.05566	-0.024027
OMEGA3	2	0.02083	0.00509	0.036574
OMEGA4	3	-0.0270	-0.04208	-0.012028
INT(4)	***** (23-23) *****			
OMEGA1	0	0.05261	0.03774	0.067502

FAC DE LOS RESIDUALES:

1	2	3	4	5	6	7	8	9	10
-0.01	0.07	0	-0.1	-0.03	0.07	0.08	0.07	-0.15	0

CORRELACION DE LOS PARAMETROS: EXISTE UNA CORRELACION DE 0.85 ENTRE LOS PARAMETROS 4 Y 3. LAS DEMAS SON INFERIORES A 0.35.

APENDICE 4.9

RESUMEN DE LOS MODELOS PARA COMPLETAR LA SERIE DE LA
BOLSA MEXICANA DE VALORES.

II. SERIE BMV2

PERIODICIDAD : DIARIA. TRANSFORMACION(Z) = $\log\left(\frac{Z}{t}\right)$

PERIODO: 20 SEPTIEMBRE, 1982 - 21 ENERO, 1983

UNA DIFERENCIA.

RESUMEN DE LA ESTIMACION DE LOS PARAMETROS:

TIPO	ORDEN	ESTIMADO	L. INF.	L. SUP.

MA(1)	1	-0.51136	-0.9710	-0.05172
INT(1)	***** (3-58) *****			
OMEGA1	0	-0.03142	-0.0507	-0.0121
OMEGA2	1	0.0342	0.0151	0.0532
INT(2)	***** (10-10) *****			
OMEGA1	0	-0.01785	-0.0278	-0.00789
INT(3)	***** (16-58) *****			
OMEGA1	0	-0.02768	-0.4364	-0.01172
OMEGA2	1	0.024	0.008829	0.03917
INT(4)	***** (23-23) *****			
OMEGA1	0	-0.01665	-0.0254	-0.00787
INT(5)	***** (33-58) *****			
OPERADOR: (1-B)				
OMEGA1	0	0.01603	-0.0044	0.036
OMEGA2	1	-0.0306	-0.5288	-0.0083
OMEGA3	3	0.0153	-0.0012	0.0319

INT(6) ***** (38-39) *****

OMEGA1	0	-0.01367	-0.02753	-0.00018
OMEGA2	1	0.02554	0.01153	0.03954

INT(7) ***** (42-48) *****

OPERADOR: (1-8)

OMEGA1	0	-0.05459	-0.0762	-0.03294
OMEGA2	1	0.087477	0.06505	0.10990
OMEGA3	3	0.063299	0.03990	0.08669

FAC DE LOS RESIDUALES:

1	2	3	4	5	6	7	8	9	10
-0.07	0.04	0.02	-0.16	0.07	0.06	-0.16	-0.16	-0.05	-0.03

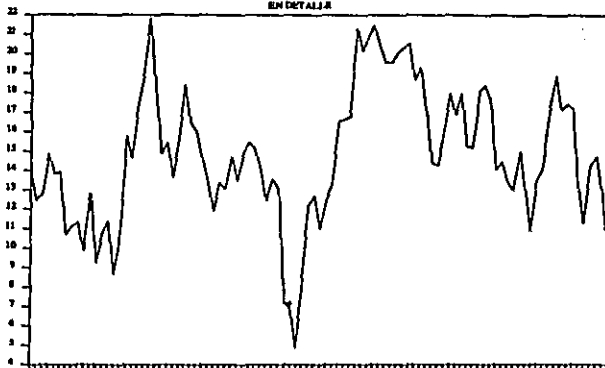
CORRELACION DE LOS PARAMETROS: NO EXISTE NINGUNA CORRELACION
MAYOR DE 0.40 EN VALOR ABSOLUTO.

GRAFICAS.

A continuación, se presentan las graficas de las tres series usadas para mostrar el ajuste que proporciona la tecnica de "Pronosticos Ponderados" como se verían originalmente. Tambien se presentan en detalle las graficas de la observación faltante contra su valor estimado.

