



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

**ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
ACATLAN**

**DISERTACION DE TECNICAS PARA AJUSTAR
CURVAS SUAVES A SERIES DE TIEMPO**



T E S I S
QUE PARA OBTENER EL TITULO DE:
A C T U A R I O
P R E S E N T A N :
CARLOS ALBERTO CADENA SANDOVAL
HUGO REYES MARTINEZ

MEXICO, D. F.

1984



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central

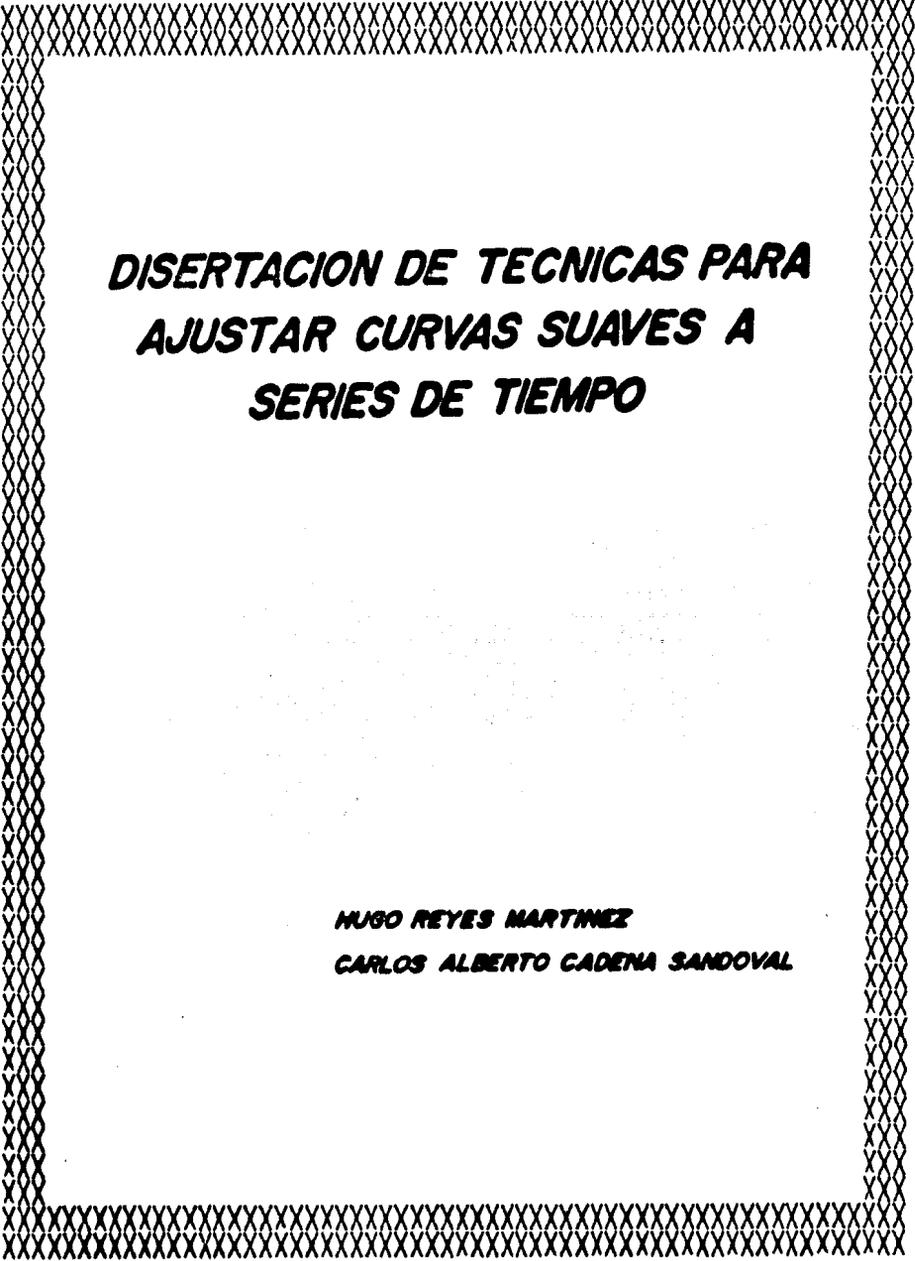


UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

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.



**DISERTACION DE TECNICAS PARA
AJUSTAR CURVAS SUAVES A
SERIES DE TIEMPO**

HUGO REYES MARTINEZ

CARLOS ALBERTO CADENA SANDOVAL

INDICE

Agradecimientos	
Antecedentes	
Introducción	2
Objetivos	6

CAPITULO UNO

SERIES DE TIEMPO

1.1 Definición y generalidades	8
1.2 Mínimos cuadrados	16
1.2.1 Modelo lineal general	16
1.2.2 Modelo no lineal	30
1.3 Descomposición	32
1.4 Promedio móvil	46
1.4.1 Promedio móvil de primer orden	47
1.4.2 Promedio móvil de orden superior	51
1.4.3 Promedio móvil por mínimos cuadrados	53
1.4.4 Propiedades	59
1.4.5 Valores en los extremos	61
1.4.6 Promedio móvil centrado	67

1.5	Autocorrelación	70
1.6	Estacionalidad	76
1.6.1	Modelos de ajuste estacional	79

CAPITULO DOS

TEORIA DE LA APPROXIMACION

2.1	Teoría de la aproximación	87
2.1.1	Áreas de aplicación	87
2.1.2	Desarrollo de la teoría	89
2.1.3	Procedimiento general de la aproximación	91

FUNCIONES SPLINE

2.2	Spline de interpolación	94
2.2.1	Introducción	94
2.2.2	Idea intuitiva de la función spline	97
2.2.3	Definición de la función spline cúbica y conceptos básicos	99
2.2.4	Propiedades básicas	101
2.2.5	Construcción de la función spline de interpolación	103
2.2.6	Razones para aproximar con funciones spline	108
2.2.7	Aplicaciones de la función spline	109
2.2.8	Otros tipos de spline	110
2.2.9	Spline generalizado	112

2.3	Spline de aproximación	113
2.3.1	Descripción del problema	113
2.3.2	Deducción de los parámetros de la función spline	118
2.3.3	Determinación del parámetro lagrangiano	123
2.3.4	Demostración de que $F(\lambda)$ es convexa y decreciente	126
2.3.5	Cálculo del parámetro lagrangiano	129
2.4	Spline por mínimos cuadrados	132
2.4.1	Desarrollo	132

CAPITULO TRES

COMPARACION DE LAS TECNICAS

3.1	Planteamiento del problema	138
3.2	Simulación de una serie	148
3.2.1	Simulación	148
3.2.2	Construcción del modelo	150
3.3	Comparación	159
3.3.1	Similitudes generales	159
3.3.2	Similitudes particulares	160
3.3.3	Diferencias generales	162
3.3.4	Diferencias particulares	164

EVALUACION

3.4	Evaluación de mínimos cuadrados	169
3.5	Evaluación del promedio móvil	173
3.6	Evaluación del spline de aproximación	188
3.7	Evaluación del spline por mínimos cuadrados	213

CONCLUSIONES	219
CONCLUSIONES PARTICULARES	219
CONCLUSIONES GENERALES	225
OBJETIVOS ALCANZADOS	228-a
APENDICE A	229
APENDICE B	242
APENDICE C	265
BIBLIOGRAFIA	272

ANTECEDENTES

Al cursar la materia de Planeación Financiera, el temario a cubrir, fue repartido entre el grupo para efectuar exposiciones por los alumnos. En esta materia conocimos el tema de Series de Tiempo, que fue el que nos tocó exponer y por lo tanto hubo que investigar acerca de los números índice y sus aplicaciones. Al estudiar el tema se hizo patente la poca información en castellano, la extensión del mismo y las variadas conexiones con otras disciplinas, situación que nos hizo pensar que era un campo propicio para elaborar una tesis.

Posteriormente, profundizamos en el estudio de las Series de Tiempo, con el objeto de buscar un problema que resultara interesante, para discutir en una tesis.

Cuando asistimos a una conferencia, en la FNEP Aca-

tlán, sobre Funciones Splines, se nos ocurrió la idea de comparar las técnicas de ajuste de curvas suaves existentes dentro de la Estadística, contra la mencionada en dicha conferencia, que al parecer tenía grandes posibilidades de suavizar a una Serie de Tiempo aún cuando pertenece al campo del Análisis Numérico.

INTRODUCCION

La Estadística y la Teoría de la Aproximación constituyen nuestro marco teórico (capítulos 1 y 2) siendo los mínimos cuadrados y los distintos tipos de promedio móvil, las técnicas utilizadas por parte de la primera, mientras que en lo que toca a la segunda, se consideraron varias clases de splines.

La tendencia, es la componente más importante de una serie de tiempo y su comportamiento no siempre es de completa suavidad. Es por esta razón que resulta interesante aplicar los métodos de ajuste de curvas a una serie de tiempo simulada, con la facilidad de que representen a la tendencia de dicha serie y ver con cuánta eficiencia lo logran.

Los procedimientos de mínimos cuadrados y promedio móvil, han sido los que tradicionalmente se han empleado para la estimación de la tendencia.

Por otro lado, una de las características más sobresalientes de la función spline es la suavidad, entonces se pensó que si éste se aplicara a una serie de tiempo, podría obtenerse una aproximación buena a la tendencia.

Este intento de aplicar herramientas fuera de la estadística a series de tiempo, nos permitirá obtener una idea más clara de la comparación y evaluación de cada una de las técnicas.

El marco teórico al cual pertenece el presente trabajo, está estructurado en una forma concreta, pero sin perder su generalidad. Sólo se tratan aspectos particulares, cuando éstos repercuten fuertemente en el desarrollo de la investigación.

Esta información fue recopilada, seleccionada, traducida, adaptada y condensada de diversos libros y revistas técnicas consultados en las bibliotecas del IIMAS y del Instituto de Matemáticas de la UNAM, para hacerla accesible a estudiantes de diversas disciplinas.

En el capítulo primero, se da una introducción a las series de tiempo contemplando algunos de sus temas más importantes, Posteriormente se detalla la idea fundamental de mínimos cuadrados en forma general, utilizando la notación matricial.

Se incluyeron algunos tipos de promedio móvil y se hizo énfasis en el presentado por Kendall (1976), cuyos pesos se obtuvieron por mínimos cuadrados, el cual tiene una ventaja sobre los demás promedios, que señalaremos en el capítulo correspondiente.

Además se trata por separado a la estacionalidad, por ser el problema principal con la estimación de la tendencia.

En el capítulo segundo, se da una breve presentación a la teoría de la aproximación y se trata a la función spline de interpolación y de aproximación. Este último basado en el algoritmo propuesto por C. H. Reinsch, el cual se desarrolló, pues los resultados expuestos en su artículo son bastante concretos y no se da mucha explicación a éstos.

Finalmente, se expone el spline por mínimos cuadrados, presentado en el libro de Carl de Boor (1978), el cual tiene como idea fundamental el principio de mínimos cuadrados, motivo que nos hizo pensar que este método daría los mejores resultados.

En el capítulo tercero, planteamos el problema y se explica cómo se genera la serie de tiempo hipotética, pues -- fue necesario simularla para conocer perfectamente sus componentes y características. Así, con el control de los parámetros, podríamos medir la eficiencia con que estiman a la tendencia cada uno de los métodos.

En la comparación se tratará de señalar las similitudes y diferencias de cada técnica; mientras que en la evaluación se establece numéricamente cuál(es) estiman mejor a la tendencia de la serie simulada.

Aquí mismo se presentan las pruebas realizadas, sus resultados y gráficas respectivas.

La secuencia de este capítulo está hecha a base de pre--guntas y respuestas que aclaran y señalan las ideas que se siguieron para llegar a las conclusiones.

OBJETIVOS

El orden en que se mencionan, constituye la secuencia que seguirá el trabajo:

- 1) Recopilar y seleccionar las principales técnicas de ajuste de curvas suaves, tanto de la estadística como de la teoría de splines.
- 2) Elaborar un esquema de las teorías a las que pertenecen cada una de las técnicas de ajuste de curvas suaves.
- 3) Presentar los métodos y sus programas con un lenguaje y notación simples, para que puedan ser empleados en forma práctica, por estudiantes y profesionistas de cualquier disciplina.

4) Simular una serie de tiempo para aplicarle cada una de las técnicas elegidas. Esto nos permitirá tener control sobre sus parámetros, lo que a su vez permite medir con precisión la eficiencia de las técnicas.

5) Probar bajo distintas suposiciones y condiciones todos los métodos de ajuste de curvas suaves aquí expuestos, para visualizar cuál o cuáles de ellos estiman mejor a la tendencia.

6) Establecer un criterio cuantitativo para la evaluación de los métodos.

7) Comparar y evaluar cada una de las técnicas seleccionadas y, analizar sus resultados. Este objetivo constituye la parte medular de nuestra tesis.

CAPITULO**U N O****SERIES DE TIEMPO**

1.1.- DEFINICION Y GENERALIDADES.

Con el fin de situarnos dentro del presente tema, enunciaremos la definición de Series de Tiempo más común, aunque no formal:

Definición. Si tenemos observaciones de algún fenómeno que varía con respecto al tiempo, éstas generan un conjunto ordenado llamado "Serie de tiempo". Es decir, estaremos asociando unidades de tiempo (generalmente iguales y unitarias) con valores reales. En algunas ocasiones se hace necesario trabajar con series multivariadas, que son las que asocian a cada unidad de tiempo, un vector de números reales.

Para ilustrar esto, daremos algunos ejemplos de series -

de tiempo, junto con sus gráficas respectivas.

Ejemplo (1).

La serie de tiempo anual llamada Tasa Bruta de Mortalidad Mexicana (1960-74) (ver fig. 1) es:

1960 — 11.1	1968 — 9.6
1961 — 10.3	1969 — 9.4
1962 — 10.4	1970 — 9.6
1963 — 10.3	1971 — 8.7
1964 — 9.8	1972 — 8.8
1965 — 9.4	1973 — 8.2
1966 — 9.6	1974 — 7.5
1967 — 9.2	TASA PROMEDIO: 9.4

Ejemplo (2).

A continuación, se muestran dos series de tiempo semanales del boletín de la Bolsa de Valores Mexicana, correspondientes a bancos (ver fig. 2):

COTIZACIONES DE ACCIONES BANCARIAS 1980

BANCOMER	50.50	50.00	49.50	48.00
BANAMEX	275.00	271.00	286.00	272.00
FECHA	8-FEB	15-FEB	22-FEB	29-FEB

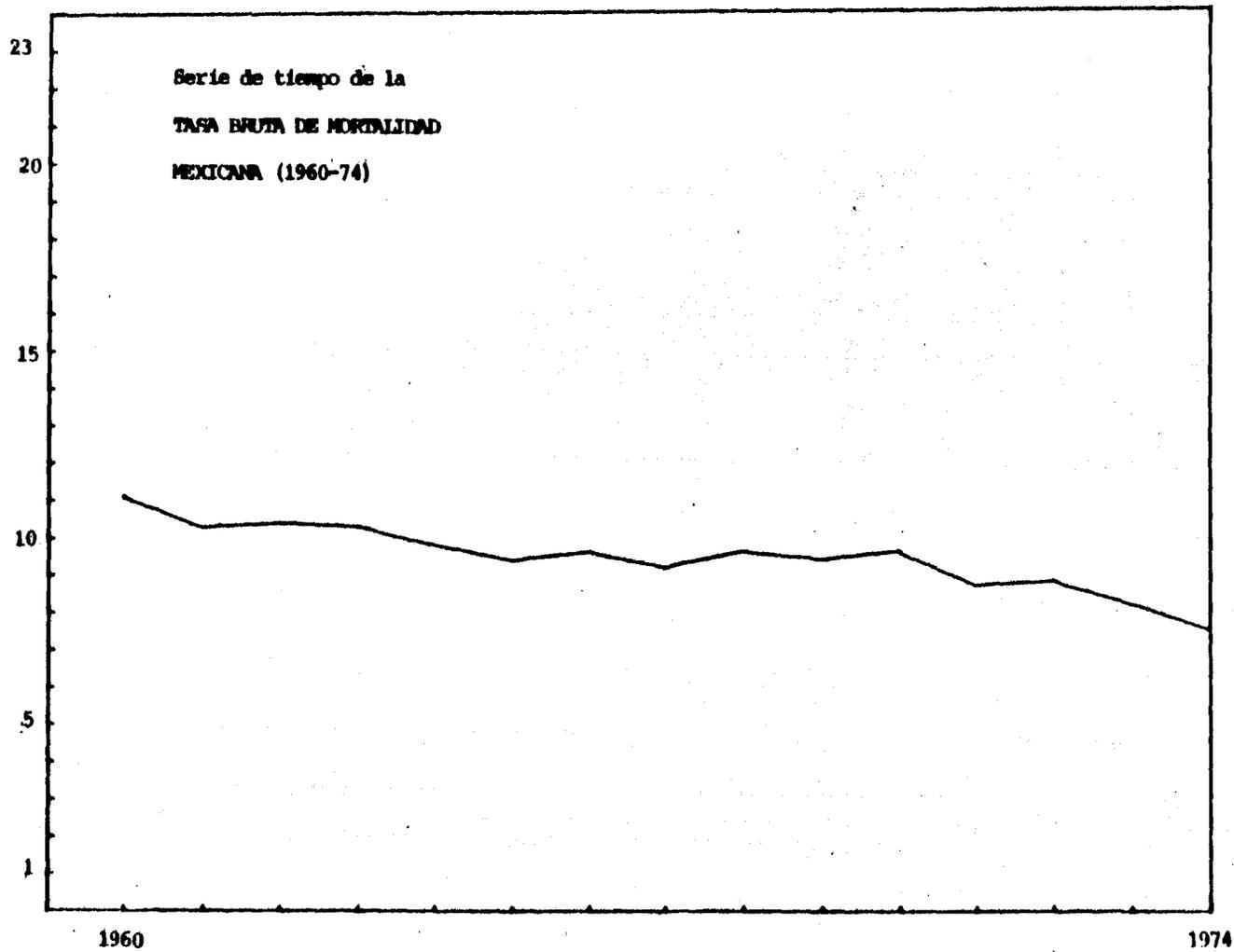
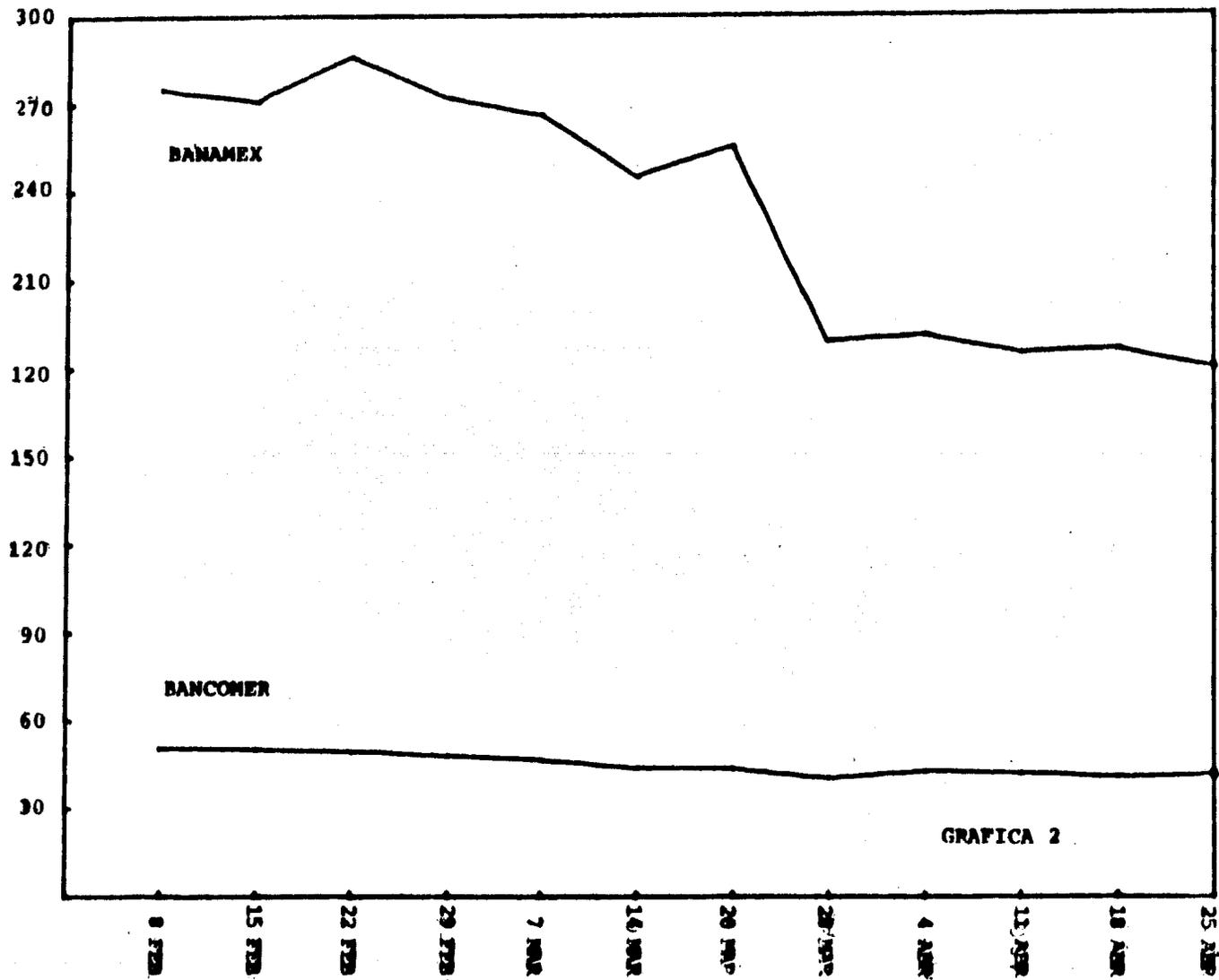


FIGURA 1



GRAFICA 2

BANCOMER	46.50	43.75	43.50	40.25
BANAMEX	266.00	245.00	255.00	189.00
FECHA	7-MAR	14-MAR	20-MAR	28-MAR
BANCOMER	42.50	42.00	40.75	41.50
BANAMEX	191.00	185.50	186.00	180.00
FECHA	4-ABR	11-ABR	18-ABR	25-ABR

Observe que para graficar el ejemplo (1), es conveniente hacer al año 1960 como el origen (año cero), al año 1961 como el año uno, al año 1962 como el año dos, ..., etc. Una convención semejante se hará siempre que se quiera graficar una serie de tiempo. Ejemplos típicos de series de tiempo son: La población económicamente activa de un país, la producción de una empresa particular, etc. Estas series se -- pueden emplear de diversas formas:

- a) en una empresa, para predecir ventas, necesidades de equipo, capacidad de almacenamiento, etc.
- b) en macroeconomía, se utilizan para la estimación de tasas futuras de crecimiento, o para cualquier otra variable económica.

Es importante mencionar desde ahora, que el tema de series de tiempo, puede alcanzar grados altos de sofisticación matemática.

Durante los últimos 25 años los métodos estadísticos de de dicados al análisis de series de tiempo, se han desarrollado considerablemente, pero especialmente el análisis espectral, que descompone una serie, en sus componentes frecuenciales, para poderla analizar detalladamente. Han sido dis tintas y muy diversas las disciplinas a donde ha beneficiado este hecho, en particular mencionaremos una muy importan te que es la econometría.

La teoría formal del análisis de las series de tiempo, - fue construída por Khintchine (1934) y Wiener (1930). A ellos se debe la aplicación del análisis armónico de Fourier, para explicar la relación que hay entre la función de autocorrelación de un proceso estocástico y la función de densi dad espectral, donde ésta última, no es más que la transfor mada de Fourier de la otra.

El objetivo principal del análisis de las variaciones de una serie de tiempo, a través de cualquier método, es bus car ciertas regularidades y patrones medibles, utilizando - los datos pasados, los cuales podríamos suponer, que se volverán a presentar, en forma semejante en el futuro. Un conocimiento claro y detallado del pasado, permite hacer buenas predicciones, que serán empleadas en la toma de decisio nes.

Un rasgo característico de las series de tiempo en compa

ración con las demás materias estadísticas, es que sus observaciones ocurren en un cierto orden, determinado por el tiempo.

Algo por lo cual podríamos estar interesados en investigar, acerca de una serie de tiempo, sería por ejemplo la relación que hay entre el valor de un elemento de la serie y el siguiente; o bien la correlación que hay a lo largo de la serie.

Llamaremos longitud de una serie de tiempo, al número de observaciones que la constituyen; donde a cada observación le llamaremos término de la serie de tiempo.

1.2.- MINIMOS CUADRADOS.

Esta sección tiene un doble propósito:

- a) Cimentar la idea esencial del principio de Mínimos Cuadrados, junto con el desarrollo de su técnica.
- b) Introducir una notación matricial, que emplearemos posteriormente y que nos permite generalizar de una manera compacta, esto nos facilitará la notación y algunas demostraciones.

En muy raras ocasiones, las variaciones de una serie de tiempo, dependen de un sólo factor, es por eso, que presentaremos a continuación un modelo general con k variables.

1.2.1.- MODELO LINEAL GENERAL.

Consideremos un modelo general que tenga k variables, mediante una relación lineal, entre la variable dependiente Y y $k - 1$ variables explicativas X_1, X_2, \dots, X_k ; además,

una serie de términos de perturbación, que denotaremos con "a".

Si contamos con una muestra de n observaciones para Y y X, podemos escribir nuestro modelo lineal general como:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_M X_{iM} + a_i \quad \text{--- (1.2.1)}$$

con $i = \overline{1, n}$ y $X_{i1} = 1$ para toda i

En donde los coeficientes " β " y los parámetros de distribución para "a", se desconocen. El problema consiste en estimar a tales incógnitas.

Así pues, como en (1.2.1) se tienen n ecuaciones, las escribiremos matricialmente como:

$$y = X\beta + a \quad \text{--- (1.2.2)}$$

donde:

$$y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & X_{21} & \dots & X_{M1} \\ 1 & X_{22} & \dots & X_{M2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{Mn} \end{pmatrix};$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_M \end{pmatrix}; \quad a = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix}; \quad \text{--- (1.2.3)}$$

Note que X_{ik} representa la i -ésima observación de la variable X_k . Esto significa que los subíndices de la matriz X , están colocados en forma invertida a la que usualmente se acostumbra.

Las suposiciones fundamentales para la estimación del vector de coeficientes " β " son:

- 1) $E(a) = 0$
 - 2) $E(aa') = \sigma^2 I_n$
 - 3) La matriz X tiene elementos fijos.
 - 4) El rango de X cumple que $k < n$.
- (1.2.4)

Con a' como el vector renglón, transpuesto al vector columna a . El producto $a \cdot a'$, nos da una matriz simétrica de orden n . I_n representa a la matriz identidad de orden n .

$$E(aa') = \begin{pmatrix} E(a_1^2) & E(a_1 a_2) & \dots & E(a_1 a_n) \\ E(a_2 a_1) & E(a_2^2) & \dots & E(a_2 a_n) \\ \dots & \dots & \dots & \dots \\ E(a_n a_1) & E(a_n a_2) & \dots & E(a_n^2) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

En esta última matriz se verifica que:

- a) $E(a_i^2) = \sigma^2$ para toda i , indicando que hay una varianza constante.
- b) $E(a_i a_{i+k}) = 0$ para $k \neq 0$, indicando que las a_i no están correlacionadas.

Usando la relación (1.2.1), bajo las condiciones fundamentales, le podremos aplicar el principio de mínimos cuadrados, para estimar los parámetros buscados, y obtener así la recta que ajuste mejor, a los datos de la muestra. Hagamos:

$$b = \{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k\}$$

un vector columna de estimadores para " β " de modo que

$$y = Xb + e \quad \text{-----} \quad (1.2.5)$$

donde e es un vector columna con n residuales. Si lo despejamos:

$$e = y - Xb \quad \text{-----} \quad (1.2.6)$$

Por tanto, la suma de residuales cuadrados será:

$$\sum_{i=1}^n e^2 = e'e \quad \text{-----} \quad (1.2.7)$$

$$= y'y - 2b'X'y + b'X'Xb \quad \text{-----} \quad (1.2.8)$$

donde e' y X' , son las transpuestas de e y X respectivamente. La matriz $X'X$, es no singular, simétrica y de orden k . Note que $e'e$ será un indicador del grado de eficiencia con que se realiza el ajuste.

Para encontrar el vector "b" que mínimice la suma de residuales cuadrados recurriremos al procedimiento de máximos y mínimos:

$$\frac{\partial}{\partial b} (e'e) = -2X'y + 2X'Xb \quad \text{-----} \quad (1.2.9)$$

igualando a cero esta expresión obtenemos:

$$X'Xb = X'y \quad \text{-----} \quad (1.2.10)$$

y de acuerdo con la cuarta suposición fundamental podremos despejar:

$$b = (X'X)^{-1} X'y \quad \text{-----} \quad (1.2.11)$$

este es un resultado muy importante para los estimadores de mínimos cuadrados.

Para encontrar la media y la varianza de "b" sustituimos (1.2.2) en (1.2.11):

$$b = (X'X)^{-1} X'(X\beta + a) \quad \text{-----} \quad (1.2.12)$$

$$b = \beta + (X'X)^{-1} X'a \quad (1.2.13)$$

Como el proceso muestral se repite, entonces los elementos de X permanecen fijos para cada muestra, pero generando una serie de "a" que serán distintas para cada muestra, y por consiguiente un vector diferente para "b".

Sacando esperanzas en ambos miembros de (1.2.13) tendremos:

$$E(b) = E(\beta) + E[(X'X)^{-1} X'a]$$

Puesto que X permanece fija entonces:

$$E(b) = \beta + (X'X)^{-1} X'E(a)$$

y como $E(a) = 0$, tenemos

$$E(b) = \beta \quad (1.2.14)$$

De aquí que, hemos probado que los estimadores de mínimos cuadrados son insesgados. Así, la recta que ajusta mejor las n observaciones en Y y X es:

$$Y_r = b_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k \quad (1.2.15)$$

A la recta Y_r , Galton le llamó línea de regresión en su trabajo "Regression Towards Mediocrity in Hereditary Stature" (1885).

Debido a la forma en que fue obtenida la línea de regresión, presenta dos propiedades importantes:

$$\text{I) } \sum_{i=1}^n (Y_i - Y_r) = 0$$

$$\text{II) } \sum_{i=1}^n (Y_i - Y_r)^2 < \sum_{i=1}^n (Y_i - Y_o)^2$$

donde Y_o es cualquier otra línea que pretenda describir el comportamiento central de los puntos Y_i . Esta segunda propiedad, nos asegura que la línea de regresión Y_r , es la que ajusta mejor a los puntos Y_i , puesto que la desviación — que hay entre ambos es mínima.

Ahora bien:

$$E|(b - \beta)(b - \beta)'| =$$

$$\begin{vmatrix} E(b_1 - \beta_1)^2 & E(b_1 - \beta_1)(b_2 - \beta_2) & \dots & E(b_1 - \beta_1)(b_n - \beta_n) \\ E(b_2 - \beta_2)(b_1 - \beta_1) & E(b_2 - \beta_2)^2 & \dots & E(b_2 - \beta_2)(b_n - \beta_n) \\ \dots & \dots & \dots & \dots \\ E(b_n - \beta_n)(b_1 - \beta_1) & E(b_n - \beta_n)(b_2 - \beta_2) & \dots & E(b_n - \beta_n)^2 \end{vmatrix}$$

————— (1.2.16)

Debido a que (1.2.14) nos muestra que $E(b_i) = \beta_i$ para $i = \overline{1, k}$, entonces podemos deducir que $E(b_i - \beta_i)^2$ es la varianza de "b"; mientras que $E(b_i - \beta_i)(b_j - \beta_j)$ con $i \neq j$, es la covarianza de β_i y β_j . A esta matriz se le conoce como matriz de varianza-covarianza y la denotaremos $\text{var}(b)$.

De la relación (1.2.13) se tiene:

$$b - \beta = (X'X)^{-1} X'a$$

y como ya se vió que:

$$\text{var}(b) = E[(b - \beta)(b - \beta)']$$

Entonces:

$$\text{var}(b) = E[(X'X)^{-1} X'aa'X(X'X)^{-1}] \quad \text{--- (1.2.17)}$$

ya que $(X'X)^{-1}$ es simétrica, entonces su transpuesta es la misma matriz. Como X es fija, entonces:

$$\text{var}(b) = (X'X)^{-1} X' E(aa')X(X'X)^{-1}$$

por la segunda suposición fundamental

$$\text{var}(b) = (X'X)^{-1} X'\sigma^2 I X(X'X)^{-1}$$

Dado que, σ^2 es un escalar, lo colocamos como coeficiente, reduciéndose la expresión a:

$$\text{var}(b) = \sigma^2 (X'X)^{-1} \quad \text{-----} \quad (1.2.18)$$

La varianza de b_i puede ser obtenida, multiplicando el i -ésimo elemento de la diagonal principal de $(X'X)^{-1}$, por σ^2 .

La covarianza de a_i para cualquier par de estimadores b_i y b_j , se calcula multiplicando σ^2 por el (i,j) -ésimo elemento de $(X'X)^{-1}$.

Regresando al tema de la suma de residuales cuadrados, a hora deseamos sustituir (1.2.2) y (1.2.12) en (1.2.6):

$$\begin{aligned} e &= X\beta + a - X[(X'X)^{-1} X'(X\beta + a)] \\ &= a - X(X'X)^{-1} X'a \quad \text{-----} \quad (1.2.19) \end{aligned}$$

Factorizando "a" se tiene:

$$e = [I_n - X(X'X)^{-1} X']a$$

Si hacemos

$$M = I_n - X(X'X)^{-1} X' \quad \text{-----} \quad (1.2.20)$$

Entonces:

$$e = Ma \quad \text{-----} \quad (1.2.21)$$

De manera que:

$$e'e = a'M'Ma$$

y como M es idempotente y simétrica:

$$e'e = a'Ma \quad \text{-----} \quad (1.2.22)$$

Sacando esperanzas en ambos miembros:

$$E(e'e) = \sigma^2 \text{tr}(M)$$

donde $\text{tr}(M)$ denota la traza de la matriz M. Esta última expresión es válida por las dos primeras suposiciones (1.2.4).

Además, por (1.2.20), sabremos que:

$$\begin{aligned} E(e'e) &= \sigma^2 \{ \text{tr}(I_n) - \text{tr}[X(X'X)^{-1}X'] \} \\ &= \sigma^2 [n - \text{tr}(I_k)] \\ &= \sigma^2 (n - k) \quad \text{-----} \quad (1.2.23) \end{aligned}$$

Finalmente concluimos que un estimador insesgado para la varianza de las perturbaciones es:

$$s^2 = \frac{e'e}{n-k} \quad \text{-----} \quad (1.2.24)$$

El cálculo de S^2 se facilita sustituyendo (1.2.10) en --
(1.2.8):

$$e'e = y'y - b'S'y \quad \text{-----} \quad (1.2.25)$$

Coefficiente de Correlación.

El grado de asociación entre dos variables de un modelo de regresión, puede medirse a través de un valor relativo.

La variación total de cada uno de los puntos Y_i , con respecto a su media aritmética \bar{Y} , está compuesta por: la variación explicada por la regresión, más la variación no explicada, es decir, en forma respectiva:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_r - \bar{Y})^2 + \sum_{i=1}^n (Y_i - Y_r)^2 \quad (1.2.26)$$

donde Y_r representa la línea de regresión (1.2.15).

De acuerdo con esto, el coeficiente de correlación, es un valor relativo que se obtiene mediante la razón de la variación explicada por la regresión, entre la variación total;

$$r^2 = \frac{\sum_{i=1}^n (Y_r - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad \text{-----} \quad (1.2.27)$$

Note que cuando los puntos Y_i caen sobre la línea de regresión, es decir, $\Sigma(Y_r - \bar{Y})^2 = \Sigma(Y_i - \bar{Y})^2$, el valor de $r^2 = 1$, lo cual se interpreta como una correlación perfecta. Por otra parte, cuando los puntos Y_i están muy alejados de la línea de regresión Y_r , entonces $\Sigma(Y_i - \bar{Y})^2$ se incrementa fuertemente y $\Sigma(Y_r - \bar{Y})^2$ ahora es muy pequeño, por lo que, el valor de r^2 tenderá a cero, lo cual significa que no hay correlación. En otras palabras, conforme r^2 se aproxima a uno, la línea de regresión ajusta cada vez mejor, a los puntos Y_i , por lo que la separación entre los puntos Y_i y la línea de regresión Y_r irá disminuyendo.

Empleando la letra "d" para denotar las desviaciones con respecto a la media aritmética, se tiene:

$$\sum_{i=1}^n d_i^2 = \sum_{i=1}^n Y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n Y_i \right)^2$$

que usando la notación matricial se expresa:

$$\sum_{i=1}^n d_i^2 = y'y - \frac{1}{n} (\Sigma Y_i)^2 \quad \text{-----} \quad (1.2.28)$$

lo cual representa la variación total en (1.2.26) y (1.2.27).

Para denotar la variación explicada por la regresión en forma matricial, nos apoyaremos en (1.2.26) y (1.2.8):

$$\sum_{i=1}^n d_i^2 - \sum_{i=1}^n e_i^2 = y'y - e'e - \frac{1}{n} \left(\sum_{i=1}^n Y_i \right)^2$$

Utilizando la relación (1.2.25) en el segundo miembro de la expresión anterior, obtenemos para la variación explicada por la regresión:

$$b'X'y - \frac{1}{n} \left(\sum_{i=1}^n Y_i \right)^2 \quad \text{-----} \quad (1.2.29)$$

Así pues, el coeficiente de correlación múltiple, como una medida de la bondad del ajuste, queda definido mediante (1.2.28) y (1.2.29) como:

$$R_{12 \dots k}^2 = \frac{b'X'y - (1/n) (\sum Y_i)^2}{y'y - (1/n) (\sum Y_i)^2} \quad \text{-----} \quad (1.2.30)$$

Aunque X permanece fija, dicho estimador resumido, lo podemos usar para analizar las tablas de varianza.

Otro uso que tiene es el de servir para el cálculo de un R^2 ajustado, especialmente cuando estamos comparando la "capacidad descriptiva" para conjuntos diferentes de variables explicativas.

El R^2 ajustado, está definido por:

$$R_{12 \dots k}^2 = 1 - \frac{\sum_{i=1}^n e_i^2 / (n-k)}{\sum_{i=1}^n y_i^2 / (n-1)} \quad \text{-----} \quad (1.2.31)$$

Siendo la relación entre ambos coeficientes la siguiente:

$$\bar{R}^2_{12,\dots,k} = \frac{n-1}{n-k} (1 - R^2_{12,\dots,k}) \quad (1.2.32)$$

El coeficiente no ajustado nunca decrecerá cuando se agregue una variable explicativa al modelo original, pero es posible que el coeficiente ajustado disminuya, en caso que la variable adicional, produzca una reducción pequeña en $(1-R^2)$ para compensar el incremento en $(n-1)/(n-k)$.

Matriz de Correlación.

Si calculamos todas las correlaciones simples (de orden cero) que hay entre las variables X_1, X_2, \dots, X_k y las colocamos en forma de matriz, obtendremos la matriz de correlación:

$$R = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1k} \\ r_{21} & r_{22} & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{pmatrix} \quad (1.2.33)$$

Donde r_{ij} es la correlación entre Y y X_j para $j = \overline{2,k}$ - mientras que $r_{ii} = 1$ para $i = \overline{1,k}$. Los coeficientes de regresión de mínimos cuadrados, pueden ser expresados como cofactores de la matriz de correlación (ver Johnston (1972] - cap. 5).

1.2.2.- MODELOS NO LINEALES.

Cuando el ajuste lineal resulta ser inadecuado para el problema con que estamos trabajando, entonces debemos recurrir a algún tipo de modelo no lineal, que logre describir mejor a las observaciones.

Entre los modelos no lineales más comunmente usados, podemos citar a los siguientes:

1. Exponencial.
2. Logarítmico.
3. Polinomial.

Las dos primeras opciones, se emplean cuando las observaciones presentan una curva evidentemente definida, en tanto que la tercera opción, se utiliza cuando no es posible elegir específicamente algún modelo matemático. En ese caso, un procedimiento razonable es examinar sistemáticamente polinomios con grados crecientes, hasta encontrar uno que haga un buen ajuste.

Supongamos:

Una relación no lineal general, entre la variable dependiente Y y la variable explicativa X , con su respectivo término de perturbación " a ". Si contamos con una muestra de n observaciones para Y y X , podemos escribir:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_n X_i^n + a_i \quad (1.2.34)$$

Así, en el caso de tener una muestra con n observaciones,

nosotros estimaremos las " β " de acuerdo con el principio de mínimos cuadrados, de manera similar a como se hizo de (1.2.6) a (1.2.11). Es decir, minimizaremos:

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \dots - \beta_k X_i^k)^2 \quad (1.2.35)$$

derivando esta última expresión con respecto a cada " β ", -- las $k-1$ ecuaciones normales representadas por:

$$\sum YX^j - \beta_0 \sum X^j - \beta_1 \sum X^{j+1} - \dots - \beta_k \sum X^{j+k} = 0 \quad (1.2.36)$$

con $j = \overline{0, k}$

La solución de tal sistema de ecuaciones, nos conducirá al valor estimado de los parámetros " β ", aunque no siempre es fácil su resolución.

Modelos aún más generales y mayor información, la podemos encontrar en Draper (1966) capítulos 5 y 10.

1.3.- DESCOMPOSICION,

Quando se tiene una serie de tiempo que representa algún fenómeno como un todo, dicha serie es el resultado de -diversas fuerzas (físicas, económicas, sociales, etc.), las cuales están interrelacionadas. Por tanto, una serie de -- tiempo es la resultante de cuatro componentes, a saber:

- a) Tendencia
- b) Estacionalidad
- c) Cíclica
- d) Irregular

Si logramos descomponer una serie de tiempo, en cada una de sus componentes (especialmente las dos primeras), esto -

nos proporcionaría una mayor información acerca de las fuerzas que originan las variaciones de la serie.

Al descomponer una serie, lo que estamos haciendo es imponer para cada componente, un cierto modelo. Como se mostrará posteriormente, este hecho, no debe dejar de inquietarnos.

En seguida describiremos cada una de las componentes que constituyen a una serie.

a) Tendencia.

Es una variación que afecta a una serie, originada por una fuerza que es persistente, uniforme y no oscilatoria, - es decir, que esa fuerza provoca cambios moderados, pero -- constantes en nuestra serie. Con ésta, podremos señalar la dirección que sigue la serie, en un periodo de tiempo largo.

Esta componente puede ser ascendente o descendente y representarse gráficamente mediante una recta, o una curva -- suave. Para su estimación, podríamos recurrir al método de regresión o al de promedio móvil.

La estimación de la tendencia, está muy lejos, de ser un proceso puramente mecánico, puesto que al elegir la recta o curva de regresión, el promedio móvil, la naturaleza de los pesos y el orden del polinomio del cual se obtuvieron los - pesos, juega un papel muy importante el criterio personal. En la sección de promedio móvil, aclararemos la idea de pe-

so y las técnicas para su cálculo.

Ejemplo.

El Producto Nacional Bruto, se incrementa debido al crecimiento de la población y al aumento de la acumulación de capital. En este caso, la serie del P.N.B. y las fuerzas que la afectan (tendencia) son población y la acumulación de capital.

El método de mínimos cuadrados (sec. 1.2), nos da los mejores estimadores lineales insesgados de los parámetros de la tendencia, cuando las desviaciones o residuales, son variables aleatorias normales e independientes, con varianza constante; pero al usar series de tiempo, estos supuestos, raras veces se cumplen. Si la serie de tiempo, a la cual ajustaremos una línea (o curva), para representar su tendencia, posee un componente cíclico, entonces las desviaciones de la serie, con respecto a la tendencia, contienen fluctuaciones cíclicas e irregulares. A causa de esto, no podemos afirmar que la línea (o curva) de tendencia estimada por mínimos cuadrados es mejor que cualquier otra; sin embargo, como una medida descriptiva, la línea (o curva) de tendencia por este método, cumple con su cometido, y tiene la ventaja de ser objetiva, por lo que es frecuentemente usada en el análisis de las series de tiempo.

b) Estacionalidad.

Es una variación que experimenta la serie en forma periódica, una o más veces al año.

Algunas de las causas principales que originan variaciones estacionales en una serie de tiempo son:

- Condiciones climatológicas
- Festividades (navidad)
- Temporada comercial, etc.

La medición de la componente estacional se realiza a través de los Índices Estacionales, los que se acostumbra a expresar en porcentajes. En la práctica, las unidades de tiempo más usuales de los términos de una serie, para medir la variación estacional, son la trimestral y la mensual.

Con el índice estacional, se pueden realizar comparaciones de un periodo de tiempo con otro.

Una serie de tiempo, puede o no, tener variación estacional, por consiguiente, antes de calcular un índice deberá examinarse detenidamente la variación de la serie, a fin de ahorrar tiempo en cálculos innecesarios.

Un método simple para detectar la variación estacional, es observar a cada uno de los términos de la serie, con respecto al promedio de los valores en ese año. Lo cual se lleva a cabo, por medio de una tabla, o una gráfica, donde

aparece el promedio anual, a ser comparado con cada uno de los términos (trimestrales o mensuales), para visualizar si persistentemente están por arriba o por debajo de él.

La variación estacional se presenta de dos maneras.

- Aquella cuyo patrón estacional permanece constante algunos años.
- Aquella cuyo patrón estacional cambia gradual o bruscamente.

La medición de la primera opción, es bastante simple, a diferencia de las variaciones estacionales cambiantes, que suponen mayores consideraciones teóricas y prácticas.

Un índice estacional, muestra el grado relativo de actividad por uno o varios años, que particularmente se ha realizado en cada mes (trimestre).

Un índice estacional puede ser:

- a) Específico.- Se aplica a cambios estacionales durante un año en particular.
- b) Típico.- Se obtiene promediando cierto número de variaciones estacionales específicas, es pues, una generalización de la variación estacional a lo largo de la serie.

Observando la variación estacional específica, puede determinarse si el patrón estacional de la serie es, o no, estable. En caso de ser cambiante, lo hace gradual, o repen-

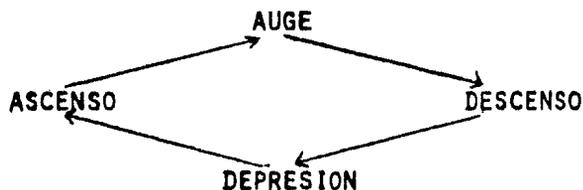
tinamente. La representatividad de un índice típico, exige que las variaciones estacionales específicas, sean estables.