SERIE 1

EN DETALLE

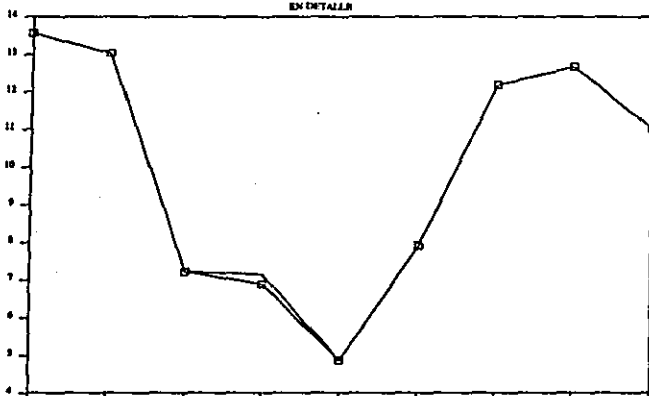


— VALOR OBSERVADO

+ VALOR ESTIMADO

SERIE 1

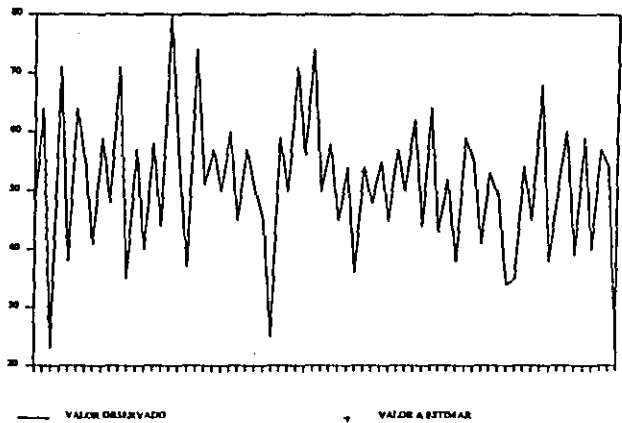
EN DETALLE



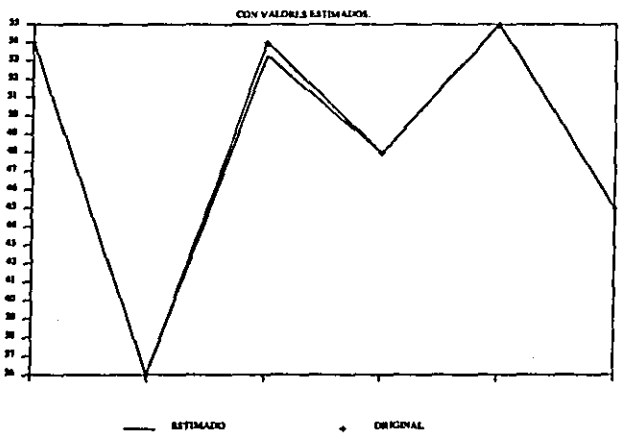
□ VALOR OBSERVADO

— VALOR ESTIMADO

SERIE 2

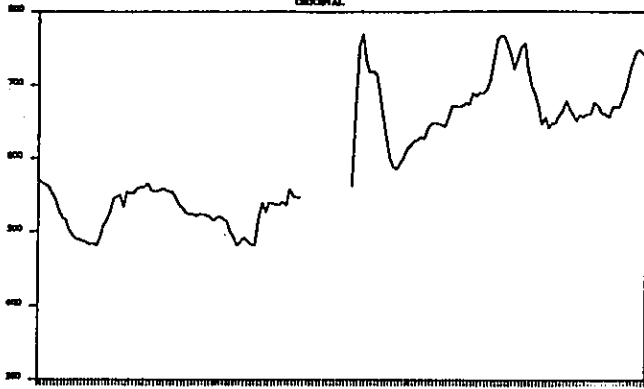


SERIE F



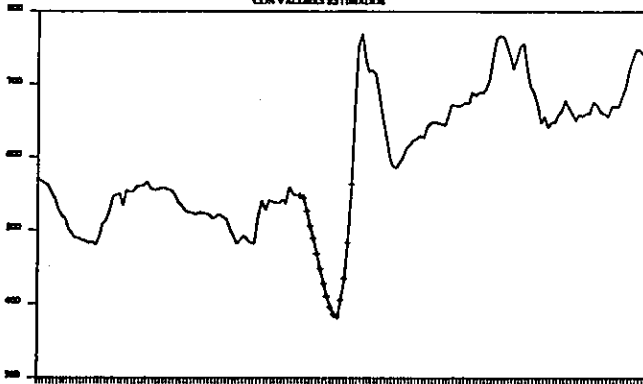
INDICE DE LA BMV.

ORIGINAL



INDICE DE LA BMV.

CON VALORES ESTIMADOS



INDICE DIARIO DE LA BMV.

(17 MAYO- 31 AGOSTO)

(20 SEPT - 21 ENE)

T	Z(T)	T	Z(T)
1	571.3	1	563
2	567.19	2	670
3	564.05	3	750.75
4	561.42	4	768.94
5	551.75	5	733.98
6	543.58	6	716.79
7	527.97	7	719.31
8	519.6	8	713.7
9	516.21	9	686.46
10	502.69	10	653.75
11	495.99	11	628.22
12	490.5	12	601.28
13	489.64	13	588.58
14	487.03	14	585.4
15	486.52	15	593.18
16	482.27	16	600.26
17	474.91	17	612.08
18	480.6	18	617.18
19	471.91	19	623.02
20	509.04	20	625.27
21	514.98	21	630.07
22	527.17	22	626.44
23	545.48	23	641.52
24	543.04	24	647.22
25	550.37	25	649.21
26	552.73	26	644.93
27	552.03	27	645.74
28	552.19	28	643.29
29	557.08	29	655.3