Con variaciones estacionales específicas cambiantes, la obtención de un índice estacional típico, es muy dudoso, y de poca importancia práctica.

Las componentes de estacionalidad y tendencia, están mezcladas en una forma muy íntima, y si éstas son eliminadas - de la serie de tiempo, obtendremos una serie fluctuante - - (con oscilaciones suaves). Para realizar proyecciones a -- corto plazo, sólo se tomarán en cuenta las componentes de - tendencia y estacionalidad.

c) Cíclica.

Es una variación semiperiódica que afecta a la serie, pero en intervalos de tiempo largos. Esta variación presenta las siguientes fases sucesivas e iterativas:



Las variaciones cíclicas, se expresan en porcentajes que están sobre o por debajo del valor real (o estimado) de tendencia.

Esta componente cambia a lo largo del tiempo en longitud e intensidad. O sea, que la serie presentará en varios intervalos largos de tiempo, las mismas crestas y valles, pero con dimensiones diferentes.

Las fluctuaciones cíclicas, pueden ser medidas de datos anuales, o de unidades de tiempo menor. Una gran parte del análisis de las fluctuaciones cíclicas, emplea datos anuales, pero deja mucho que desear, pues las series anuales, - obscurecen dos de los aspectos más críticos de los ciclos:

- a) Localizar los puntos de giro
- b) Medir la amplitud de las fluctuaciones entre tales puntos.

Para proporcionar la información relevante de los ciclos, es mucho más conveniente, usar datos mensuales o trimestrales. En especial convienen los mensuales, pero tan pronto como son usados los datos que registran las variaciones durante el año, entra en acción la influencia de las componentes estacional e irregular; así que, para medir los ciclos en una serie mensual, primero deberá eliminarse la tendencia y la variación estacional, es decir, nos ocuparemos de estimar la componente cíclica e irregular, como una sola entidad, debido a que los métodos estadísticos presentan ineeficiencias en su estimación.

La naturaleza irregular de las variaciones cíclicas, de-

saffa cualquier intento de usar un ciclo medio, para representar sus efectos en una serie de tiempo. El mejor enfoque, mejor dicho, el menos insatisfactorio, es medir las fluctuaciones cíclicas en forma indirecta, eliminando primero, efectos de tendencia y variación estacional. Entonces, las fluctuaciones restantes, se les toma como variaciones cíclicas, aunque no podemos olvidar, que también reflejan influencia de la componente irregular. Este procedimiento indirecto, es imperfecto, no sólo porque un ciclo estimado así, no puede ser proyectado para el futuro, sino también porque tal estimación dependerá de los métodos empleados para calcular la tendencia y la estacionalidad, lo cual nos acarrea, un error sistemático, pero a fin de cuentas logramos el objetivo inicial, que era aislar la componente cíclica.

Aún cuando la componente cíclica no es fácil de controlar, resulta de gran interés para los economistas.

Para que las fluctuaciones de una serie, se les considere como ciclos, las expansiones y contracciones, deben durar un mínimo de doce meses, pues en otro caso, se les tomará como una variación irregular.

Cualquiera que sea el método de estimación de la componente cíclica, consume mucho tiempo y requiere bastante información (al menos 30 años). Situaciones que la convierten en impráctica, por lo que es omitida frecuentemente en

los modelos, aún en muchos paquetes estadísticos. A la tendencia, estacionalidad y ciclos, se les considera en cierto sentido, como funciones de tiempo estables, a diferencia de la componente irregular que no lo es.

Las fuerzas que originan las variaciones cíclicas, son - principalmente de tipo económicas. Por ejemplo: los altibajos de una economía están afectados estrechamente por los - niveles de inversión, producción, consumo y gastos del go--bierno. Por lo que en este ejemplo, las variaciones cíclicas, en cierta forma, reflejan los ciclos para la economía total de un país.

d) Irregular.

Esta componente la constituyen todas aquellas variaciones que afectan a la serie y que son distintas a las ya descritas. Por tanto, las fuerzas que las originan son de naturalaleza aleatoria, es decir, son perturbaciones inusitadas (ruido). Su control es muy difícil pues no hay métodos plenamente satisfactorios para su estimación. Por tales razones, esta componente es relativamente de poca importancia, además, las fuerzas aleatorias se producen por lo general - de fuentes desconocidas. Así que, no es posible ni necesario identificarlas, pero en cambio, es mejor tratarlas como parte del conjunto de factores que crean el rasgado curso - cíclico, que observamos en la realidad.

Para estimar la componente irregular, los términos de la

serie, deben estar expresados en unidades de tiempo menores a un año (mensuales o trimestrales), pues de lo contrario, en unidades anuales, se pierde su influencia. Ejemplos de esta componente, los caracterizan las huelgas, guerras, incendios, o cualquier fenómeno natural no recurrente.

Modelos.

La representación de una serie de tiempo mediante un modelo, la podemos realizar por:

- 1) Modelo aditivo.- cuando la serie es la resultante de la suma de sus componentes:

$$U = T + E + C + I \quad \text{-----} \quad (1.3.1)$$

- 2) Modelo multiplicativo.- cuando la serie es la resultante del producto de sus componentes:

$$U = T \cdot E \cdot C \cdot I \quad \text{-----} \quad (1.3.2)$$

El empleo de estos modelos se lleva a cabo de acuerdo -- con aquél que logre representar mejor al fenómeno en estudio, y no de acuerdo con las causas que determinan a la serie.

El modelo aditivo supone que las componentes de una serie de tiempo, son independientes entre sí. Es decir, que por alto o bajo que sea el nivel de la tendencia, no afecta

rá a las componentes restantes.

Por otra parte, el modelo multiplicativo supone que las fuerzas generadoras de las componentes están estrechamente relacionadas. Es decir, que aunque las componentes se deben a causas diferentes, tienen una relación entre sí.

En la práctica se suele utilizar al modelo multiplicativo para analizar a las series de tiempo pues en base a éste, se han desarrollado las técnicas clásicas de estimación de las componentes.

Debemos percatarnos que la descomposición de una serie, a través de cualquier procedimiento, tan sólo son estimaciones de cada una de las componentes que la forman.

Las fuerzas que afectan a las series económicas o sociales, son muy complejas, en consecuencia, resulta improbable que se desarrollen técnicas perfectas, para aislar las cuatro componentes.

Cuando los términos de una serie de tiempo son anuales, desaparecen los efectos de la variación estacional; casi lo mismo, sucede con la componente irregular, por lo que los modelos (1.3.1) y (1.3.2) se reducen a:

$$U = T + C \quad \text{-----} \quad (1.3.3)$$

$$U = T \cdot C \quad \text{-----} \quad (1.3.4)$$

De modo que las fluctuaciones cíclicas, serán las desviau

ciones alrededor de los valores de tendencia. Entonces, --
tendremos respectivamente:

$$U - T = C \quad \text{-----} \quad (1.3.5)$$

$$U / T = C \quad \text{-----} \quad (1.3.6)$$

que expresen los residuales cíclicos (1.3.5) y los relativos cíclicos (1.3.6).

Puede sostenerse que la tendencia y la componente cíclica, están sujetas por lo regular, a muchos factores en común. Tales como: Ingreso nacional, población y cambios en las preferencias de los consumidores.

Al determinar la tendencia de una serie, se persiguen - los objetivos siguientes:

- a) Extrapolación
- b) Ajuste estacional de una serie.

Ambos, se interfieren en sus pretensiones, ya que podemos hallar una ecuación de tendencia que ajuste satisfactoriamente a los ciclos, pero a su vez sea inadecuada para -- pronosticar, o viceversa. Además, cuando difieren los años inicial y final escogidos, para obtener la tendencia, llegamos a una ecuación diferente. Un efecto semejante se produce cuando cambiamos el periodo total comprendido. En consecuencia, los relativos cíclicos, son afectados al elegir la función de tendencia y al escoger el periodo al cual la ajustamos.

El ajuste estacional también tiene cierta arbitrariedad, pues los patrones estacionales de muchas series, se hallan sujetos a modificaciones sin advertencia, y aún en el caso de que permanezcan relativamente constantes, los valores de los índices estacionales típicos, siempre difieren de acuerdo con los distintos métodos de cálculo. Esto significa -- que los valores de los datos ajustados estacionalmente son afectados por:

- a) El uso de índices estacionales típicos o cambiantes
- b) El método de cálculo.

El $C \cdot I$ resultante es influido por el supuesto bajo - el cual fue descompuesta la serie (aditivo o multiplicati--vo). Al escoger uno, somos incapaces de determinar, si el otro se acerca más a la realidad de la serie.

De las consideraciones anteriores, podemos decir que la estimación de $C \cdot I$ incluye no sólo las fluctuaciones cíclicas e irregulares, sino también, los errores causados en la estimación de la tendencia, patrón estacional y el su- - puesto bajo el que se descompuso la serie original.

Puesto que estos errores nunca pueden ser determinados - con exactitud, siempre existe especulación e incertidumbre en la interpretación y uso, de los resultados estadísticos llamados ciclos.

La tendencia, estacionalidad y ciclos, son reales y significativos para muchas de las series económicas y sociales. A pesar de que sus mediciones sean tan sólo aproximaciones, aún así, pueden ser de gran utilidad, en la investigación económico-social y en la administración de negocios a gran escala, ya que nos permiten pronosticar las series de tiempo, y con ello se nos facilitará la planeación.

1.4.- PROMEDIO MOVIL.

Entre los métodos más usuales que hay para ajustar curvas suaves a series de tiempo, podemos señalar a :

- Promedio móvil
- Exponencial

En particular, sólo nos ocuparemos de discutir el primero, por ser el de nuestro interés, pues son muy diversos -- los tipos que existen de un promedio móvil. Si deseamos saber sobre el segundo método, puede consultarse el Thomopoulos (1980) Cap. 10.

Aquí estudiaremos algunos de los procedimientos más im--

portantes de un promedio móvil, puesto que, en la práctica, a veces conviene aplicar varios de ellos, con el fin de seleccionar uno cuyos resultados se apeguen al objeto del análisis.

En esencia, todos los procedimientos de promedio móvil, tienen la misma idea, sólo cambian algunos aspectos técnicos, otorgándoles así, características específicas a cada uno; motivo por el cual será utilizado para una clase específica de problema.

Algunos autores, suelen llamar a un promedio móvil como un "filtro", pues elimina de la serie, los elementos periódicos de alta frecuencia, y deja casi sin tocar, los de frecuencia baja. Además consideran a la serie no ajustada, como la entrada a un filtro lineal, y a la serie ya ajustada como la salida. Así que, debemos interpretar la representación de una serie, mediante un filtro lineal, tan sólo como una aproximación.

1.4.1.- PROMEDIO MOVIL DE PRIMER ORDEN.

Este es el tipo más sencillo de un promedio móvil, motivo por el que también se le llama Promedio Móvil Simple. En este procedimiento, la tendencia es descrita por la fluctuación de los términos de una serie nueva. Esto se logra aplicando un promedio móvil simple, para obtener una serie de medias aritméticas, con los términos de la serie origi--

nal. A esta última, se le debe agrupar sistemáticamente, - para asignarle a cada grupo, un valor representativo " \bar{U} ", - que es calculado con el promedio de los términos que integran al grupo.

Suponga que tenemos una serie de tiempo con n términos - que denotaremos por:

$$U_1, U_2, \dots, U_n \quad \text{-----} \quad (1.4.1)$$

Tomando de ella los primeros N términos ($N < n$), obtenga su media aritmética; enseguida, repita el proceso, tomando ahora el grupo siguiente que comienza en el segundo término, y finaliza en el $(N + 1)$ -ésimo término. Este algoritmo debe repetirse sucesivamente, hasta agotar el último grupo -- que sea posible formar con la serie original. Al número de términos que constituye cada grupo, le llamaremos extensión del promedio. Todo esto, matemáticamente se expresa:

Suponga una serie como en (1.4.1), a la que le aplicaremos un promedio móvil simple de extensión N . Aquí, nos interesan en particular que $N = 2m + 1$; es decir, que cada grupo contenga un número impar de términos.

Por consiguiente, para calcular la serie nueva, empleemos la fórmula:

$$\bar{U}_t = \frac{\sum_{j=0}^{n+1} U_{t+j}}{N} \quad (1.4.2)$$

donde: $N = 2m + 1$

$t = m + 1, m + 2, \dots, n - m$

$j = 0, 1, 2, \dots, n - 2m - 1$

Advierta que j es una variable auxiliar que sirve para cambiar de un grupo a otro variando simultáneamente con t . Así para $t = m + 1$ tenemos $j = 0$, etc.; también se debe notar, que la serie nueva carece de los primeros y últimos m términos. La razón para preferir una extensión impar en el promedio, es que de esta manera el primer \bar{U}_t extraído, quedará precisamente a la mitad del primer grupo, coincidiendo con el término U_{m+1} de la serie original; de modo que es posible comparar a las dos series, en sus términos respectivos.

En otro caso, si tomáramos una extensión par ($2m$), entonces el primer \bar{U}_t quedaría en medio de U_m y U_{m+1} del primer grupo, razón por la que no podríamos comparar ambas series directamente y esto ocasionaría algunos problemas de inexactitud.

Conforme la extensión del promedio sea mayor, así también será mayor la suavidad adquirida por la serie nueva, pero en consecuencia ésta, habrá perdido más valores de ten

dencia en sus extremos.

Al emplear un promedio móvil de extensión corta, se logra una mejor estimación de la tendencia y un mejor ajuste. En el caso contrario, correremos el riesgo de obtener una tendencia adulterada, o sea con la influencia de otras componentes.

Las variaciones de la serie, causadas por la componente cíclica, pueden ser parcialmente eliminadas, haciendo coincidir la extensión del promedio móvil, con la longitud del ciclo que presenta la serie original. En tal forma, que -- las variaciones cíclicas sean moldeadas uniformemente y con ello se obtenga una buena estimación de la tendencia.

Se pueden investigar los efectos que provoca la inclusión de variaciones estacionales o cíclicas en una serie de tiempo, recurriendo a una simulación.

1.4.2.- PROMEDIO MOVIL DE ORDEN SUPERIOR.

Promedio Móvil Doble.

Esta técnica es usada cuando sucede que, después de haber aplicado un promedio móvil simple a una serie, no hemos conseguido la suavidad deseada.

Es obvio que tal técnica, es una extensión al procedimiento anterior. Así que en caso de tener una serie como en (1.4.1) a la que ya hemos aplicado un promedio móvil simple (1.4.2), entonces tendremos una serie nueva que es:

$$\bar{U}_{m+1}, \bar{U}_{m+2}, \dots, \bar{U}_{n-m} \quad \text{-----} \quad (1.4.3)$$

A esta última serie le aplicaremos nuevamente un promedio móvil con extensión $N' = 2m' + 1$, al que denotaremos con $\bar{U}_t^{(2)}$

El promedio móvil doble, lo calcularemos con la fórmula:

$$\bar{U}_t^{(2)} = \frac{\sum_{i=m'+j+1}^{N'+j+1} \bar{U}_i}{N'} \quad \text{-----} \quad (1.4.4)$$

donde: m es del polinomio simple

$$N' = 2m' + 1$$

$$t = m + m' + 1, m + m' + 2, \dots, n - m - m'$$

$$j = 0, 1, 2, \dots, n - 2(m + m')$$

Observe que esta última fórmula admite la posibilidad de usar una extensión del promedio diferente a la usada anteriormente.

Por otra parte, aunque se logró una mayor suavidad para nuestra serie, ahora se tiene la desventaja de haber perdido aún más términos en los extremos de la serie nueva.

Para resolver el problema de la pérdida de valores de --tendencia, en los extremos de la serie, plantearemos una solución posterior llamada "Valores en los extremos".

Si quisiéramos pronosticar en forma práctica, aplicando un promedio móvil, podemos consultar Thomopoulos (1980); auque debe aclararse que no es un buen método para hacerlo.

1.4.3.- PROMEDIO MOVIL POR MINIMOS CUADRADOS.

Esta técnica cuenta con mucha preferencia, debido a -- que casi cualquier serie, puede ser representada a través de un polinomio, alcanzando una gran exactitud. Motivo por el cual, hemos elegido esta técnica como la apropiada para compararla con otra, que expondremos en el capítulo siguiente.

El método requiere de una serie del tipo (1.4.1), para ajustar un polinomio de orden propuesto. Se toma un primer grupo de términos (digamos $2m + 1$) de la serie y le determinamos el valor de tendencia para U_{m+1} , que está precisamente a la mitad de este primer grupo. A continuación, ajustaremos el mismo polinomio al grupo de términos siguiente, que comienza en U_2 y termina con U_{2m+2} , de la serie original. Con este grupo, determinaremos nuevamente el valor de la tendencia para U_{m+2} . Así sucesivamente, hasta agotar el último grupo posible.

En realidad, no es necesario ajustar un polinomio para cada grupo, puesto que el procedimiento es equivalente al tomar combinaciones lineales de las observaciones, con pesos bien especificados.

En forma inductiva ilustraremos el procedimiento a partir de un ejemplo, para después hacer la generalización.

Ejemplo (3).

Se desea ajustar un polinomio cúbico, a una serie de - - tiempo con siete términos.

Sin que perdamos generalidad, las unidades de tiempo las tomaremos como:

$$t = -3, -2, -1, 0, 1, 2, 3 \quad \text{-----} \quad (1.4.5)$$

El polinomio que usaremos es:

$$U_t = C_0 + C_1 t + C_2 t^2 + C_3 t^3 \quad \text{-----} \quad (1.4.6)$$

donde la constante C_0 será determinada por el principio de Mínimos Cuadrados (ver sec. 1.2). Es decir, minimizaremos:

$$\sum_{t=-3}^3 (U_t - C_0 - C_1 t - C_2 t^2 - C_3 t^3)^2 \quad \text{-----} \quad (1.4.7)$$

lo cual ya se explicó antes. Así pues llegaremos a las cuatro ecuaciones normales:

$$\left. \begin{array}{l} \text{I.- } \sum U_t = 7 C_0 + 28 C_2 \\ \text{II.- } \sum t U_t = 28 C_1 + 196 C_3 \\ \text{III.- } \sum t^2 U_t = 28 C_0 + 196 C_2 \\ \text{IV.- } \sum t^3 U_t = 196 C_1 + 1588 C_3 \end{array} \right\} \quad \text{-----} \quad (1.4.8)$$

en donde las sumas de potencias impares de $t = -3$ hasta - -

$t = 3$, se han anulado.

Por el momento, sólo nos interesa C_0 , que es el valor de la serie para $t = 0$ (en el polinomio). El cálculo de ese valor, sólo requiere del empleo de las ecuaciones I y III:

$$\begin{aligned}
 C_0 &= \frac{1}{21} \left\{ 7 \sum_{t=-3}^3 U - \sum_{t=-3}^3 t^2 U_t \right\} \\
 &= \frac{1}{21} \left\{ 7(U_{-3} + U_{-2} + U_{-1} + U_0 + U_1 + U_2 + U_3) \right. \\
 &\quad \left. - (9U_{-3} + 4U_{-2} + U_{-1} + 0 + U_1 + 4U_2 + 9U_3) \right\} \\
 &= \frac{1}{21} \left\{ -2U_{-3} + 3U_{-2} + 6U_{-1} + 7U_0 + 6U_1 + 3U_2 \right. \\
 &\quad \left. - 2U_3 \right\} \text{-----} \quad (1.4.9)
 \end{aligned}$$

De modo que el valor de tendencia en cualquier punto de la serie original, es el promedio de un grupo de siete términos, donde tal valor de tendencia, está en el centro del grupo y para ser obtenido se utilizan los pesos:

$$\frac{1}{21} \{ -2, 3, 6, 7, \dots \} \text{-----} \quad (1.4.10)$$

Esta última expresión denota a un promedio móvil con pesos, en una forma abreviada, pues se aprovechó que los pesos presentan simetría central, además de omitir las "U".

Note que los pesos, son ponderaciones asignadas a los -- términos de un grupo, según la importancia relativa que pre senta el término para dicho grupo. A estos pesos, les afec ta directamente la extensión que tenga el promedio.

Para generalizar los resultados obtenidos, consideremos un polinomio de orden k que ajustaremos a una serie, toman-- do grupos con $2m + 1$ términos. Entonces debemos minimi-- zar la expresión:

$$\sum_{-m}^m (U_t - C_0 - C_1 t - C_2 t^2 - \dots - C_k t^k)^2 \quad \text{--- (1.4.11)}$$

de acuerdo a como lo hicimos en (1.2.9) y (1.2.10), para te ner $k + 1$ ecuaciones.

El resultado de C_0 depende de los valores numéricos da-- dos por la suma $\sum t^j$ y las funciones lineales de las "U" especificadas por $\sum t^j U$. El valor de la tendencia en el punto $t = k$, es entonces un promedio de los valores U_{k-m} al U_{k+m} .

En la tabla (1) se muestran los pesos calculados, para a justar una serie, mediante polinomios de segundo y tercer - orden. Con extensión del promedio de 5 a 21 términos.

Bajo el mismo procedimiento, es posible construir una ta bla para polinomios de cuarto y quinto grado, consulte - - Kendall (1976) pág. 31.

T A B L A (1)

E	POLINOMIOS CUADRADOS-CUBICOS
5	$\frac{1}{35} \mid -3, 12, 17, \dots \mid$
7	$\frac{1}{21} \mid -2, 3, 6, 7, \dots \mid$
9	$\frac{1}{23} \mid -21, 14, 39, 54, 59, \dots \mid$
11	$\frac{1}{429} \mid -36, 9, 44, 69, 84, 89, \dots \mid$
13	$\frac{1}{145} \mid -11, 0, 9, 16, 21, 24, 25, \dots \mid$
15	$\frac{1}{1105} \mid -78, -13, 42, 87, 122, 147, 162, 167, \dots \mid$
17	$\frac{1}{323} \mid -21, -6, 7, 18, 27, 34, 39, 42, 43, \dots \mid$
19	$\frac{1}{2261} \mid -136, -51, 24, 89, 144, 189, 224, 249, 264, 269, \dots \mid$
21	$\frac{1}{3059} \mid -171, -76, 9, 84, 149, 204, 249, 284, 309, 324, 329, \dots \mid$

En lo que concierne a la interpolación de los pesos negativos en un promedio móvil, diremos que no siempre es fácil su justificación, pero en general, indican que algunos términos de la serie son nocivos.

1.4.4.- PROPIEDADES.

1. La suma de los pesos en un promedio móvil, siempre es igual a la unidad.
2. Los pesos, presentan una simetría central respecto al término U_q .
3. Al ajustar hacia adelante o hacia atrás del tiempo, obtendremos indistintamente, los mismos valores de tendencia.
4. Las fórmulas para calcular los pesos con polinomios de orden $2k$, son las mismas para los polinomios de orden $2k + 1$.
5. La manera en que hemos conseguido hasta ahora los promedios, deja sin valor de tendencia a los primeros y últimos m términos de la serie original. Esta dificultad, será superada en la sección 1.4.5, "Valores en los extremos".
6. Aunque es posible llegar a fórmulas semejantes cuando se elige una extensión par en el promedio, como resultado tenemos valores de tendencia centrados entre dos términos de la serie original. Lo cual es inconveniente, como ya se ha señalado. Pero en la práctica, algunas veces esto es necesario, en cuyo caso se sugiere utilizar un Promedio Móvil Centrado, los que trataremos en la sec

ción 1.4.6.

Si los datos originales muestran una fluctuación periódica uniforme, un promedio móvil de extensión igual o múltiplo al periodo, eliminará gran parte de las variaciones periódicas. Por pequeña que sea la diferencia, entre la duración de la periodicidad de la serie original, y la extensión del promedio móvil, no puede borrar definitivamente -- las variaciones periódicas en la serie original. El proceso de aplicar un promedio móvil, sólo tiende a suavizar un poco, las fluctuaciones a corto plazo.

1.4.5.- VALORES EN LOS EXTREMOS.

Hasta el momento, las formas en que se han obtenido -- los diferentes tipos de promedio, todos ellos, originan una serie nueva que carece de valores en sus puntos extremos, - es decir, que los primeros y últimos m términos de la serie original, han perdido sus valores correspondientes de ten-- dencia debido al procedimiento empleado.

Existen razones que justifican el hecho de querer tener los valores en los extremos de la serie nueva, pues son de gran importancia y utilidad, como se observa en cierto tipo de pronóstico.

En esta sección, hablaremos de algunas técnicas para solucionar el problema, para cada uno de los procedimientos - de promedio.

1. Para un Promedio Móvil de Primer Orden.

Sean U_1, U_2, \dots, U_m los primeros m términos de la serie original, a los cuales no les fueron asignados valores de - tendencia. Para lograr ésto, emplearemos la tabla (2).

En ningún sentido, este procedimiento es el mejor, pero sí representa una alternativa, sencilla de calcular, para - el supuesto caso de necesitar dichos valores en los extre-- mos de la serie; además, mucho se debe tomar en cuenta, las características propias de la serie original, para saber si es o no aplicable el procedimiento.

T A B L A (2)

	EXTREMOS	TENDENCIA ASIGNADA
P R I M E R O S	U_1	$\bar{U}_1 = U_1$
	U_2	$\bar{U}_2 = (U_1 + U_2)/2$

	U_m	$\bar{U}_m = (U_1 + U_2 + \dots + U_m)/m$
U L T I M O S	U_n	$\bar{U}_n = U_n$
	U_{n-1}	$\bar{U}_{n-1} = (U_n + U_{n-1})/2$

	U_{n-m}	$\bar{U}_{n-m} = (U_n + U_{n-1} + \dots + U_{n-m})/m$

Otra alternativa sería simular términos de la serie original (hacia adelante y hacia atrás), de tal "suerte", que, al aplicar el promedio móvil, los términos simulados en la serie original, sean precisamente los que se pierden al obtener los valores de tendencia.

2. Para un Promedio Móvil de Segundo Orden.

En caso de interesarnos por tener el mismo número de términos de la serie original, después de haber aplicado un -- promedio móvil doble, primero debe calcularse la tendencia (1.4.2), junto con sus valores extremos (Tabla 2) de la se-

rie original (1.4.1); a continuación se usará (1.4.4) para encontrar las $\bar{U}_t^{(2)}$ y finalmente se vuelve a utilizar la tabla 2, para hallar los nuevos valores extremos de la serie.

No debe olvidarse que cada vez que empleamos la tabla 2, estamos perdiendo precisión, pero en cambio, hemos logrado nuestro objetivo inicial que era tener el mismo número de términos que en la serie original.

3. Para un Promedio Móvil por Mínimos Cuadrados.

El cálculo de los valores de tendencia, en los extremos de una serie para este caso, sólo requerirá de la adaptación del método de mínimos cuadrados para lograrlo.

En el ejemplo 3 (1.4.5), se ajustó un polinomio cúbico a una serie con siete términos; ahora se les considerará como los siete últimos términos de una serie dada.

Para encontrar los valores de dicho polinomio en $t = 1, 2, 3$, necesitaremos los valores C_1, C_2 y C_3 de la expresión (1.4.6), que no hemos usado.

Resolviendo directamente las ecuaciones (1.4.8) nos daría:

$$\begin{aligned}
 C_1 &= \frac{1}{1512} \left\{ 397 \sum_{-3}^3 t U_t - 49 \sum_{-3}^3 t^3 U_t \right\} \\
 C_2 &= \frac{1}{84} \left\{ \sum_{-3}^3 t^2 U_t - 4 \sum_{-3}^3 U_t \right\} \\
 C_3 &= \frac{1}{216} \left\{ \sum_{-3}^3 t^3 U_t - 7 \sum_{-3}^3 t^3 U_t \right\}
 \end{aligned} \quad (1.4.12)$$

Si expresamos estas ecuaciones como un promedio de los siete últimos términos:

$$\begin{aligned}
 U_t &= \frac{1}{21} [-2, 3, 6, 7, 6, 3, -2] \\
 &+ \frac{1}{252} [22, 67, -58, 58, 67, -22]t \\
 &+ \frac{1}{84} [-5, 0, -3, -4, -3, 0, 5]t^2 \\
 &+ \frac{1}{36} [-1, 1, 1, 0, -1, -1, 1]t^3 - (1.4.13)
 \end{aligned}$$

Para $t = 1, 2, 3$, la expresión anterior, toma los valores siguientes:

$$U_1 = \frac{1}{42} [1, -4, 2, 12, 19, 16, -4]$$

$$U_2 = \frac{1}{42} [4, -7, -4, 6, 16, 19, 8]$$

$$U_3 = \frac{1}{42} [-2, 4, 1, -4, -4, 8, 39]$$

Debemos fijarnos que la suma de los pesos, sigue siendo igual a la unidad, pero en cambio, ya no existe simetría -- central.

En el apéndice A daremos los pesos correspondientes para los valores extremos de una serie, al aplicar diferentes ex tensiones del promedio, esto, de acuerdo con la tabla (1).

En caso de necesitar tablas más completas que las del Apéndice A, puede consultarse: Cowden D. J. (1962).

Consecuencias.

El proceso de acercamiento al extremo de una serie, -- provoca que los pesos se alejen cada vez más, uno del otro. Esto se debe a que la suma de sus cuadrados, tiende a incre mentarse.

Como podríamos esperar, mientras más cerca estemos de -- los extremos, menor será la confiabilidad que tenga el punto de tendencia. Así que, una curva ajustada, tiende a des viar sus extremos.

Al calcular los valores de tendencia para los extremos - de una serie, no se pueden usar las mismas fórmulas para po línomios de orden par, que para los de orden impar.

1.4.6.- PROMEDIO MOVIL CENTRADO.

Ya hemos enfatizado la conveniencia de utilizar una ex tensión impar en un promedio móvil. Pero suele ocurrir, so bretodo en las series económicas, que necesitamos tomar una extensión par. Tal sería el caso si tuviéramos una serie mensual, y elegimos una extensión de doce, con el objeto de representar una tendencia anual.

Nosotros deseamos que los valores de tendencia, coinci-- dan con las fechas en que ocurrieron las observaciones. Una alternativa para lograrlo, sería sacando una media aritmética, entre los dos valores de tendencia más cercanos, a cada una de las fechas en que ocurrieron las observaciones.

Otra alternativa más usual, sería tomar en el promedio - móvil, una extensión impar, pero considerando a una misma - fecha, dos veces en cada grupo, la que tendrá la mitad del peso de las fechas restantes. Por tanto, si en un promedio móvil, elegimos una extensión trece, para una serie mensual, se emplearán los pesos siguientes:

$$\frac{1}{24} [1, 2, 2, \dots, 2, 2, 1] \quad \text{---} \quad (1.4.15)$$

Observe que los dos eneros considerados en la serie men sual, tienen la mitad del peso que los demás términos.

Kendall afirma que en la práctica, es recomendable, bajo

ciertas circunstancias, el empleo de promedios centrados antes de estimar a la tendencia, a manera de filtro. Esto supone pues, una mejor estimación de la tendencia. En el libro de Kendall (1976), se exponen estas ideas, y se desarrollan para información mensual, aunque son completamente generalizables.

Hay razones para creer que un promedio móvil con extensión de doce meses, eliminará casi todas las influencias -- de la estacionalidad y la mayor parte de la componente irregular de la serie, pero no lo puede hacer por completo. Esto se debe a que las variaciones estacionales son periodicidades, y sus amplitudes pueden cambiar un poco de un año -- con otro.

1.5.- AUTOCORRELACION.

Introduciremos ahora, un concepto básico para el análisis de una serie de tiempo, denominado la Autocorrelación, éste se aplicará al referirnos a una serie teórica, siendo denotada por ρ_k . Cuando se trate de una serie muestral cambiaremos el nombre a correlación serial.

La correlación serial es un coeficiente que mide la relación que existe entre dos o más términos de una misma serie, por tanto, entre U_t y U_{t+k} hay $k-1$ términos relacionados, - le llamaremos "la correlación serial de orden k " denotándola con r_k .

Para calcular la correlación serial o dicho de otra manera

ra, estimar la autocorrelación de cualquier serie de tiempo, primero se obtendrán las estimaciones para las autocovarianzas (C_k), que definiremos por:

$$C_{k-} = \frac{1}{n} \sum_{k=0}^{n-k} (U_t - \bar{U})(\bar{U}_{t+k} - \bar{U}) \quad (1.5.1)$$

$$\text{donde } \bar{U} = \frac{1}{n} \sum_{t=1}^n U_t$$

de tal forma que, la correlación serial se puede calcular con:

$$r_k = C_k / C_0 \quad \text{con } k = 1, 2, \dots \quad (1.5.2)$$

El Correlograma.

La obtención y ordenación de los coeficientes ρ_1, ρ_2, \dots ó sus coeficientes muestrales equivalentes, junto con sus respectivas gráficas, constituyen un Correlograma; donde k serán las abscisas y ρ_k las ordenadas, que deberán cumplir las condiciones

$$-1 < \rho_k < 1 \quad \text{con } k > 1 \quad \text{y } k = 1, 2, \dots$$

El correlograma, nos refleja la distribución de la estructura interna de la serie. Además nos puede proporcionar un diagnóstico muy útil para distinguir diferentes ti--

pos de series, o bien para elegir un modelo que represente adecuadamente a la serie.

La Autocorrelación en un Modelo Lineal General.

El modelo lineal general, fué tratado en la sección -- (1.2), ahora veremos cómo afectan las perturbaciones autocorrelacionadas a un modelo de esta naturaleza.

Recordemos que la segunda suposición fundamental de -- (1.2.4) es de gran relevancia para el modelo lineal general. En tal suposición, los elementos que están fuera de la diagonal principal, son ceros, es decir, que los términos de perturbación, tienen una covarianza cero.

Para un modelo que posee perturbaciones distribuidas normalmente, esto significa que dichas perturbaciones, son parejas e independientes, en tanto que para los términos de una serie de tiempo, esto mismo indica la independencia serial de las perturbaciones.

En el caso de que exista alguna correlación serial entre los valores de X , para el modelo (1.2.1), entonces, tendremos en consecuencia, una correlación serial en el elemento de perturbación compuesto. Este es el caso particular de las variables omitidas puesto que casi siempre por comodidad y simplicidad sólo incluiremos en el modelo de regresión, las variables de mayor repercusión, por lo que el elemento de perturbación reflejará la influencia de éstas.

La correlación serial de las variables omitidas, no necesariamente implica, un término de perturbación serialmente correlacionado. Un elemento de perturbación, también puede ser originado por errores de medición en variables explicativas del modelo de regresión.

En el modelo (1.2.1), las perturbaciones a_t , siguen un esquema autoregresivo de primer orden que es:

$$a_t = \rho a_{t-1} + \varepsilon_t \quad \text{-----} \quad (1.5.3)$$

donde $|\rho| < 1$ y ε_t satisfacen las condiciones siguientes:

$$\left. \begin{array}{l} 1. E(\varepsilon_t) = 0 \\ 2. E(\varepsilon_t \cdot \varepsilon_{t+k}) = \sigma_\varepsilon^2 \quad \text{para } k = 0 \\ E(\varepsilon_t \cdot \varepsilon_{t+k}) = 0 \quad \text{para } k \neq 0 \end{array} \right\} \text{----- para toda } t$$

Apoyándonos en (1.5.3) se tiene:

$$a_t = \rho(\rho a_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

repetiendo el proceso indefinidamente:

$$a_t = \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \dots$$

así que:

$$a_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \quad \text{-----} \quad (1.5.4)$$

Por consiguiente : $E(a_t) = 0$ ya que $E(\epsilon_t) = 0$ para toda t . Además por (1.5.4), se puede afirmar que:

$$E(a_t^2) = \left(\sum_{i=0}^{\infty} \rho^{2i} \right) E(\epsilon_{t-1}^2)$$

y como las " ϵ " son serialmente independientes:

$$E(a_t^2) = \left(\sum_{i=0}^{\infty} \rho^{2i} \right) \sigma_{\epsilon}^2$$

Observe que las rhoes, forman una serie geométrica, las cuales convergen para $|\rho| < 1$.

Usando notación de cálculo, veremos la expresión a la -- que converge:

$$S_n - \rho^2 S_n = \sum_{i=0}^n (\rho^{2i} - \rho^{2(i+2)})$$

$$(1 - \rho^2) S_n = 1 - \rho^{2n+2}$$

entonces:

$$S_n = \frac{1 - \rho^{2n+2}}{1 - \rho^2}$$

Como $|\rho| < 1$, se cumple que:

$$\begin{aligned} \lim_{n \rightarrow \infty} S_n &= \lim_{n \rightarrow \infty} \frac{1 - \rho^{2n+2}}{1 - \rho^2} \\ &= \frac{1}{1 - \rho^2} \end{aligned}$$

ya que $\{\rho^{2n+2}\}$ converge a cero.

Finalmente, regresando al problema que nos ocupaba:

$$\sigma_a^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \quad \text{para toda } t \quad \text{-----} \quad (1.5.5)$$

Similarmente, apoyándonos en (1.5.4) llegamos a:

$$E(a_t \cdot a_{t-k}) = E\left[\left(\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}\right) - \left(\sum_{i=k}^{\infty} \rho^{i-k} \varepsilon_{t-i}\right)\right]$$

en la expresión anterior, separaremos los $k-1$ términos iniciales del primer factor para factorizar a rho, de tal forma, que en ambos factores den sumas idénticas:

$$E(a_t \cdot a_{t-k}) = E\left[\left(\sum_{i=0}^{k-1} \rho^i \varepsilon_{t-i} + \rho^k \sum_{i=k}^{\infty} \rho^{i+k} \varepsilon_{t-i}\right) \left(\sum_{i=k}^{\infty} \rho^{i-k} \varepsilon_{t-i}\right)\right]$$

por las suposiciones de (1.5.3) y como rho es una constante:

$$\begin{aligned} E(a_t \cdot a_{t-k}) &= \rho^k E\left(\sum_{i=k}^{\infty} \rho^{i-k} \varepsilon_{t-i}\right)^2 \\ &= \rho^k + \sigma^2 a \end{aligned}$$

De modo que el modelo (1.2.1), no satisface la suposición de tener un elemento de perturbación serialmente independiente. Cabe mencionar que, el esquema autorregresivo empleado (1.5.3), es el más simple de todos, por lo que algunas veces se le llama Esquema de Markoff (vea Kendall - 1976 pág. 71). Pero aún los de mayor complicación, fallan al no satisfacer la suposición de independencia serial.

1.6.- ESTACIONALIDAD.

Cuando una serie de cierta actividad, expresada con -- términos mensuales durante n años, muestra persistentemente la más alta cantidad de actividad en diciembre, y la más baja en febrero, en cada uno de los n años, se dice que tal - serie presenta estacionalidad.

Las series de tiempo típicamente estacionales, son las - de los viajeros, pues están afectadas fuertemente por los - períodos de vacaciones.

Para analizar e interpretar la tendencia y variaciones - cíclicas en fenómenos económicos, es frecuentemente neces- rio, ajustar estacionalmente las series de tiempo, es decir,

eliminar de la serie, la componente estacional.

Los métodos espectrales, han ganado mucho terreno en la estimación de la variación estacional, puesto que lo hacen, sin tomar en cuenta a las demás componentes de la serie. Esta ventaja resulta inalcanzable para la mayoría de los métodos tradicionales de la estadística, que requieren primeramente de la estimación de la tendencia, para posteriormente estimar la variación estacional.

En seguida, daremos algunas de las razones que motivan la estimación de la componente estacional:

- a) Para comparar una variable en distintas etapas del año, como un fenómeno exclusivamente interanual.
- b) En caso de querer eliminar las variaciones estacionales de una serie, para poder estimar posteriormente las componentes restantes, sin que ésta les afecte.
- c) Si deseamos saber una cifra actual de la serie, primero debemos eliminar la variación estacional.

Cada uno de estos objetivos, tiene una naturaleza diferente, por lo que podemos afirmar que, un mismo método para la determinación de la componente estacional, no siempre será el apropiado para cualquier serie.

Así que, de acuerdo con las características que presente la serie, aplicaremos un método determinado, para convertirla en una serie ajustada estacionalmente (desestacionaliza-

da). Para lo cual primero tenemos que estimar la componente estacional y posteriormente removerla de los datos originales. Tal ajuste estacional posee dos aspectos sobresalientes:

- a) El ajuste histórico de los datos pasados, esto es, usando toda la información disponible.
- b) El ajuste actual de cada nueva observación. Desde el punto de vista práctico, la última observación es la de mayor importancia, puesto que porta el valor actual ajustado, que es de gran relevancia, para el establecimiento de políticas económicas, empresariales, etc.

Los pasos que se siguen para el ajuste estacional son:

- a) Estimación de la tendencia. Esto se hace usualmente a través de un filtro de promedi móvil, o bien por mínimos cuadrados.
- b) Estimación de la componente estacional. La metodología se fundamenta en las diferencias de las observaciones, menos los valores de la tendencia estimados ($U - X$); o bien, la razón de las observaciones, con respecto a los valores de tendencia (U / X), según si el modelo es aditivo o multiplicativo respectivamente y donde X es la estimación de la tendencia.

Para ampliar sobre estos modelos e ilustrar con ejemplos, consulte Chou (1977), cap. 20.

De modo que, una vez determinada la estacionalidad, en forma de un Índice Estacional, podemos generar una serie desestacionalizada, dividiendo la serie original, entre el

Indice Estacional. Si posteriormente le aplicamos a la serie un promedio móvil, obtendremos una segunda y mejor estimación de la tendencia.

1.6.1.- Modelos de Ajuste Estacional.

El método de ajuste estacional, basado en las desviaciones de las observaciones, con respecto a la tendencia, se relaciona implícitamente con el modelo aditivo:

$$U_t = \xi_t + \eta_t + \varepsilon_t \quad \text{-----} \quad (1.6.1)$$

mientras que, el de las razones de las observaciones, con respecto a la tendencia, se relaciona implícitamente con el modelo multiplicativo:

$$U_t = \xi_t \cdot \eta_t \cdot \varepsilon_t \quad \text{-----} \quad (1.6.2)$$

donde U_t son las observaciones, ξ_t es la tendencia, η_t es la componente estacional, ε_t es la componente irregular, y además, todas están en un tiempo t .

A (1.6.1) y (1.6.2) se les conoce como modelos puros.

En la práctica, suele ocurrir que sea suficiente con elegir entre una variación estacional de carácter aditivo, o bien una variación estacional con carácter multiplicativo; sin embargo, para algunas series, ninguno de los modelos pu

ros resultan satisfactorios, entonces un modelo alternativo sería, aquel que posea elementos aditivos y multiplicativos simultáneamente.

Un modelo mixto, podrá ser así:

$$U_t = \xi_t \cdot \eta_t + \varepsilon_t \quad \text{-----} \quad (1.6.3)$$

Modelos más sofisticados para trabajos muy finos, de ajuste estacional, los podremos encontrar en el artículo de J. Durbin y M. J. Murphy. (1965).

Los modelos (1.6.1) y (1.6.2), suponen que las variaciones ascendentes y descendentes de los ciclos en una serie, están claramente balanceadas, es decir, cada ciclo tiene la misma duración y amplitud. De manera que, si incluimos un número suficientemente grande de términos en el proceso de promediación (promedio móvil de cualquier tipo) de datos para cada unidad de tiempo, (trimestre o mes), las fluctuaciones cíclicas pueden ser canceladas por el proceso.

Cabe señalar aquí, un aspecto teórico de relevancia para nosotros, concerniente en el hecho de que si ε_t en (1.6.2), fuera una variable con función de frecuencia asignable, no podríamos permitir que tuviera signo negativo, puesto que -causaría que U_t también fuera negativo. En la realidad, la circunstancia de que U_t tenga signo negativo, sólo se conc

be en el caso muy improbable, de que un evento aleatorio, afecte fuertemente a la serie. Para resolver este inconveniente, el modelo (1.6.2), lo transformaremos a una forma lineal, aplicando logaritmos en ambos miembros:

$$\ln U_t = \ln \xi_t + \ln \eta_t + \ln \varepsilon_t \quad \text{-----} \quad (1.6.4)$$

El elemento residual $\ln \varepsilon_t$ en (1.6.4), tendrá una función de frecuencia diferente, a la que tenía ε_t , pero seguirá siendo una variable aleatoria de media cero. Si llamamos λ_t al elemento residual y aplicamos la función exponencial en ambos miembros de (1.6.4), obtenemos una relación comparable con (1.6.2) que es:

$$U_t = \xi_t \cdot \eta_t \cdot \exp(\lambda_t) \quad \text{-----} \quad (1.6.5)$$

a través de la cual, ya hemos excluido la posibilidad de que U_t tome un valor negativo a consecuencia del elemento residual. Este modelo lo utilizaremos posteriormente para hacer una simulación.

Como ya sabemos, tanto en el modelo aditivo, como en el multiplicativo, para estimar a η_t , primero tenemos que estimar a ξ_t . Si para ello empleamos un promedio móvil con pesos iguales tomados a lo largo de un año, no afectaríamos mucho a los residuales, puesto que casi, por definición, la suma de los efectos estacionales en un año es cero. Es decir, cualquier exceso que se presente en la suma de los e-

fectos estacionales, puede ser absorbido por ξ_t . Para aclarar más, hablaremos de una serie mensual, entonces tendremos que formar un promedio móvil centrado (ver sec. 1.4.6), además le impondremos al modelo la condición:

$$\sum_{j=1}^{12} r_j = 0 \quad \text{-----} \quad (1.6.6)$$

En lo que toca a la componente irregular, no necesariamente haremos las mismas consideraciones, pues en caso de querer describir a la serie en forma suave, podríamos preferir eliminar a la componente irregular; pero si estamos interesados en las oscilaciones a corto plazo, es mejor probablemente, extraer la tendencia y dejar la componente irregular, para que la serie no se vea afectada por nuestros métodos de eliminación, ya que, es poco lo que sabemos acerca del mecanismo generador de las oscilaciones, y podría suceder que la componente irregular, jugara un papel muy importante.

El efecto causado por el ajuste estacional de una serie, es que la serie ajustada, es de alguna forma más suave, que la serie aún no ajustada, es decir, que ocurre una atenuación en la magnitud de las variaciones que presentaba la serie original.

Cambios en la variación estacional.