30	557.5	30	671.72
31	561.01	31	671.2
32	561.04	32	689.91
33	562.07	33	671.58
34	563.71	34	676.63
35	565.1	35	673.7
36	567.53	36	695.27
37	567.49	37	694.14
38	567.39	38	697.16
39	565.35	39	699.11
40	564.5	40	697.46
41	564.93	41	695.95
42	565.39	42	695.69
43	565.99	43	691.08
44	565.45	44	689.13
45	565.75	45	685.61
46	564.77	46	682.68
47	564.45	47	683.39
48	564.55	48	681.33
49	564.29	49	681.41
50	562.37	50	681.36
51	562.43	51	682.53
52	564.1	52	680.39

53	519
54	521.09
55	517.2
56	514.24
57	500.61
58	491.96
59	481.35
60	486.81
61	492.08
62	487.21
63	482.53
64	481.83
65	519.96
66	539.71
67	527.38
68	540.3
69	538.6
70	537.22
71	536.62
72	541.62
73	535.65
74	558.47
75	548.01
76	547.57
77	546.45

53	497.03
54	487.16
55	489.26
56	446.6
57	456.02
58	641.1
59	447.76
60	448.01
61	457.55
62	464.06
63	477.75
64	467.7
65	450.3
66	451.02
67	458.6
68	456.49
69	440.4
70	460.33
71	474.35
72	471.32
73	462.44
74	460.43
75	485.81
76	449.04
77	447.47
78	470.04
79	471.25
80	475.31
81	473.14
82	471.54
83	477.16
84	477.19
85	477.16
86	477.19