El problema más agudo para el ajuste estacional, aparece cuando, la variación estacional cambia con el tiempo. Hasta ahora, hemos asumido que el promedio de la variación estacional es constante año tras año, con lo que se simplifica el cálculo de la estimación del patrón estacional. Tal patrón, tiene una regularidad mucho más clara que el de cualquier otra de las componentes de la serie. No obstante, el patrón estacional puede modificarse de una forma gradual o repentina, motivado por alteraciones en las políticas de los negocios, innovaciones tecnológicas, actividades gubernamentales, etc. Cuando el patrón estacional de una serie, cambia repentinamente, es mejor usar un índice estacional - específico para el tiempo particular en el cual se estudia la variación estacional. Si el patrón estacional se modifica gradualmente, hay una manera de describirlo, consistente en calcular el índice estacional de cambio, basado en una - tendencia en vez de un promedio de la estacionalidad específica, para cada unidad de tiempo.

Mayor información y ejemplos, se encuentran en el Shao - (1960) capítulo 19.

Existen otros dos métodos alternativos para el manejo de la componente estacional, aunque son un poco más complejos. Uno de ellos, está fundamentado en el Análisis Armónico, y el otro en Regresión. Un desarrollo del método armónico, -

debe ser tratado cuidadosamente, puesto que pertenece al terreno del Análisis Espectral, que está fuera de los propósitos de nuestro trabajo; el método de regresión, podemos encontrarlo en el Kendall (1976) pág 66. Este método no es de uso común entre los estadísticos, puesto que requiere de mucha información y la mayoría de las veces, no es éste el caso.

Los estadísticos experimentados, indican que casi siempre es recomendable que se observe detalladamente, la naturaleza de la serie de tiempo en estudio, así como el establecimiento claro de los propósitos que nos inducen al análisis. Además, señalan que es mejor deshacerse desde un principio, de los efectos conocidos, antes de efectuar cualquier estudio. Otra costumbre buena, es "limpiar" la serie, con algún tipo de filtro apropiado, antes de proceder al tratamiento. En forma similar, es preferible primeramente, estimar y eliminar los efectos estacionales, para evitar su presencia en la serie puesto que distorsionarían nuestras conclusiones. Esta medida se justifica por el hecho de que la estacionalidad se debe a causas plenamente identificadas. Cuando la serie con la que se está trabajando presenta un comportamiento estacional, con periodicidad conocida "p", es importante agrupar los datos, en forma de tabla con p columnas, lo que facilitará en muchos aspectos al análisis. En la práctica, lo más usual es construir tablas con cuatro o doce columnas, según sea la información trimestral o men---

sual respectivamente.

En la actualidad, ya se cuentan con paquetes de programación que se utilizan para el análisis de las series de tiempo en una forma más eficiente, pues con ellos podemos separar a cada una de las componentes, según el criterio del usuario. Esto significa que podemos manipular la serie, conforme a nuestras necesidades; además, se pueden obtener las gráficas respectivas de cada uno de los resultados. Tales paquetes, cuentan con un gran número de opciones, como pruebas de hipótesis, correlación, etc., que facilitan y reducen el tiempo de nuestro análisis considerablemente.

En el Instituto de Matemáticas Aplicadas U.N.A.M., se encuentran algunos de estos paquetes.

CAPITULO**D O S****TEORIA DE LA
APROXIMACION**

2.1.- TEORIA DE LA APROXIMACION.

La función Spline ha basado su desarrollo en la Teoría de la Aproximación, por esta razón se presentará un marco - de referencia breve sobre esta teoría.

La idea de la aproximación es muy antigua y aún en la actualidad no es posible explicar claramente en qué consiste, pues se desconocen características distintivas que puedan - ayudar a definirla.

2.1.1.- AREAS DE APLICACION.

Esta teoría trata de resolver distintos problemas, pero parece haber dos en las que centra casi todo su interés.

El primero, es cuando se tiene una función muy compleja y se desea sustituirla por otra que presenta las mismas características, pero que sea más simple en su manejo para facilitar los cálculos. Tal función por ejemplo, podría ser un polinomio que tenga valores aproximados a la función original.

El segundo problema se refiere a determinar una función de cierta clase, que se ajuste con la mayor precisión a una serie de datos en el tiempo y pueda usarse para representar a tal serie. El problema anterior corresponde, por ejemplo, al ajuste por mínimos cuadrados.

Otra área importante de la Teoría de la Aproximación, son los problemas de interpolación y cuadratura. Consulte Hildebrand (1974).

La estrecha relación que existe entre la Teoría de la Aproximación y la Teoría de la Interpolación, se debe a que la primera ha aplicado y modificado varias técnicas de interpolación para aproximar una curva a una función continua o a una serie de datos.

2.1.2.- DESARROLLO DE LA TEORIA.

En los últimos años, la Teoría de la Aproximación se ha desarrollado en dos áreas diferentes:

En la computación y en el análisis funcional.

La primera, ha dado un gran impulso a esta teoría, y sus aportaciones más importantes se pueden resumir de la siguiente manera:

- 1) Se ha desarrollado un extraordinario potencial numérico, a tal grado que es posible atacar problemas donde aún la teoría no es bien conocida, sin embargo, las soluciones prácticas son bastante aceptables. Tal es el caso del trabajo de Hans Strauss, en el que desarrolla la técnica Spline en el sentido de mínimos cuadrados (Karlsruhe).
- 2) Debido a las limitaciones físicas de la computadora, se ha creado la necesidad de aproximaciones.

En los últimos quince años, la Teoría de la Aproximación se ha escrito en base al análisis funcional, unificando de esta forma el lenguaje, esto permite tratar problemas difíciles debido al grado de abstracción que presentan sus distintos conceptos. Así, con la ayuda del análisis funcional, se trata de encontrar mejores procedimientos, estimaciones de error y más clases de funciones.

De este modo, los métodos de solución van desde las matemáticas puras hasta conceptos de matemáticas aplicadas, tales como, análisis numérico, programación matemática y lineal, simulación, etc.

En los últimos diez años, esta teoría ha presentado las siguientes características:

- 1) Ha tenido un gran avance en la investigación de los métodos de interpolación por curvas suaves.
- 2) Un creciente interés en la aproximación numérica de varias variables. Consulte Hayes (1970).
- 3) Se ha trabajado con buenos resultados en problemas de aproximación bajo condiciones en la frontera, específicamente los trabajos de Rudin con condiciones convexas. Consulte Rudin (1967), Jones & Karlovitz (1967) y - - - Duffin & Karlovitz (1967).

2.1.3.- PROCEDIMIENTO GENERAL DE LA APROXIMACION.

El procedimiento general que se ha seguido para dar principio a una aproximación es el siguiente:

Se selecciona un cierto número de funciones $\phi_k(x)$ cuyas combinaciones lineales den origen a los elementos de un conjunto S , se escoge una técnica que seleccione la mejor función de aproximación, entre todas las funciones de S , cuyas propiedades se identifiquen lo más cercanamente posible, con las características conocidas de la función verdadera $F(x)$. Así, por ejemplo, una técnica sería seleccionar un polinomio de grado bajo (los polinomios son las funciones que presentan mayor facilidad en su manejo), que ajuste a los datos, o bien, a una función continua.

Otra forma de alcanzar una buena aproximación, es hacer una partición del intervalo, donde está definida la función verdadera, construyendo un polinomio de grado bajo en cada subintervalo, tal es el caso de la función Spline.

La exactitud de la aproximación obtenida, depende de la cantidad de información disponible de la función verdadera $F(x)$.

Un hecho importante de esta teoría, es que existe una función que aproxima uniformemente a otra, siempre y cuando cumpla con el siguiente teorema:

Teorema de Aproximación de Weierstrass.- Sea f una función de valor real y continua en un intervalo cerrado $[a,b]$. Para toda $\epsilon > 0$ existe un polinomio p (que podría depender de ϵ) tal que:

$$|f(x) - p(x)| < \epsilon$$

para toda x en $[a,b]$. La demostración puede consultarse en el Apostol (1974), pág. 322.

F U N C I O N E S S P L I N E

LAS FUNCIONES SPLINE SE HAN USADO CON DOS PROPOSITOS:

- INTERPOLACION

- APROXIMACION

EN LAS SIGUIENTES SECCIONES SE DESARROLLARAN POR SEPARADO

2.2.- SPLINE DE INTERPOLACION.

2.2.1.- INTRODUCCION.

El problema de ajustar curvas suaves a una serie de tiempo, se puede resolver mediante interpolación, cuyo objetivo es hacer pasar una función a través de ciertos puntos, para determinar valores entre ellos.

Teóricamente, existe una función $F(x)$, que llamamos función verdadera, la cual pasa a través de los puntos conocidos (x_k, Y_k) con $a \leq k \leq b$, explicando perfectamente los fenómenos estudiados.

De esta manera se tratará de obtener una función $f(x)$ -- que aproxime lo mejor posible a la función verdadera - - -

$F(x)$, imponiendo restricciones a dicha función de interpolación.

Entre las distintas restricciones se mencionan las dos siguientes:

1) Restricción de interpolación. Si:

$$F(x_k) = f(x_k) \quad \text{Para } a \leq k \leq b$$

2) Restricción de interpolación y suavidad. Cuando se cumple que:

a) $F(x_k) = f(x_k)$ con $a \leq k \leq b$

b) $F'(a) = f'(a)$ y $F'(b) = f'(b)$

c) $f(x)$ es n veces continuamente diferenciable.

Una característica importante de la interpolación la da el siguiente teorema:

Si x_1, x_2, \dots, x_n son puntos distintos, y $F(x_1), F(x_2), F(x_3), \dots, F(x_n)$ son datos dados, entonces existe exactamente un polinomio para el cual $p(x_k) = F(x_k)$, con $k = \overline{1, n}$. La demostración se puede consultar en el libro Carl de Boor (1978), pág. 2.

Entre los métodos de interpolación más usados podemos citar los siguientes:

I. Basados en polinomios.

- 1.1) Fórmula prospectiva y retrospectiva de Newton.
- 1.2) Fórmula de Stirling.
- 1.3) Interpolación iterada.
- 1.4) Interpolación lineal.
- 1.5) Interpolación de Chebychev.

II. Basados en razones de polinomios.

- 2.1) Fracciones continuas.

III. Basados en funciones seno y coseno.

- 3.1) Interpolación de Fourier.
- 3.2) Interpolación de Chebychev.

IV. Basados en funciones spline.

- 4.1) Spline cúbicos.

Los métodos anteriores se pueden consultar en Hildebrand (1974).

1.2.2.- IDEA INTUITIVA DE LA FUNCION SPLINE Y SUS RASGOS HISTORICOS.

"Un compromiso entre una poligonal y un polinomio de interpolación de grado mayor es necesario: Uno combina grados bajos y por lo tanto polinomios de oscilación débil -- con una función que es continuamente diferenciable en el - intervalo cerrado $[a, b]$ ". (Späth (1974).

La idea anterior es el origen de una de las técnicas de interpolación y aproximación más recientes, dicha técnica recibe el nombre de Spline.

El uso de polinomios por partes, precede a los métodos de interpolación de Newton. La restricción para suavizar con fórmulas de interpolación por subintervalos, apareció en la literatura actuarial del siglo XIX, bajo el nombre de "interpolación osculatoria" (interpolación repetida en un mismo punto), consulte de Boor (1978); y parece haber sido Sprage (1880) quien la introdujo; un ejemplo de este tipo de interpolación lo da la fórmula de Hermite y de - - Newton. Desarrollos posteriores en métodos de interpola-- ción osculatoria, incluyendo la deducción de una fórmula - de suavizamiento, que interpola por funciones spline, los llevó a cabo Jenkins (1927). Sin embargo, el término "fun-- ción spline" fue introducido al campo de las matemáticas - por I. J. Schoenberg en 1946 en el artículo "Contributions to the problem of approximation of equidistant data by ana-- lytic funtions" (1946).

El spline, fue usado mucho tiempo antes como dispositivo mecánico (tiras flexibles de algún material elástico), para hacer trazos por los dibujantes y diseñadores.

2.2.3.- DEFINICION DE FUNCION SPLINE CUBICA Y CONCEPTOS BASICOS.

$f(x)$ es una función spline de tercer grado con uniones en los nodos: $x_1 < x_2 < \dots < x_n$ si y sólo si posee las propiedades siguientes:

- 1) $f_k(x)$ es un polinomio cúbico en cada uno de los subintervalos $[x_k, x_{k+1}]$ con $k = \overline{1, N-1}$
- 2) $f(x) \in C^2$.

$C^2[a, b]$ es la clase de funciones que son dos veces -- continuamente diferenciables en todo punto del intervalo cerrado $[a, b]$.

Uno de los conceptos básicos de la teoría de spline, -- consiste en la existencia de un intervalo cerrado $[a, b]$ sobre el cual se define la partición:

$$\Delta : \{ a = x_1 < x_2 < \dots < x_n = b \} \quad \text{--- (1.2.1)}$$

formando de esta manera $(N-1)$ subintervalos $[x_k, x_{k+1}]$ -- con $k = \overline{1, N-1}$. Cuando los subintervalos están uniformemente espaciados, el cálculo del spline se simplifica como se mostrará más adelante.

En cada uno de los subintervalos se determina un polinomio cúbico, el cual se une en el nodo x_k con $k = \overline{2, N-1}$, -- con el siguiente polinomio por la condición de continuidad en la primera derivada (o bien, en la segunda derivada).

Sean $f'(x_k^-) = f'(x_k^+)$ y $f''(x_k^-) = f''(x_k^+)$ las condiciones de continuidad.

Estas condiciones garantizan que la función spline, sus primeras y segundas derivadas sean continuas en todo punto del intervalo $[a, b]$.

La función $f_k(x)$ que se construye en cada subintervalo, es de la forma:

$$f_k(x) = a_k + b_k \phi_1 + c_k \phi_2 + d_k \phi_3$$

donde las " ϕ_i " con $i = 1, 2, 3$ son funciones conocidas y fáciles de evaluar, por ejemplo:

$$\phi_i = h_k = (x_{k+1} - x_k)^i \quad \text{con } k = \overline{1, N-1}$$

y las a_k, b_k, c_k, d_k son los parámetros a determinar bajo las siguientes restricciones de interpolación y suavidad:

- 1) $f(x_k) = y_k$ con $k = \overline{1, N}$
- 2) $f_k(x)$ es dos veces continuamente diferenciable.

2.2.4.- PROPIEDAD BASICA,

La propiedad más importante de la función spline es - la "propiedad mínima" dada por la desigualdad:

$$\int_0^b |f''(x)|^2 dx \leq \int_0^b |g''(x)|^2 dx \quad \text{--- (2.2.2)}$$

donde $f''(x)$ es la segunda derivada de la función spline y $g''(x)$ es cualquier función de interpolación.

Esta propiedad se conoce también como "curvatura mínima" ya que el integrando se aproxima a la curvatura definida - por $f''/(1 + f'^2)^{3/2}$ (Consulte Courant (1974) pag. 372), -- cuando $f'(x)$ es muy pequeña. Esto adquiere importancia en la función spline natural, es decir, $f'(a) = f'(b) = 0$ y $f''(a) = f''(b) = 0$, pues es la única función entre todas - las funciones de interpolación que al ir cambiando de di-- rección lo hace suavemente, presentando las curvaturas más pequeñas y minimizando el número de puntos de inflexión. Por la razón anterior, es la función más suave entre todas las funciones que tienen segundas derivadas continuas.

En el caso de splines generalizados, la propiedad anterior recibe el nombre de "norma mínima", pues en este caso deja de existir el concepto de curvatura.

Esta propiedad fue descubierta diez años después de que Schoenberg introdujo la función spline al campo de las matemáticas. Sin embargo, tal propiedad fue más o menos desarrollada en la Teoría de Vigas (Beam Theory), de aquí, -

la estrecha relación que existe entre ambas teorías.

El teorema de Holladay que demuestra la relación (2.2.2) ha servido también para demostrar la similitud que existe entre la teoría de splines y el trabajo sobre aproximación óptima de funciones lineales de Golomb y Weiberger, en donde las funciones $\bar{U}(x)$ aparecen como solución de un problema variacional. Tal semejanza fue descubierta por Secret y la publicó con el nombre "Best integration formulas and best errors bounds".

La demostración de la expresión (2.2.2) se da en el teorema de Holladay del apéndice B.

2.2.5.- CONSTRUCCION DE LA FUNCION SPLINE DE INTERPOLACION.

En la construcción del spline cúbico (Consulte Späth (1974) cap. 3) se selecciona un polinomio de la forma:

$$f_k(x) = a_k + b_k h_k + c_k h_k^2 + d_k h_k^3 \quad \text{-----} \quad (2.2.3)$$

donde $h_k = x - x_k$ con $k = \overline{1, N-1}$

ya que facilita los cálculos por su simplicidad, y donde a_k, b_k, c_k, d_k son parámetros a determinar, en total $4(N-1)$ para formar los $(N-1)$ polinomios cúbicos.

Considerando el subintervalo $[x_k, x_{k+1}]$, derivando dos veces $f(x)$ y evaluado en x_k y x_{k+1} obtenemos:

$$\begin{array}{l} f_k(x_k) = a_k \\ f_{k+1}(x_{k+1}) = a_k + b_k h_k + c_k h_k^2 + d_k h_k^3 \end{array} \quad \left| \quad \text{-----} \quad (2.2.4) \right.$$

$$\begin{array}{l} f'_k(x_k) = b_k \\ f'_{k+1}(x_{k+1}) = b_k + 2c_k h_k + 3d_k h_k^2 \end{array} \quad \left| \quad \text{-----} \quad (2.2.5) \right.$$

$$\begin{array}{l} f''_k(x_k) = 2c_k \\ f''_{k+1}(x_{k+1}) = 2c_k + 6d_k h_k \end{array} \quad \left| \quad \text{-----} \quad (2.2.6) \right.$$

donde $h_k = x_{k+1} - x_k$ con $k = \overline{1, N-1}$

Se calcularán los coeficientes caracterizados por la terna (x_k, y_k, f''_k) a partir de las ecuaciones (2.2.4) y (2.2.6);

$$a_k = f_k \quad \text{donde } f_k = f_k(x_k) \quad \text{-----} \quad (2.2.7)$$

$$b_k = (f_{k+1} - f_k - c_k h_k^2 - d_k h_k^3) / h_k \\ = hy_k / hx_k - (f''_{k+1} + 2f''_k) hx_k / 6 \quad \text{-----} \quad (2.2.8)$$

$$\text{donde } hy_k = f_{k+1} - f_k$$

$$c_k = f''_k / 2 \quad \text{-----} \quad (2.2.9)$$

$$d_k = (f'''_{k+1} - f'''_k) / 6h_k \quad \text{-----} \quad (2.2.10)$$

Existe una forma alternativa de obtener los mismos coeficientes usando las expresiones (2.2.4) y (2.2.5) como se hizo anteriormente.

Ahora, imponiendo la condición de continuidad en la primera derivada:

$$f'_{k-1}(x_k) = f'_k(x_k) \quad \text{-----} \quad (2.2.11)$$

nos lleva a:

$$3d_{k-1}h_{k-1}^2 + 2c_{k-1}h_{k-1} + b_{k-1} = b_k \quad \text{-----} \quad (2.2.12)$$

y sustituyendo los parámetros se tiene:

$$hx_{k-1}f''_{k-1} + 2(hx_{k-1} + hx_k)f''_k + hx_k f''_{k+1} = 6 \left(\frac{hy_k}{hx_k} - \frac{hy_{k-1}}{hx_{k-1}} \right) \\ \text{con } k = \overline{2, N-1} \quad \text{-----} \quad (2.2.13)$$

Esta ecuación da un sistema de (N-2) ecuaciones con 'N' incógnitas f''_k $k=\overline{2, N-1}$.

Para que el sistema sea consistente, es necesario conocer los valores de f''_1 y f''_N . Así, el sistema estará dado por:

$$\begin{pmatrix} 2(hx_1+hx_2) & hx_2 & & & & & \\ hx_2 & 2(hx_2+hx_3) & hx_3 & & & & 0 \\ & & & \dots & & & \\ & 0 & & & hx_{n-2} & 2(hx_{n-2}+hx_{n+1}) & \\ & & & & & & \end{pmatrix} \begin{pmatrix} f''_2 \\ f''_3 \\ \dots \\ f''_{n-1} \end{pmatrix} =$$

$$\begin{pmatrix} 6 \left\{ \frac{hy_2}{hx_2} - \frac{hy_1}{hx_1} \right\} - hx_1 f''_1 \\ 6 \left\{ \frac{hy_3}{hx_3} - \frac{hy_2}{hx_2} \right\} - hx_2 \\ \dots \\ 6 \left\{ \frac{hy_{n-1}}{hx_{n-1}} - \frac{hy_{n-2}}{hx_{n-2}} \right\} - hx_{n-1} f''_n \end{pmatrix} \quad \text{--- (2.2.14)}$$

Como se dijo anteriormente, cuando $f''(x_1) = f''(x) = 0$ se tiene un spline natural, el cual minimiza la integral:

$$\int_0^b |f''(x)|^2 dx$$

Hay otros métodos para obtener el sistema de $(N-2)$ ecuaciones:

- 1) A partir de la ecuación de interpolación de Hermite, consulte Hildebrand (1974), pág. 48.
- 2) Otra forma es cuando tenemos una expresión que combina una interpolación lineal y un término de corrección cúbico para dar suavidad a la función. Consulte Forsythe & Moler (1967) Cap. 4.

Es conveniente, desde el punto de vista computacional, usar el método de eliminación de Gauss o de Cholesky para resolver el sistema de ecuaciones. Una vez obtenidos los valores de los parámetros, se sustituyen para llegar al spline cúbico.

La matriz del sistema tiene las características siguientes:

- 1) Tridiagonal
- 2) Simétrica
- 3) Diagonalmente dominante, con elementos diagonales positivos.

Estas propiedades hacen que la matriz sea no singular y por lo tanto existe su inversa, lo que garantiza la existencía y unicidad del spline.

Además de hacer $f''(a) = f''(b) = 0$, hay otros tipos de condiciones en la frontera, consulte De Boor (1978) pág. 54, que aseguran la existencia y unicidad del spline, las cuales afectan ligeramente al sistema, de hecho, sólo al primer renglón y al último de la matriz, y además al vector --del lado derecho (Y). Sin embargo, no cumplen con la pro propiedad de curvatura mínima. Consulte Spath (1974) pág. 39, 40 y 42.

Las condiciones en la frontera afectan fuertemente a los intervalos cercanos a esas fronteras, y este hecho debe to

marse muy en cuenta.

Cuando por alguna razón se está trabajando con abscisas equidistantes, particularmente, si $hx_k = 1$, con $k = \overline{1, N-1}$, entonces las ecuaciones se transforman en:

$$f''_{k-1} + 4f''_k + f''_{k+1} = 6hy^2_{k-1} \quad \text{con } k = \overline{2, N-1} \quad (2.2.15)$$

Cuando se usan polinomios de grado par, pueden aparecer inestabilidades numéricas, destruyendo la dominancia diagonal, ya que no es posible considerar condiciones en la - - frontera simétricas, como en los polinomios de grado impar. Por esta razón no se ha trabajado con este tipo de splines.

2.2.6.- RAZONES PARA APROXIMAR CON FUNCIONES SPLINE.

La técnica spline se ha venido usando con dos propósitos:

1) Aproximación de funciones de variable real.- Cuando se tiene una función $F(x)$ que presenta oscilaciones fuertes, -- discontinuidades en algunos de sus puntos, o bien es bastante complicada para ser derivada, integrada, etc., entonces -- podría ser conveniente reemplazar esa función por un spline para facilitar los cálculos.

A pesar de la restricción de suavidad impuesta al spline, éste puede aproximarse bastante bien, a funciones que tienen un comportamiento muy irregular.

2) Ajuste de información empírica.- En este caso no se tiene una función continua $F(x)$ a lo largo del intervalo $[a,b]$, sino que se conocen un conjunto de valores (datos) de una -- función tabulada, algunos de los cuales pueden presentar errores aleatorios.

2.2.7.- APLICACIONES DE LA FUNCION SPLINE.

Las funciones spline polinomiales han logrado un lugar importante en la solución numérica de muchos problemas, como por ejemplo:

- 1) Aproximación de derivadas e integrales.
- 2) Soluciones aproximadas de ecuaciones integrales.
- 3) Soluciones aproximadas de ecuaciones diferenciales.
- 4) Aproximación simultánea de funciones y de sus derivadas - de orden superior.
- 5) Aproximación de áreas verdaderas de histogramas o distribuciones probabilísticas empíricas.

Todas estas aplicaciones se deben a las diversas propiedades algebraicas y analíticas que presentan las funciones - spline, pueden consultarse en: Ahlberg, Nilson & Walsh (1967) Späth (1974) y Carl de Boor (1978).

2.2.8.- OTROS TIPOS DE SPLINE.

Existen otros tipos de spline cúbicos simples (Consulte Späth (1974) Capítulo 4), cuyo uso depende del problema específico que se tenga, así:

1) El spline cúbico con restricciones de cota periódica se usa cuando se tiene:

$$f_{, (j)}(x_1) = f_{n-, (j)}(x_n) \quad \text{con } j=0,1,2 \quad (2.2.16)$$

representando una función suave y cerrada.

La expresión de la cual se obtiene el sistema de ecuaciones, se deduce de la condición de continuidad, que de hecho, sigue siendo la (2.2.13), aunque las condiciones (2.2.16) añaden un nuevo renglón a la matriz y también modifican el último renglón. De esta manera se tiene un sistema de ecuaciones con $N-1$ variables $f_1'', f_2'', \dots, f_{N-1}''$. La matriz de coeficientes es tridiagonal, cíclica y definida positiva.

2) El spline paramétrico se usa cuando las abscisas no siguen el orden acostumbrado, es decir:

$$x_1 < x_2 < \dots < x_n$$

en tal caso será necesario suponer que hay un conjunto T cuyos elementos tienen la propiedad:

$$t_1 < t_2 < \dots < t_n$$

de esta forma se restablecerá la condición perdida.

Los valores t_n , se obtienen haciendo $t_1 = 0$ y usando la fórmula recursiva:

$$t_k = t_{k-1} + d_{k+1}$$

$$\text{donde } d_{k+1} = \sqrt{(hx_{k-1})^2 + (hy_{k-1})^2}$$

Se ajustan dos spline cúbicos naturales uno a los puntos (t_k, x_k) y el otro a (t_k, y_k) con $k = \overline{2, N}$. Los spline así obtenidos $X(t)$ y $Y(t)$, son las funciones paramétricas requeridas que sirven como puntos $(X(t), Y(t))$ por los que pasará el spline $f(x)$ que nos interesa.

3) El spline de quinto grado no es muy usado en la práctica, pues requiere mayor número de operaciones y memoria de la computadora, para lograr tan sólo un poco de más precisión que el spline cúbico, y con la misma aplicación del spline de Hermite.

Este tipo de spline considera la partición:

$$\Delta x : \{ x_1 < x_2 < \dots < x_n \}$$

el algoritmo es similar al spline cúbico.

4) El spline de Hermite se usa para aproximar áreas verdaderas de histogramas, con gran precisión, en éste es necesario conocer, además de los puntos (x_k, y_k) con $k = \overline{1, N}$, las derivadas f'_k en cada punto x_k y usar polinomios de quinto grado cuyos parámetros dependerán de los valores de los puntos y de sus primeras y segundas derivadas.

La expresión que origina las $(N-2)$ ecuaciones del sistema, se obtiene de la ecuación de continuidad de la tercera derivada.

2.2.9.- SPLINE GENERALIZADO.

Debido a la generalización que se ha hecho a los resultados obtenidos por Schoenberg, se han desarrollado funciones spline de carácter muy diferente a los splines simples, es decir, $f(x) \in C^{2k} [a, b]$ que satisfacen la ecuación diferencial:

$$D^{2k+2}f = 0$$

cuando se trata de polinomios de grado $2k + 1$.

Algunos de estos splines generalizados reciben el nombre de L-splines y Splines cardinales, consultar Varga (1971); g-splines, Lg-splines, pLg-splines y λ -splines, consultar Karlsruhe (1975); M-splines, consultar Carl de Boor (1978); etc.

La mayoría de los splines anteriores aparecen como solución a ciertos problemas de ajuste de datos mediante funciones racionales, logarítmicas, exponenciales y trigonométricas.

Hasta ahora, no se tiene una teoría algebraica de splines generalizados bien desarrollada, la razón principal es que se ha dado mayor importancia a los splines polinomiales.

Otra dirección que han tomado los splines, además de su generalización, es la extensión a varias dimensiones, consultar Ahlberg, Nilson & Walsh (1967) y Späth (1974).

2.3.- SPLINE DE APROXIMACION.

En la presente sección, nos ocuparemos del spline cúbico de aproximación, el cual tiene como objetivo reemplazar la interpolación estricta por un cierto tipo de suavizamiento, cuando se conoce que los valores de las ordenadas presentan errores aleatorios. Si se tiene alguna idea que sugiera un posible modelo para la curva a suavizar, digamos $f(x) = a + b \exp(-cx)$, se recomienda que los parámetros se determinen con el procedimiento de mínimos cuadrados, de otro modo, la función spline es conveniente.

2.3.1.- DESCRIPCION DEL PROBLEMA.

Al igual que en la función spline de interpolación, --

partiremos de los valores dados (x_k, y_k) con $k = \overline{0, N}$ y de la condición:

$$a = x_0 < x_1 < \dots < x_n = b$$

para construir una función suave $f(x)$, que minimice a la -- funcional (en el sentido de máxima suavidad):

$$J[g] = \int_a^b |g''(x)|^2 dx \quad \text{-----} \quad (2.3.1)$$

entre todas las funciones $g(x)$ bajo la restricción:

$$K[g] = \sum_{k=0}^n \left[\frac{g(x_k) - y_k}{\Delta y_k} \right]^2 < S \quad \text{-----} \quad (2.3.2)$$

donde $g(x) \in C^2[x_0, x_n]$, $S \geq 0$ y $\Delta y_k > 0$ con $k = \overline{0, N}$, son -- los valores dados. Note que S es el coeficiente de suavi-- dad y Δy_k son los pesos, que se dan a cada dato y_k .

Si es posible estimar la desviación estándar del error -- de la ordenada y_k , entonces ésta puede servir como el va-- lor para Δy_k . En este caso S puede tomar cualquier valor dentro del intervalo de confianza correspondiente al lado -- izquierdo de (2.3.2):

$$N - \sqrt{2N} \leq S \leq N + \sqrt{2N}$$

Deducción del intervalo de confianza.

Si $y_k \sim N(g(x_k), (\Delta y_k)^2)$, entonces:

$$u_k = \frac{y_k - g(x_k)}{\Delta y_k} \sim N(0, 1)$$

y $u_k^2 \sim \chi^2_{(1)}$ de aquí que:

$$S \geq \sum_{k=0}^n u_k^2 \sim \chi^2(n+1, n+1, \sqrt{2(n+1)})$$

Sea $N = n + 1$, entonces el intervalo de confianza para S es:

$$N - \sqrt{2N} \leq S \leq N + \sqrt{2N} \quad \text{-----} \quad (2.3.3)$$

El coeficiente de confianza del intervalo anterior, para 100 grados de libertad es de 0.6731. Si S se distribuye bajo una normal, entonces el coeficiente de confianza del intervalo es de:

$$\begin{aligned} P(N - \sqrt{2N} \leq S \leq N + \sqrt{2N}) &= P(-1 \leq \frac{S-N}{\sqrt{2N}} \leq 1) \\ &= \phi(1) - \phi(-1) \\ &= \phi(1) - (1 - \phi(1)) \\ &= 0.6826 \end{aligned}$$

El lagrangiano asociado al problema de minimizar $J[g]$ bajo la restricción $K[g] \leq S$ está dado por:

$$L[g] = J[g] + \lambda(K[g] + z^2 - S) \quad \text{-----} \quad (2.3.4)$$

donde z es una variable de holgura que permite pasar de las

restricciones del tipo (2.3.2), a la del tipo:

$$\kappa[g] = s \quad \text{-----} \quad (2.3.5)$$

La minimización de (2.3.4) establece una clase de compromiso entre desear quedar cerca de los datos y_k , dado por la ecuación (2.3.2), y querer obtener una función suave, dado por (2.3.1).

Este problema de minimización se resuelve a través del cálculo de variaciones, consulte Funk (1962), y que aquí sólo presentamos los resultados.

$$\begin{array}{l}
 \text{I) } g''(x) \equiv 0 \quad \text{con } x_k < x < x_{k+1} \quad \text{para } k = \overline{0, N-1} \\
 \text{II) } \left\{ \begin{array}{l} -g''(x_0^+) = 2\lambda(g(x_0) - y_0) / \Delta y_0^2 \\ g''(x_n^-) = 2\lambda(g(x_n) - y_n) / \Delta y_n^2 \\ g''(x_k^-) = g''(x_k^+) = 2\lambda(g(x_k) - y_k) / \Delta y_k^2 \quad \text{con } k = \overline{1, N-1} \end{array} \right. \\
 \text{III) } \left\{ \begin{array}{l} g''(x_0^+) = g''(x_n^-) = 0 \\ g''(x_k^-) = g''(x_k^+) \quad \text{con } k = \overline{1, N-1} \end{array} \right. \\
 \text{IV) } g^{(j)}(x_k^-) = g^{(j)}(x_k^+) \quad \text{con } k = \overline{1, N-1} \quad \text{y } j = \overline{0, 1}
 \end{array} \quad (2.3.6)$$

Las relaciones anteriores (2.3.6) son las ecuaciones de Euler-Lagrange que determinan el mínimo de (2.3.1) bajo las restricciones ó condiciones subsidiarias (2.3.2).

Los resultados expuestos en (2.3.6) son precisamente los necesarios para construir la función spline natural (cúbica

ca por partes), y en lo sucesivo se denotará por $f(x)$:

$$f(x) = a_k + b_k(x-x_k) + c_k(x-x_k)^2 + d_k(x-x_k)^3 \quad \text{--- (2.3.7)}$$

$$x_k < x < x_{k+1}$$

2.3.2.- DEDUCCION DE LOS PARAMETROS DE LA FUNCION SPLINE.

Sea:

$$f''(x) = 2c_k + 6d_k(x - x_k) \quad \text{con } x_k \leq x < x_{k+1} \quad \text{y } k = \overline{0, N-1}$$

la segunda derivada de la función spline, por (III) de - -
(2.3.6), obtenemos:

$$f''(x_0^+) = f''(x_n^-) = 0$$

$$2c_0 + 6d_0(x_0^- - x_0) = 2c_n + 6d_n(x_n^+ - x_n) = 0$$

$$\text{con } x_0^- = x_0 \quad \text{y} \quad x_n^+ = x_n$$

$$2c_0 = 2c_n = 0$$

entonces:

$$c_0 = c_n = 0$$

también por (III) de (2.3.6)

$$f''_k(x_{k+1}^-) = f''_{k+1}(x_{k+1}^+)$$

$$2c_k + 6d_k(x_{k+1}^- - x_k) = 2c_{k+1} + 6d_{k+1}(x_{k+1}^+ - x_{k+1})$$

Sea:

$$x_{k+1}^+ = x_{k+1} \quad \text{y} \quad x_{k+1}^- = x_{k+1}$$

$$d_k = \Delta c_k / 3h_k \quad \text{donde } h_k = x_{k+1} - x_k \quad (2.3.8)$$

$$\Delta c_k = c_{k+1} - c_k$$

De (IV) de (2.3.6) y de la función spline cúbica se tiene que:

$$a_k + b_k (x_{k+1}^- - x_k) + c_k (x_{k+1}^- - x_k)^2 + d_k (x_{k+1}^- - x_k)^3 = a_{k+1} +$$

$$b_{k+1} (x_{k+1}^+ - x_{k+1}) + c_{k+1} (x_{k+1}^+ - x_{k+1})^2 + d_{k+1} (x_{k+1}^+ - x_{k+1})^3$$

aquí $x_{k+1}^+ = x_{k+1}$ y $x_{k+1}^- = x_{k+1}$ con $k = \overline{0, N-1}$.

$$b_k = \Delta a_k / h_k - c_k h_k - d_k h_k^2 \quad (2.3.9)$$

También de (IV) de (2.3.6) y para la primera derivada de $f(x)$ se tiene:

$$b_{k-1} + 2c_{k-1} (x_k^- - x_{k-1}) + 3d_{k-1} (x_k^- - x_{k-1})^2 =$$

$$b_k + 2c_k (x_k^+ - x_k) + 3d_k (x_k^+ - x_k)^2$$

con $x_k^- = x_k^+ = x_k$ y $k = \overline{1, N-1}$

y usando los resultados (2.3.8) y (2.3.9):

$$\frac{c_{k-1} h_{k-1}}{3} + \frac{2}{3} (h_{k-1} + h_k) c_k + \frac{h_k c_{k-1}}{3} =$$

$$\frac{a_{k+1}}{h_{k-1}} - \left(\frac{1}{h_{k-1}} + \frac{1}{h_k} \right) a_k + \frac{a_{k+1}}{h_k}$$

o bien, en notación de matrices:

$$Tc = Q^T a \quad \text{-----} \quad (2.3.10)$$

donde T es una matriz definida positiva, tridiagonal de orden N-1;

$$t_{kk} = 2(h_{k-1} + h_k)/3$$

$$t_{k,k+1} = t_{k+1,k} = h_k/3$$

Q es una matriz de dimensiones (N+1) por (N-1), tridiagonal:

$$q_{k-1,k} = 1/h_{k-1}$$

$$q_{kk} = -1/h_{k-1} - 1/h_k$$

$$q_{k+1,k} = 1/h_k$$

$$c = (c_1, \dots, c_{N-1})^T$$

$$a = (a_0, \dots, a_N)^T$$

Por último, de (II) de (2.3.6) y de la tercera derivada se tiene:

$$\frac{6(c_k - c_{k-1})}{3h_{k-1}} - \frac{6(c_{k+1} - c_k)}{3h_k} = 2\lambda \frac{(f(x_k) - y_k)}{\Delta y_k^2}$$

$$2 \left[\frac{c_k - 1}{h_{k-1}} - \left(\frac{1}{h_{k-1}} + \frac{1}{h_k} \right) + \frac{c_{k+1}}{h_k} \right] = 2\lambda \Delta^{-2} y_k (y_k - a_k)$$

pasando a la notación de matrices:

$$Qc = \lambda D^{-2} (y - a) \quad \text{-----} \quad (2.3.11)$$

$$\text{donde } y = (y_1, \dots, y_n)^T$$

$$D = \text{diag} (\Delta y_1, \dots, \Delta y_n)$$

λ es el parámetro lagrangiano y Q, c, a tal como se definieron antes.

Finalmente, multiplicando por la izquierda a la expresión (2.3.11) por $Q^T D^2$ separamos la variable c :

$$Qc = \lambda D^{-2} (y - a)$$

$$Q^T D^2 Qc = \lambda Q^T D^2 D^{-2} (y - a)$$

$$= \lambda Q^T (y - a)$$

$$= \lambda Q^T y - \lambda Q^T a$$

por (2.3.10), sabemos que $Tc = Q^T a$, entonces:

$$Q^T D^2 Qc + \lambda Tc = \lambda Q^T y$$

$$(Q^T D^2 Q + \lambda T)c = \lambda Q^T y \quad \text{-----} \quad (2.3.12)$$

Ahora, despejando a de (2.3.11), tenemos:

$$a = y - \lambda^{-1} D^2 Qc \quad \text{-----} \quad (2.3.13)$$

Conociendo el valor de λ se obtiene c de (2.3.12), a de (2.3.13), y después los vectores restantes d y b . La matriz (2.3.12) es una matriz de banda con cinco diagonales distintas de cero, y definida positiva, si $\lambda \geq 0$.

2.3.3.- DETERMINACION DEL PARAMETRO LAGRANGIANO.

La expresión (2.3.4) tiene que ser también minimizada, pero ahora con respecto a z y λ , dando las siguientes dos -- condiciones:

$$\begin{aligned} \frac{\partial L}{\partial z} &= \frac{\partial}{\partial z} \int_0^b f^{n2}(x_k) dx + \frac{\partial}{\partial z} \left\{ \lambda \left[\sum_{k=0}^n \frac{(f(x_k) - y_k)^2 + z^2 - S}{\Delta y_k} \right] \right\} \\ &= 0 + \lambda \left[\frac{\partial}{\partial z} \sum_{k=0}^n \frac{(f(x_k) - y_k)^2 + z^2 - S}{\Delta y_k} - \frac{\partial S}{\partial z} \right] \quad (2.3.14) \\ &= 2\lambda z \end{aligned}$$

Iguando a cero $\partial L / \partial z$ se tiene la primera condición:

$$\lambda z = 0 \quad \text{-----} \quad (2.3.15)$$

Ahora bien:

$$\begin{aligned} \frac{\partial L}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \int_0^b f^{n2}(x_k) dx + \frac{\partial}{\partial \lambda} \lambda \left[\sum_{k=0}^n \frac{(f(x_k) - y_k)^2 + z^2 - S}{\Delta y_k} \right] \\ &= 0 + \left[\sum_{k=0}^n \frac{(f(x_k) - y_k)^2 + z^2 - S}{\Delta y_k} \right] \frac{\partial \lambda}{\partial \lambda} \\ &= \sum_{k=0}^n \frac{(f(x_k) - y_k)^2 + z^2 - S}{\Delta y_k} \quad \text{-----} \quad (2.3.16) \end{aligned}$$

Iguando a cero el segundo miembro y despejando la suma se obtiene la segunda condición:

$$\sum_{k=0}^n \frac{(f(x_k) - y_k)^2}{\Delta y_k} = S - z^2 \quad \text{-----} \quad (2.3.17)$$

De la última condición, y con $f(x_k) = a_k$, se llega a:

$$\left(\frac{a_0 - y_0}{\Delta y_0}\right)^2 + \dots + \left(\frac{a_n - y_n}{\Delta y_n}\right)^2 = S - z^2$$

o en notación matricial:

$$\|D^{-1}(a - y)\|_2^2$$

De (2.3.13), tenemos:

$$\|D^{-1}(-\lambda^{-1}D^2Qc)\|_2^2 = \|-\lambda^{-1}DQc\|_2^2$$

Despejando a c de (2.3.12), tenemos:

$$c = (Q^T D^2 Q + \lambda T)^{-1} \lambda Q^T y \quad \text{-----} \quad (2.3.18)$$

y sustituyendo en la relación anterior:

$$\| -DQ(Q^T D^2 Q + \lambda T)^{-1} \lambda Q^T y \|_2^2$$

o bien:

$$\| DQ(Q^T D^2 Q + \lambda T)^{-1} \lambda Q^T y \|_2^2 = F^2(\lambda) \quad \text{-----} \quad (2.3.19)$$

De la condición $\lambda z = 0$, se tiene que $\lambda = 0$ o $z = 0$.

El primer caso sólo es posible si $F \leq \sqrt{S}$, como consecuencia se tiene que $c = 0$, por (2.3.12), y por un proceso de límite:

$$\begin{aligned} \lim_{\lambda \rightarrow 0} a &= \lim_{\lambda \rightarrow 0} (y - \lambda^{-1} D^2 Q c) \\ &= \lim_{\lambda \rightarrow 0} (y - \lambda^{-1} \lambda D^2 Q (Q^T D^2 Q + \lambda T)^{-1} Q^T y) \\ a &= y - D^2 Q (Q^T D^2 Q)^{-1} Q^T y \quad \text{————— (2.3.20)} \end{aligned}$$

reduciendo el spline cúbico a una recta.

Si $F > \sqrt{S}$, tenemos que $\lambda \neq 0$ y $z = 0$, por lo tanto, se mantiene la igualdad en (2.3.2) y λ debe determinarse de la ecuación:

$$F(\lambda) = \sqrt{S}$$

que tiene al menos una raíz negativa, sin embargo, la expresión (2.3.1) se hace mínima para la única raíz positiva. Este último resultado puede consultarse en el artículo de C. H. Woodford (1970), en donde, además, presenta el desarrollo del spline de suavizamiento para polinomios impares y el programa correspondiente en algol.

2.3.4.- DEMOSTRACION DE QUE $F(\lambda)$ ES CONVEXA Y DECRECIENTE.

$F(\lambda)$ es una función convexa, estrictamente decreciente si $\lambda \geq 0$, y $f(\lambda) \rightarrow 0$ conforme $\lambda \rightarrow \infty$.

Se toma como punto de partida a $F^2(\lambda)$ definida en - - - (2.3.19), se demostrará que $F'' > 0$ y que $F(\lambda_1) > F(\lambda_2)$ si $\lambda_1 < \lambda_2$.

Como T es una matriz simétrica definida positiva, entonces, por la descomposición de Cholesky se tiene que:

$$T = LL^T$$

donde L es una matriz triangular inferior, y no singular, entonces:

$$\begin{aligned} Q^T D^2 Q + \lambda T &= Q^T D^2 Q + \lambda LL^T \\ &= L(L^{-1} Q^T D^2 Q (L^{-1})^T + \lambda I) L^T \end{aligned}$$

Hagamos: $A = DQ(L^{-1})^T$

tomando transpuesta de lo anterior: $A^T = L^{-1} Q^T D$

y sustituyendo:

$$Q^T D^2 Q + \lambda T = L(A^T A + \lambda I) L^T$$

de aquí que:

$$F^2(\lambda) = \|DQ(L^{-1})^T (A^T A + \lambda I)^{-1} L^{-1} Q^T y\|_2^2$$

si $y = Dq$, se tiene:

$$F^2(\lambda) = \|A(A^T A + \lambda I)^{-1} A^T q\|_2^2$$

Ahora, por descomposición singular de una matriz:

$$A = V\Delta U^T$$

tomando la transpuesta de lo anterior:

$$A^T = U\Delta^T V^T$$

donde V y U son matrices ortogonales y $\Delta = \text{diag}(\sigma_1, \dots, \sigma_k)$,
y donde las σ son valores propios de A . Sustituyendo:

$$\begin{aligned} F^2(\lambda) &= \|V\Delta U^T(U\Delta^T\Delta U^T + \lambda I)^{-1}U\Delta^T V^T q\|_2^2 \\ &= \|V\Delta U^T(U(\Delta^T\Delta + \lambda I)U^T)^{-1}U\Delta^T V^T q\|_2^2 \end{aligned}$$

Como U es una matriz ortogonal, entonces $UU^T = I$, de aquí que:

$$F^2(\lambda) = \|V(\Delta(\Delta^T\Delta + \lambda I)^{-1}\Delta^T)V^T q\|_2^2$$

Como V es una matriz ortogonal, entonces:

$$F^2(\lambda) = \|(\Delta(\Delta^T\Delta + \lambda I)^{-1}\Delta^T)V^T q\|_2^2$$

Si hacemos $w = V^T q$, la expresión se reduce a:

$$F^2(\lambda) = \|\Delta(\Delta^T \Delta + \lambda I)^{-1} \Delta^T w\|_2^2 \\ = \sum_{k=1}^{n-1} \left(\frac{\sigma_k^2 w_k}{\sigma_k^2 + \lambda} \right)^2$$

entonces:

$$h_k(\lambda) = \sigma_k^4 w_k^2 / (\sigma_k^2 + \lambda)^2$$

Si $0 \leq \lambda_1 < \lambda_2$ entonces:

$$|\sigma_k^4 w_k^2 / (\sigma_k^2 + \lambda_1)^2| > |\sigma_k^4 w_k^2 / (\sigma_k^2 + \lambda_2)^2|$$

queda demostrado que $F(\lambda)$ es estrictamente decreciente.