BIBLIOGRAFIA

- [1] BOX, GEORGE E.P., JENKINS, GWILYM M.
" TIME SERIES ANALYSIS: FORECASTING AND CONTROL".
HOLDEN-DAY
OAKLAND, CA.
1976.
USA.
- [2] CHOW, GREGORY., LIN, AN-LOH.
"BEST LINEAR UNBIASED ESTIMATION OF MISSING OBSERVATIONS
IN AN ECONOMIC TIME SERIES".
JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION.
VOL. 71, NO. 335.
P. 719-721.
SEPTIEMBRE, 1976.
- [3] DAMSLETH, ERWIND.
"INTERPOLATING MISSING VALUES IN A TIME SERIES"
SCANDINAVIAN JOURNAL OF STATISTICS.
VOL. 7. NO. 7.
P. 33-39
1980.
- [4] DE ALBA, ENRIQUE., MENDOZA, YOLANDA.
"DISAGGREGATION AND FORECASTING: A BAYESIAN ANALYSIS".
(ARTICULO INEDITO PRESENTADO EN LA
UNIVERSIDAD DE CHICAGO).
1985.
COPIA EN MI PROPIEDAD.
- [5] FERREIRO, OSVALDO M.
"ESTIMATING UNOBSERVED VALUES IN TIME SERIES"
(EXAMEN PRELIMINAR DE DOCTORADO.
UNIVERSIDAD DE WISCONSIN.
DEPARTAMENTO DE ESTADISTICA.)
JULIO 15, 1981.
INEDITO. COPIA EN MI PROPIEDAD.
- [6] HARVEY, C., PIERSE, R. G.
"ESTIMATING MISSING OBSERVATIONS IN ECONOMIC TIME SERIES".
JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION.
VOL. 79. NO. 385.
P. 126-131.
MARZO, 1984.
- [7] JONES, RICHARD H.
"MAXIMUM LIKELIHOOD FITTING OF ARMA MODELS TO TIME SERIES
WITH MISSING OBSERVATIONS".
TECHNOMETRICS.
VOL. 22. NO. 3.
AGOSTO, 1980.

- [8] KOHN, ROBERT., ANSLEY, CRAIG.
 "ESTIMATION, PREDICTION AND INTERPOLATION FOR ARIMA
 MODELS WITH MISSING DATA".
 JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION.
 SEPTIEMBRE 1986.
 VOL. 81. NO. 395.
 P. 751-761.
- [9] MOOD, ALEXANDER ET AL.
 "INTRODUCTION TO THE THEORY OF STATISTICS".
 MC. GRAW-HILL.
 1980.
- [10] MENZENFRICKE, ULRICH., GUTMANN, IRVING.
 "BAYESIAN INFERENCE IN MULTIVARIATE REGRESSION WITH
 MISSING OBSERVATIONS ON THE RESPONSE VARIABLE".
 JOURNAL OF BUSINESS & ECONOMIC STATISTICS
 VOL. 1, NO. 3
 JULIO , 1983 .
- [11] ROBINSON, P. M.
 "ESTIMATION OF A MODEL FOR ELECTRICITY UTILITY
 DEMAND IN THE PRESENCE OF MISSING OBSERVATIONS".
 PROC. DOE STATISTICAL SYMPOSIUM.
- [12] WILCOX, JAMES A.
 "DISAGREGATING DATA USING RELATED SERIES".
 JOURNAL OF BUSINESS & ECONOMIC STATISTICS .
 VOL. 1, NO. 3.
 JULIO , 1983.
- [13] WILSON, G.T., BRUBACHER, E.R.
 "INTERPOLATING TIME SERIES WITH APPLICATION
 TO THE ESTIMATION OF HOLIDAY EFFECTS ON ELEC-
 TRICITY DEMAND".
 JRSS C (APPL. STATISTICS).
 VOL .25. NO. 2
 P. 107-116.
- [14] ZELLNER, A.
 "ON THE ANALYSIS OF FIRST ORDER AUTOREGRESSIVE
 MODELS WITH INCOMPLETE DATA".
 INTERNATIONAL ECONOMIC REVIEW.
 VOL. 7. NO. 2.
 P. 72-76.
- [15] "D'NIAL USER'S GUIDE".
 QUEEN'S UNIVERSITY.
 KINGSTON, CANADA.
 DICIEMBRE, 1983.