Ahora hay que demostrar que la segunda derivada es positiva:

$$h_k'(\lambda) = -(2\sigma_k^4 w_k^2) / (\sigma_k^2 + \lambda)^3$$

$$h_k''(\lambda) = 6\sigma_k^4 w_k^2 / (\sigma_k^2 + \lambda)^4 > 0$$

donde $h_k''(\lambda)$ es positiva para $\lambda \geq 0$; $h_k(\lambda)$ es convexa, estrictamente decreciente y $\lim_{\lambda \rightarrow \infty} h_k(\lambda) = 0$.

Por lo tanto, $F^2(\lambda)$ es convexa, estrictamente decreciente y $\lim_{\lambda \rightarrow \infty} F^2(\lambda) = 0$

2.3.5.- CALCULO DEL PARAMETRO LAGRANGIANO.

Para calcular la raíz positiva de la función $F(\lambda)$, se usa el método de Newton. El uso de (2.3.18) es importante en el siguiente desarrollo:

Sea:

$$F(\lambda) = \|DQ(Q^T D^2 Q + \lambda T)^{-1} Q^T Y\|_2$$

por comodidad;

$$\begin{aligned} \mu &= \lambda^{-1} c \\ &= (Q^T D^2 Q + \lambda T)^{-1} Q^T Y \end{aligned}$$

entonces:

$$\begin{aligned} F(\lambda) &= \|DQ\mu\|_2 \\ F^2(\lambda) &= \mu^T Q^T D^2 Q \mu = F(\lambda) \cdot F(\lambda) \end{aligned}$$

Ahora:

$$\frac{FdF}{d\lambda} = \mu^T Q^T D D Q \frac{d\mu}{d\lambda} \quad \text{-----} \quad (2.3.21)$$

de la definición de μ se sigue que:

$$\frac{d\mu}{d\lambda} = -\lambda^{-2} c + \lambda^{-1} \frac{dc}{d\lambda} \quad \text{-----} \quad (2.3.22)$$

derivando con respecto a λ la relación (2.3.18):

$$\begin{aligned} \frac{dc}{d\lambda} &= (Q^T D^2 Q + \lambda T)^{-1} Q^T Y - (Q^T D^2 Q + \lambda T)^{-2} \lambda T Q^T Y \\ &= (Q^T D^2 Q + \lambda T)^{-1} Q^T Y - T ((Q^T D^2 Q + \lambda T)^{-1} c) \\ &= (Q^T D^2 Q + \lambda T)^{-1} (Q^T Y - Tc) \end{aligned} \quad (2.3.23)$$

sustituyendo (2.3.23) en (2.3.22) se obtiene:

$$\begin{aligned} \frac{d\mu}{d\lambda} &= -\lambda^{-2} c + \lambda^{-1} ((Q^T D^2 Q + \lambda T)^{-1} (Q^T Y - Tc)) \\ &= -\lambda^{-2} c + \lambda^{-2} c - \lambda^{-1} (Q^T D^2 Q + \lambda T)^{-1} Tc \\ &= -(Q^T D^2 Q + \lambda T)^{-1} T\mu \end{aligned}$$

o bien:

$$(Q^T D^2 Q + \lambda T) \frac{d\mu}{d\lambda} = -T\mu$$

sustituyendo este último resultado en (2.3.21) se tiene:

$$\frac{FdF}{d\lambda} = \mu^T (-\lambda T \frac{d\mu}{d\lambda} - T\mu)$$

y de aquí:

$$\frac{FdF}{d\lambda} = \lambda \mu^T T (Q^T D^2 Q + \lambda T)^{-1} T\mu - \mu^T T\mu$$

Algoritmo.

Se inicia con $\lambda = 0$.

a) LLevamos a cabo la descomposición de Cholesky $R^T R$ de $(Q^T D^2 Q + \lambda T)$.

b) Calculamos μ de $R^T R \mu = Q^T y$ y $v = D Q \mu$ y acumulándose $e = v^T v$.

c) Si $e > S$

Se calcula $f = \mu^T T \mu$ y $g = w^T w$ donde $R^T w = T \mu$
 se reemplaza λ por $\lambda + \frac{e - \sqrt{5e}}{f - \lambda g}$
 regresamos al paso (a)

sino

d) Se calcula $a = y - Dv$, $c = \lambda \mu$ y b_h, d_h , de acuerdo a (2.3.9) y (2.3.8) respectivamente.

2.4.- SPLINE POR MINIMOS CUADRADOS.

Esta aproximación es muy adecuada cuando se trata de recuperar una función suave que está afectada por algún tipo de "ruido".

2.4.1.- DESARROLLO.

Sea $F(x)$ la función que deseamos aproximar (función verdadera) y sólo conocemos su valor en los puntos $x_k \in [a, b]$ con $x_k < x_{k+1}$, para $k = \overline{1, n-1}$. Si generamos un espacio lineal de dimensión finita de funciones spline (polinomios de cualquier grado, aunque nosotros nos referiremos a los cúbicos únicamente), a través de una base de splines-B (consulte -- Carl de Boor (1978), cap. 9,10,11), lo más probable es que

la función $F(x)$ no sea una combinación lineal de la base.

La base generada por los splines-B tiene la propiedad de estar relativamente bien condicionada, al menos para grados moderados del polinomio. Consulte Carl de Boor (1978) pág. 155.

Si queremos determinar una función que pertenezca al espacio de las funciones spline cúbicas (\mathcal{S}_4), es decir:

$$f(x) \in \mathcal{S}_4$$

que aproxime mejor a $F(x)$, lo podemos hacer usando el principio de mínimos cuadrados, pues la esencia de este método es minimizar una distancia, o sea, que se busca una función en \mathcal{S}_4 tal que:

$$\|F(x) - f(x)\|_2 = \min_{g(x) \in \mathcal{S}_4} \|F(x) - g(x)\|_2 \quad (2.4.1)$$

La dimensión finita de \mathcal{S}_4 garantiza que la función $f(x)$ existe.

La distancia a la que nos referimos está dada por:

$$e(x) = F(x) - f(x)$$

a la cual llamaremos función error.

De esta manera, $f(x)$ es la proyección de $F(x)$ sobre el espacio lineal \mathcal{S}_n , siendo la función error ortogonal a \mathcal{S}_n , es decir:

$$\langle g, F - f \rangle = \sum_{k=1}^n [g(x_k)] [F(x_k) - f(x_k)] w_k = 0 \quad (2.4.2)$$

los w_k son pesos no negativos y asignados comunmente de la siguiente manera:

$$w_k = \begin{cases} \Delta x_k / 2 & \text{para } k = 1 \\ (\Delta x_{k-1} + \Delta x_k) / 2 & \text{para } k = 2, \dots, n-1 \\ \Delta x_{n-1} / 2 & \text{para } k = n \end{cases} \quad (2.4.3)$$

La idea geométrica anterior, puede resumirse en el siguiente lema:

La función $f(x)$ es una mejor aproximación de \mathcal{S}_n a $F(x)$ con respecto a $\|\cdot\|_2$ si y sólo si la función $f(x)$ está en \mathcal{S}_n y la función $F-f$, el error, es ortogonal a \mathcal{S}_n , es decir, para toda $f \in \mathcal{S}_n$, $\langle f, F-f \rangle = 0$.

La demostración puede consultarse en Carl de Boór (1978)-pág. 250-251.

Como toda función en el espacio \mathcal{S}_n es una combinación lineal de la base (B_1, B_2, \dots, B_n) de spline-B, con algunos coeficientes α , entonces la relación (2.4.2) es equivalente a:

$$\langle B_k, F - \sum_{j=1}^n \alpha_j B_j \rangle = 0 \quad \text{con } f = \sum_{j=1}^n \alpha_j B_j \quad \text{--- (2.4.4)}$$

el sistema anterior, tiene solución única (α) . Estas últimas ecuaciones, constituyen las ecuaciones normales escritas generalmente como:

$$\sum_{j=1}^n \langle B_k, B_j \rangle \alpha_j = \langle B_k, F \rangle \quad \text{con } k = \overline{1, n} \quad \text{--- (2.4.5)}$$

la matriz de coeficientes, para las ecuaciones normales anteriores, es una matriz de bandas, con ancho igual a tres, pues:

$$\langle B_k, B_j \rangle = 0 \quad \text{para } |k-j| \geq 4$$

la matriz es también simétrica y definida positiva. Por esta razón puede efectuarse la factorización de Cholesky para resolver el sistema:

$$Ax = b$$

de la siguiente manera:

La matriz A se factoriza en RDR^T , donde R es una matriz triangular inferior y D es diagonal. El sistema triangular:

$$Ry = b$$

se resuelve para y , y el sistema $DR\alpha = y$

da la solución de $A\alpha = b$ obteniendo finalmente la función buscada:

$$f = \sum_{j=0}^M \alpha_j B_j$$

Si $\|\cdot\|_2$ no es una norma en S_4 , entonces la base spline-B para S_4 , no es linealmente independiente. Como consecuencia, el sistema de ecuaciones normales es singular y como siempre existe solución, entonces habrá un número infinito de soluciones. En esta situación se debe elegir una "mejor" aproximación en el sentido de que sea la más pequeña entre todas las aproximaciones por mínimos cuadrados. El método selecciona esta "mejor" aproximación forzando a los coeficientes de los splines-B dependientes a ser iguales a cero.

CAPITULO**T R E S****COMPARACION DE
LAS TECNICAS**

3.1 PLANTEAMIENTO DEL PROBLEMA

Hasta ahora, sólo hemos formulado las hipótesis, propiedades y limitaciones de cada una de las técnicas expuestas en los dos capítulos anteriores. La comparación y evaluación propiamente dichas, las lograremos -- probándolas bajo condiciones experimentales controladas, es decir, simulando una serie de tiempo y aplicando las técnicas para analizar su eficiencia y características.

Este capítulo presenta un aspecto importante para -- nuestro trabajo, que consiste en la discusión de las -- técnicas:

- Mínimos cuadrados y promedio móvil contra
- La función spline.

que pertenecen a dos disciplinas distintas de la matemática. Las dos primeras técnicas se han usado tradicionalmente para estimar a la tendencia de una serie de tiempo, mientras que la función spline se usa por primera vez (hasta donde logramos investigar) para este propósito.

El desarrollo del problema será en base a un conjunto de preguntas y respuestas, que enmarcarán nuestro objetivo para lograr visualizar las distintas alternativas que hay al atacar el problema propuesto.

1. ¿Por qué se generó una sólo serie de tiempo?

Resulta difícil simular y manejar un conjunto suficientemente grande de series de tiempo, para poder obtener conclusiones contundentes, dado que si se quisiera hacer un diseño de experimentos, se tendrían que considerar los siguientes factores:

1.- Modelos	{	Aditivo Multiplicativo Mixto
-------------	---	------------------------------------

2.- Tendencia	{	Lineal Polinomial Exponencial Logarítmica
---------------	---	--

3.- Estacionalidad { Estable
Variable

4.- Irregular { Suave
Fuerte
Errático

El número de series requeridas es mayor a setenta, a las que habría de aplicarle cada uno de los métodos discutidos con sus variantes respectivas, para entonces realizar un análisis exhaustivo y poder así concluir algo válido en forma general. Esto resulta muy costoso, de mucha dedicación y tiempo, por lo que nosotros nos limitaremos a hacer una ilustración comparativa de donde se refleje el comportamiento que observó cada una de las técnicas al ser aplicadas a una sólo serie de tiempo simulada y aunque no se lograrán conclusiones contundentes, -- pensamos que es posible que el presente trabajo pueda servir como punto de partida para investigaciones posteriores, en donde se trate de analizar alguna técnica en particular, contando ya con la facilidad de disponer de la teoría desarrollada, de los programas para la computadora y de un ejemplo práctico de cómo pueden manejarse tales métodos.

2. ¿Qué tipo de serie de tiempo es la indicada para lograr nuestro objetivo?

Ante esta pregunta, se podría elegir entre dos alternativas:

- a) Series empíricas
- b) Series teóricas.

En la primera alternativa están incluidas las series de ventas, precios, producción, etc., pero estas presentan muchos inconvenientes. Por ejemplo, el hecho de que algunas series oficiales o de empresas privadas tienen la particularidad de que en ocasiones están encaminadas con fines políticos o personales; además, la mayoría de las veces se cuenta con poca información y no se les ha dado un tratamiento estadístico adecuado por personal -- técnico capacitado, esto traerá consigo una imprecisión que valdría más no tener.

No debemos olvidar que al tratar con series empíricas carecemos de las componentes verdaderas de la serie de tiempo, por lo que sólo se podrá estimar a cada una de ellas, esta razón las elimina por completo.

En cuanto a la segunda alternativa se debe subrayar, que es posible describir matemáticamente (a través de un modelo) el comportamiento de algunos fenómenos económicos como índices de inflación, gasto público (Fox Jenkins-1969). De modo que si se toma un modelo multiplicativo (sección 1.6) de los que describen usualmente a se-

ries económicas y se les alimenta en una forma adecuada, se obtendrá como resultado, una serie teórica con ciertas características económicas, con la ventaja de que la serie así generada, puede ser tan grande como se quiera, con tendencia no lineal conocida, una estacionalidad estable y una componente irregular cuyos elementos sigan una distribución normal con media cero y varianza uno.

3. ¿Por qué la serie teórica debe tener rasgos económicos?

Esta decisión se tomó debido a que las series económicas son las que tienen un uso más conocido, y en donde podemos encontrar, una gran variedad de series, cada una de las cuales tiene características propias, que las diferencian de las demás.

No es posible tratar de generar una serie de tiempo económica, que posea todos los rasgos que caracterizan a los fenómenos económicos, por tal razón se simuló una serie que podría ser del tipo común, como ventas, producción, etcétera.

En este trabajo, simularemos una serie de tiempo económica para poder observar lo que podría suceder al tratar de representar la tendencia con cada una de las técnicas.

Para efectuar la comparación se requieren considerar los siguientes aspectos:

a) Asignar los parámetros al modelo para conocer a cada una de las componentes de la serie.

b) A tal serie, le aplicaremos las técnicas de suavizado con distintas modificaciones para cuantificar sus efectos.

c) Decir cuál de las técnicas es la mejor y bajo qué condiciones.

d) Por las razones expuestas, nos resulta conveniente generar una serie de tiempo teórica con ciertas características económicas.

4. De todas las clases de mínimos cuadrados, de promedios móviles y de splines ¿Cuáles resultan apropiados para ser comparados?

Para contestar esta pregunta, primero se tuvieron que revisar detalladamente cada uno de los métodos, para determinar aquellos que pudieran, con mayor ventaja, ajustar una curva suave a una serie de tiempo.

Selección de un modelo de mínimos cuadrados.

En la sección 1.2 se expuso el modelo lineal general y el modelo no lineal, de ellos se tomó el modelo cúbico:

$$Y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

Como el promedio móvil y el spline también pueden desarrollarse a partir de este polinomio cúbico, resulta

que las cuatro técnicas quedan en igualdad de circunstancias para ser comparadas.

Selección de un promedio móvil.

Inicialmente se realizó una investigación para recolectar en su mayoría todas las técnicas que tenían como esencia un promedio móvil. Como resultado se obtuvo que algunos promedios tenían usos específicos, como por ejemplo, - el promedio móvil de Spencer de 15 y 21 puntos empleado - para suavizar series (Kendall-1976), sin embargo, tenían una desventaja en común, que al estimar la tendencia de - una serie, siempre pierden valores de tendencia en ambos extremos de la serie nueva (como se puede notar en algunos de los procedimientos de mayor relevancia expuestos - en la sección 1.4). Hay un procedimiento que cuenta con una extensión de sus propios elementos para encontrar esos valores extremos, que son los promedios móviles con pesos calculados bajo el principio de mínimos cuadrados (sección 1.4.5), motivo suficiente para elegirlo como el indicado - para la comparación. Cabe mencionar que los pesos pueden obtenerse mediante un polinomio cúbico o bien de quinto - grado. Aunque en este último caso el procedimiento se complica, ganando poca precisión.

Hay ciertas técnicas conocidas para estimar la tendencia que están implantadas en paquetes estadísticos -----

que se manejan a través de una computadora, por ejemplo X-11, SABL, BAYSEA, etc., sin embargo, requieren de una máquina de gran capacidad que posea tales paquetes, lo cual no estaba a nuestro alcance.

Selección de las dos funciones spline.

En lo que corresponde a splines, encontramos que los de mayor uso en la práctica son los cúbicos simples y - que éstos pueden ser de interpolación o aproximación. El algoritmo desarrollado por Reinsch puede usarse cuando los datos están influidos por un cierto error, es decir, $y_k = g(t_k) + \epsilon_k$ donde $g(t_k)$ se supone una función suave y ϵ_k es el error ($k = 0, 1, \dots, n$). En nuestro caso, tal error lo podemos relacionar como la estacionalidad y error aleatorio de la serie de tiempo; por esta razón lo seleccionamos para la comparación.

También se seleccionó el spline por mínimos cuadrados, pues cuenta con la característica de recuperar una función suave que está afectada por algún tipo de "ruido".

Los splines de quinto grado logran un poco más de precisión, pero requieren más cálculos y por lo tanto más tiempo de procesamiento para su solución. Hay otras clases de splines que son desarrollados en distintas dimensiones y espacios funcionales para casos muy específicos, requiriendo conocer temas muy avanzados en mate-

máticas (como el análisis funcional), razón por la que fueron eliminados, considerando los objetivos del presente trabajo.

En lo sucesivo, cada vez que se mencione el principio de mínimos cuadrados, el promedio móvil y cualquiera de los dos splines, nos referimos en forma específica a los elegidos anteriormente.

5. ¿Cómo podríamos medir objetivamente la eficiencia de un método de suavizamiento, con respecto a los otros?

Ya se dijo anteriormente, que se pretende suavizar una serie de tiempo, aunque no en el sentido de suavizamiento máximo, sino más bien para representar a la tendencia. Un buen suavizamiento entonces, será aquel que se aproxime mejor a la tendencia de la serie, esto hace fundamental el conocimiento exacto de esta componente para poder medir la eficiencia de cada método.

Una vez que se apliquen los métodos a la serie de tiempo, se compararán las curvas suaves ajustadas contra la tendencia teórica ξ_t (sección 3.2.2) de la serie, siendo la mejor aquella que minimice la raíz cuadrada del error cuadrático medio (Row - Miller, 1971).

$$\text{E.C.M.} = \left(\frac{\sum_{k=1}^n (T_k - \xi_k)^2}{n} \right)$$

donde:

T_k - La tendencia obtenida con alguno de los métodos

ξ_k - La tendencia verdadera

n - Número de observaciones.

6. ¿Qué es el ajuste de curvas?

Es la elección de una función matemática junto con -
la determinación de sus parámetros, para representar a -
un conjunto de observaciones.

3.2 SIMULACION DE UNA SERIE

3.2.1 SIMULACION

DEFINICION.- Es el proceso de diseño de un modelo matemático, para imitar un fenómeno real, con el propósito de realizar experimentos controlados que nos den información de su comportamiento.

La simulación es sólo una herramienta que ayuda al análisis de la conducta de los fenómenos reales bajo condiciones especificadas por el investigador. Esta disciplina implica por sí sola el empleo de distribuciones de probabilidad.

Los resultados de una simulación nos forman un criterio acerca del problema, la cual nos ayuda a la toma de decisiones.

7. ¿Cuándo se debe usar la simulación?

a) Si el problema que se tiene es muy complejo para resolverlo.

b) La segunda razón incluye a aquellos problemas en el que el investigador necesita ganar experiencia e información de una situación real compleja y tener la posibilidad de manejarla en forma similar.

c) El tercer tipo de situación, es aquella en que no existe el problema en el mundo real, pero uno desea anticiparse y saber cómo tratarlo si éste apareciera.

Una serie simulada difiere de una serie real, en que las componentes de la primera reflejan las características generales de las componentes reales pero están desprovistas de un contenido económico-social, en el sentido de que no representa un fenómeno en particular.

En oposición a los fenómenos experimentales, los fenómenos sociales son empíricos, pues no se tiene control de los parámetros que los originan, por esta razón, nosotros simularemos una serie de tiempo con algunos rasgos económicos, a través de un modelo multipli-

cativo, en donde se tiene control de los parámetros que actúan sobre la tendencia y la componente irregular, en tanto que la estacionalidad se asignará en forma determinista. De esta manera, la simulación nos permite conocer con precisión a cada una de las componentes.

Al experimentar con el modelo, se fueron controlando sus parámetros para reproducir una serie de tiempo hipotética lo más cercana posible a las del tipo común.

Por lo tanto, la serie simulada es entonces el resultado de multiplicar las componentes de acuerdo con el modelo elegido.

La facilidad de reproducir y controlar las características de comportamiento de las componentes en forma independiente, nos permite conocer perfectamente la tendencia de la serie. Este hecho es fundamental para el objetivo de nuestro trabajo, pues al aplicar los métodos de suavizamiento a la serie hipotética estimaremos a la tendencia, tal estimación podrá ser comparada posteriormente, con el comportamiento de la tendencia simulada, para determinar cuál la representa mejor.

3.2.2 CONSTRUCCION DEL MODELO.

Como ya se dijo, el modelo indicado para simular una serie de tiempo económica, con las características re-

queridas es del tipo multiplicativo:

$$U_k = \xi_k \eta_k \epsilon_k \quad k = 1, 2, \dots, 120$$

donde cada una de las componentes se representará en la forma más sencilla.

La razón por la que se consideran ciento veinte observaciones fue porque se requería suficiente información - para poder hacer un análisis en tres etapas distintas de la serie, es decir, una inicial en los primeros doce meses, una central en los siguientes noventa y seis y una final en los restantes doce meses.

La tendencia ξ se generó a partir de un polinomio de segundo grado dado por:

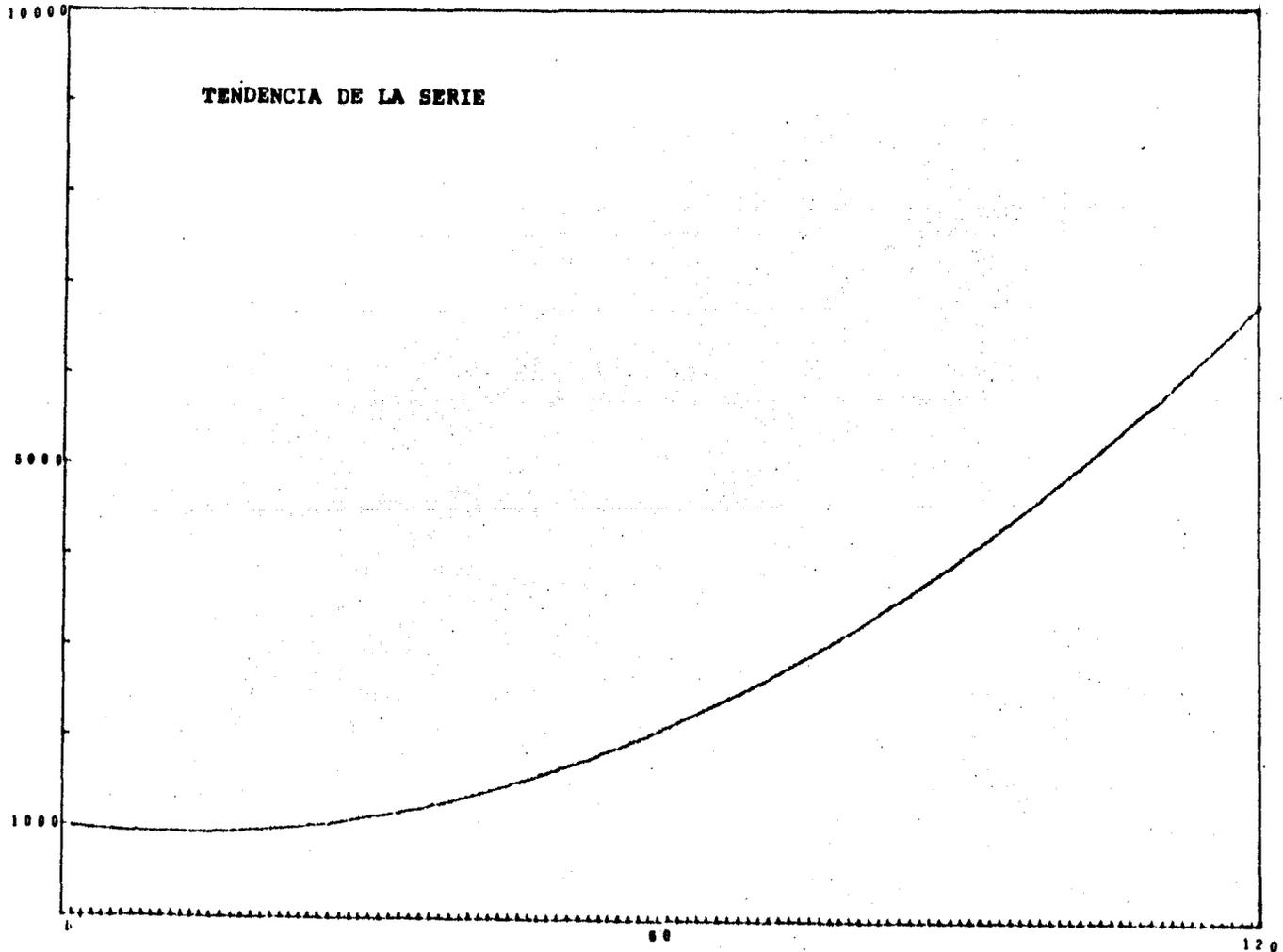
$$\xi_k = 0.5t^2 + 12t + 1000$$

(Ver gráfica 3).

$$k = t = 1, 2, \dots, 120$$

9. ¿Por qué la tendencia se representó por un polinomio cuadrático?

Si se hubiera seleccionado una recta como la tendencia caeríamos en el caso más simple, que rara vez ocurre en la realidad, puesto que la mayoría de los fenómenos socio-económicos siguen un comportamiento no lineal. Por lo anterior, es conveniente para fines prácticos tomar un polinomio de segundo grado.



TENDENCIA DE LA SERIE

GRAFICA (3)

La estacionalidad η se asignó en forma determinista y estable, satisfaciendo la siguiente condición:

$$\prod_{k=1}^{12} \eta_k = 1$$

Donde:

$\eta_k = 0.6999$	$k = 1,$	(12), 109	(ENERO)
$\eta_k = 0.8$	$k = 2,$	(12), 110	(FEBRERO)
$\eta_k = 0.9$	$k = 3,$	(12), 111	(MARZO)
$\eta_k = 1.1$	$k = 4,$	(12), 112	(ABRIL)
$\eta_k = 1.2$	$k = 5,$	(12), 113	(MAYO)
$\eta_k = 1.12$	$k = 6,$	(12), 114	(JUNIO)
$\eta_k = 1.1$	$k = 7,$	(12), 115	(JULIO)
$\eta_k = 0.896$	$k = 8,$	(12), 116	(AGOSTO)
$\eta_k = 0.97$	$k = 9,$	(12), 117	(SEPTIEMBRE)
$\eta_k = 0.9$	$k = 10,$	(12), 118	(OCTUBRE)
$\eta_k = 1.2$	$k = 11,$	(12), 119	(NOVIEMBRE)
$\eta_k = 1.3$	$k = 12,$	(12), 120	(DICIEMBRE)

La manera de interpretar el índice estacional es como sigue:

Para los eneros tenemos un índice de 0.6999 lo que significa que la serie está por debajo de la tendencia en un 30.01%, o sea, como el valor de tendencia para el primer enero es† 988.5 X 0.6999 = 691.851, o bien, el caso de diciembre con $\eta_k = 1.3$ nos dice que está por arriba de la tendencia en un 30%, es decir, como el valor -

†(errata) 988.5 este valor se reduce a

de tendencia para el primer diciembre es de 928 entonces se incrementa a $928 \times 1.3 = 1206.4$.

(Ver gráfica 4).

La componente irregular ϵ se generó a partir de una distribución de probabilidad normal estandarizada, la fórmula* que se usó fue:

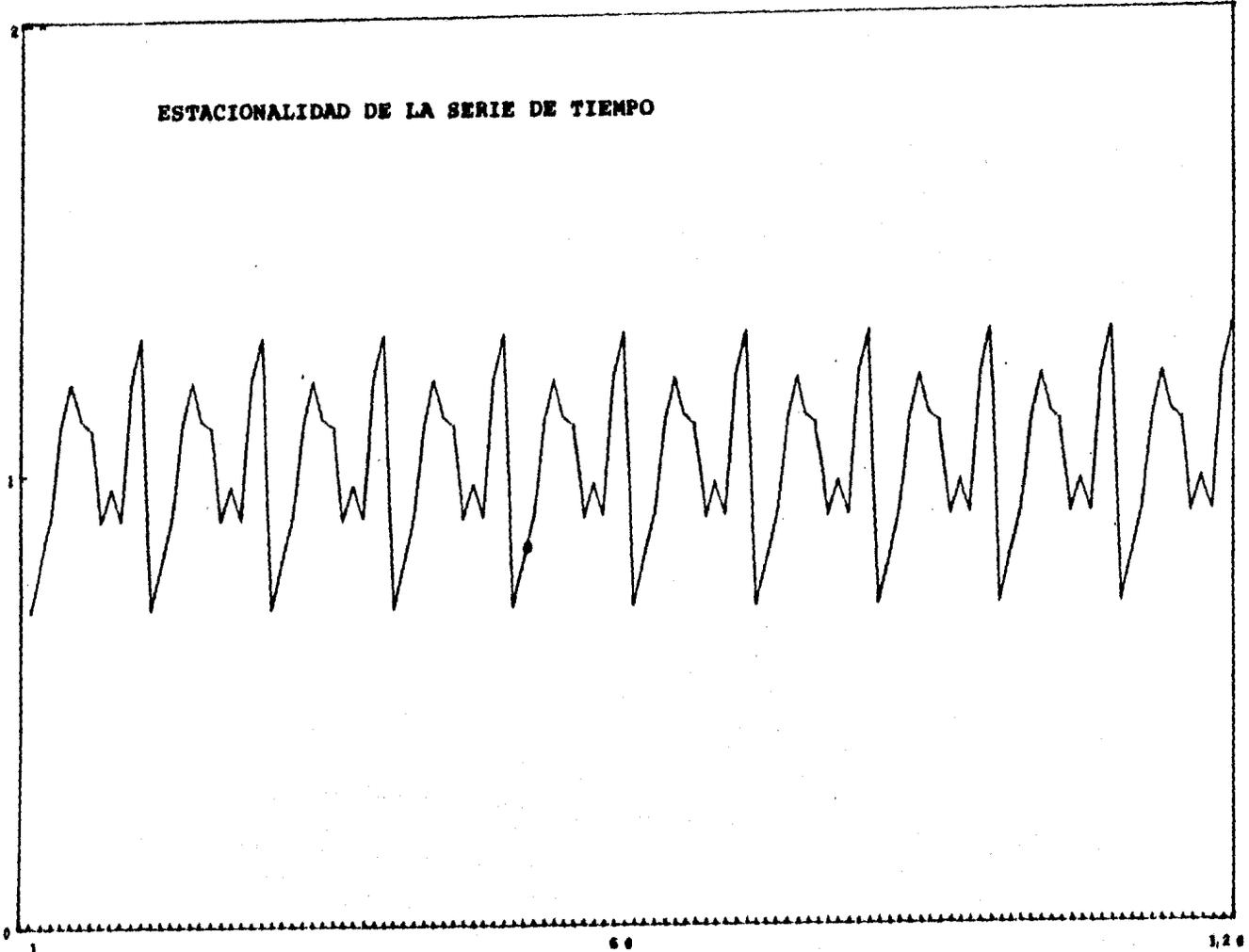
$$I_k = (-2 \log(u))^{1/2} \sin(2\pi v) \quad k = 1, 2, \dots, 120$$

y semillas $u = 0.999999999999$

$v = 0.12541254129$

Con esta fórmula, se obtuvieron 256 números aleatorios, esta cantidad es debido a que la longitud máxima del vector de la máquina (HP - 9830A) sólo acepta hasta ese subíndice, de aquí se tomaron los últimos ciento veinte, pues estos valores se aproximan mejor a una distribución normal. La media y varianzas muestrales que se obtuvieron fueron de 6.1 y 32 respectivamente. Como era de esperarse algunos números calculados por la fórmula anterior, resultaron ser negativos y como ya se explicó (sección 1.6) que esto resulta inconveniente para un modelo multiplicativo, como el que se está cons

 * Técnicas de simulación en computadoras, por Naylor --
 Bacinty, Burdier Kon y Chu, Editorial Limusa.



ESTACIONALIDAD DE LA SERIE DE TIEMPO

GRAFICA (4)

truyendo, se hizo una transformación a través de la función exponencial como se vió en la sección 1.6.5. La media y varianza nuevas debidas a la transformación fueron 0.05 y 1.12 respectivamente. De esta manera ya todos los números resultantes fueron positivos, sin embargo, todavía faltaba un pequeño problema a resolver, el cual consistía en que estos errores afectaban fuertemente a la serie, entonces, para reducir su efecto se dividió a cada argumento de la exponencial entre 25, dando como resultado errores aleatorios más pequeños, alrededor de 1. Estos errores aleatorios tienen la misma interpretación que el factor estacional, pero a una escala mucho menor (ver gráfica 5).

Finalmente el modelo apropiado para generar la serie económica deseada fue:

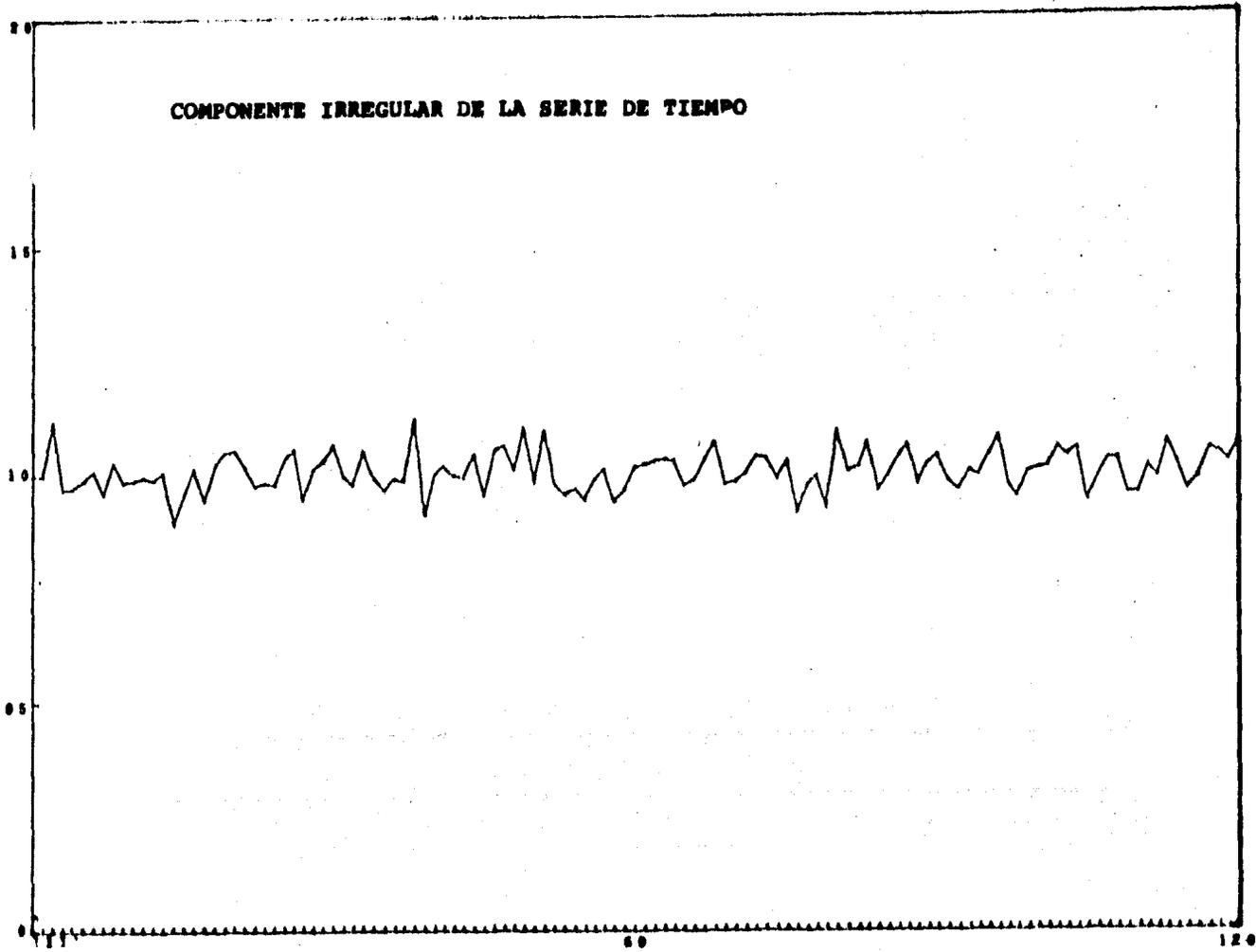
$$U_k = T_k E_k \exp(I (k + 136) / 25) ; k=1,2,\dots,120$$

T_k - Es la tendencia

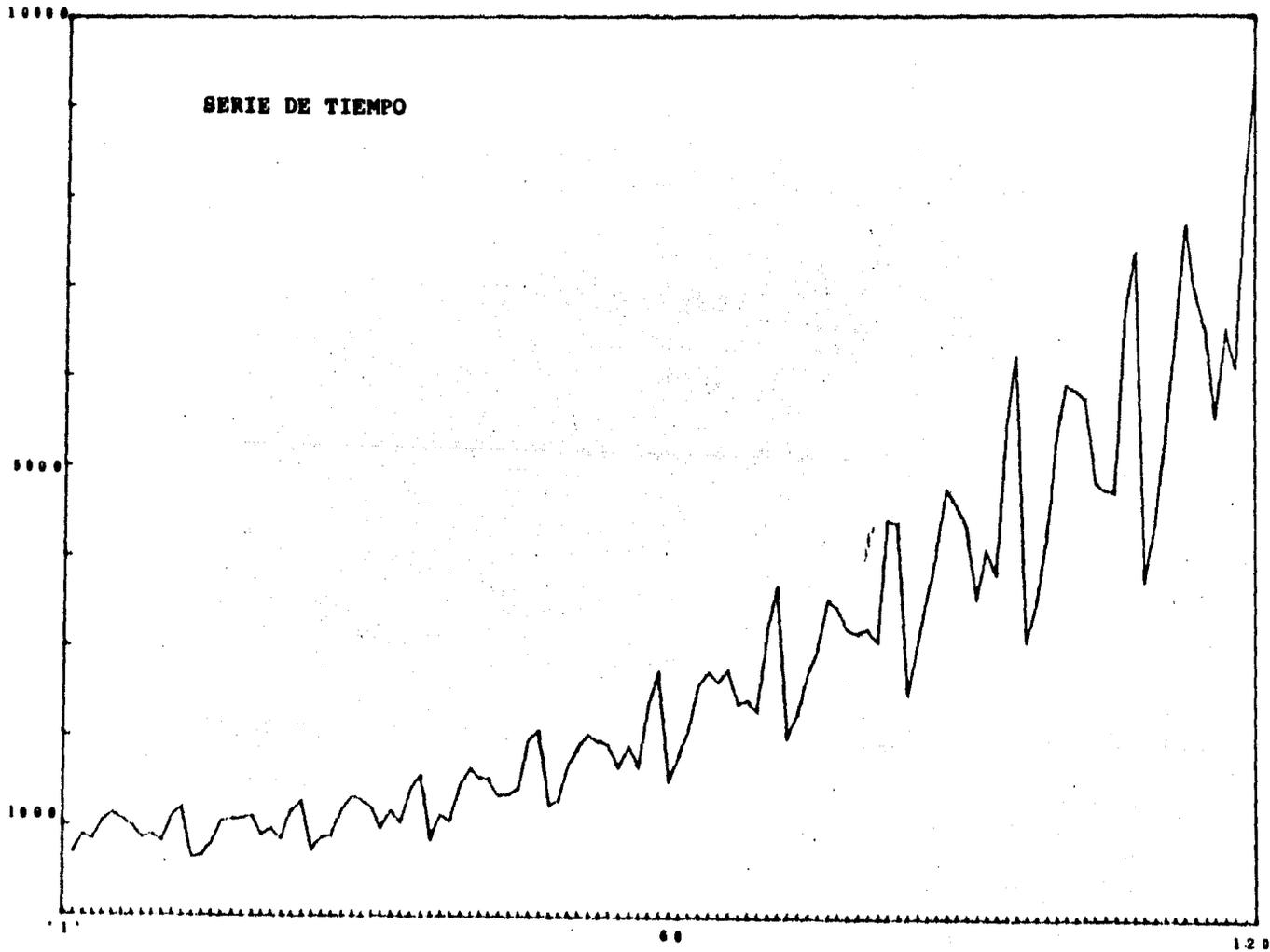
E_k - Es la estacionalidad

I_k - Es el error aleatorio.

De esta manera, la serie simulada es la resultante de las tres componentes relacionadas que actúan en forma sistemática (ver gráfica 6).



GRAFICA (5)



GRAFICA (6)

3.3 COMPARACION

A continuación se mencionará las relaciones, similitudes y diferencias que existen entre las técnicas.

La relación entre los métodos, de acuerdo al objetivo de la tesis, es la representación de la tendencia de la serie de tiempo.

En seguida, separaremos las similitudes y diferencias generales de las particulares en los cuatro métodos, para evitar repeticiones.

3.3.1 Similitudes generales para los cuatro Métodos.

a) Las cuatro técnicas usan polinomios de la siguien

te manera:

- El promedio móvil los emplea para calcular los pe sos mediante los cuales suaviza a la serie de tiempo.

- Mínimos cuadrados debe calcular los parámetros para un sólo polinomio de cualquier grado.

- El spline de aproximación (Reinsch) requiere de una serie de cálculos para obtener los parámetros que determinan a los polinomios impares y que al unirlos a cada punto t_k constituyen a la función spline de --- aproximación.

- Splines por mínimos cuadrados también debe calcular los parámetros para uno o más polinomios de cualquier grado según el número de subintervalos elegidos.

3,3,2 Similitudes particulares.

a) Promedio móvil y función spline de aproximación:

1) Tanto con este método como en el otro, se puede controlar la suavidad de la serie que generan, así, para el promedio móvil se usarán las extensiones de cinco a veintiuno y para el spline el coeficiente de suavidad S (sección 2.3.1).

2) Los dos métodos presentan mayor pérdida de precisión al calcular los valores de tendencia en los extremos de la serie de tiempo que en su parte central como se puede apreciar en las tablas 6 y 11.

3) Las extensiones del promedio móvil así como el grado del polinomio de la función spline conviene que sean impares, por las razones expuestas en las secciones 1.4 - y 2.2.8 respectivamente.

b) Promedio móvil y mínimos cuadrados.

Los dos pueden obtenerse por calculadora.

c) Promedio móvil y spline por mínimos cuadrados.

Los dos tienen forma de controlar la suavidad, el promedio móvil por la extensión del promedio y el spline por mínimos cuadrados a través de los puntos de quiebre.

d) Spline de aproximación y mínimos cuadrados.

El spline minimiza a una integral (2.3.1) y mínimos cuadrados; minimiza a una suma de cuadrados (1.2.9).

e) Spline de aproximación y spline por mínimos cuadrados.

1) Ambas técnicas emplean pesos con las mismas características, enteros o fraccionarios pero positivos.

2) Los dos tienen forma de controlar la suavidad: el spline de aproximación mediante el coeficiente de suavidad S y el spline por mínimos cuadrados con la sucesión de puntos de quiebre.

3) Los dos splines presentan más o menos el mismo ----

grado de complejidad en su desarrollo y cálculo.

4) Tanto un método como el otro requieren ser resueltos por computadora.

5) Ambas técnicas presentan restricciones en su desarrollo.

6) Ninguna de las dos técnicas provee de un procedimiento seguro para asignar pesos.

7) Los dos procedimientos presentan matrices tridimensionales, definidas positivas y los sistemas de ecuaciones lineales se resuelven por el método de Cholesky.

f) Spline por mínimos cuadrados y mínimos cuadrados.

Las dos técnicas tienen por objetivo minimizar la distancia que existe entre las observaciones y la curva a ajustar, aunque cada una tiene fundamentos distintos. Así el spline por mínimos cuadrados parte de conceptos de álgebra lineal, mientras que mínimos cuadrados lo hace a través del cálculo diferencial.

3.3.3 Diferencias generales para los cuatro métodos.

a) Una ventaja del promedio móvil sobre los otros tres métodos, es que al añadir un dato no es necesario recalcular todos los valores de tendencia nuevamente -- (tiene permanencia), sino tan solo, los del extremo donde fue añadido calculando el término U_{N+M} y recalculando lo los últimos M términos de la serie. (sección 1.4.5).

b) Los valores de tendencia obtenidos con promedio móvil son discretos, no se tiene una función matemática que los represente. Para los otros tres métodos estos valores si están definidos en una función continua (y diferenciable), tal que nos permite conocer valores entre los puntos t_k ($k = 1, 2, \dots, 120$).

c) El desarrollo del promedio móvil no tiene ninguna restricción, a diferencia de los otros métodos, los cuales cuentan con varias restricciones de la función, como se puede apreciar en sus respectivas secciones.

d) Una característica relevante que se notó en el promedio móvil, es que se puede aplicar en forma iterativa (segundo orden), mientras que para la función spline tendríamos el problema de asignar nuevamente los pesos y para mínimos cuadrados no tiene sentido.

e) Hay una clase especial de promedio móvil para desestacionalizar la serie de tiempo (promedio móvil centrado, sección 1.4.6), sin embargo en los otros tres métodos no hay una clase de función particular que ataque directamente este problema.

3.3.4 Diferencias particulares:

a) Promedio móvil y spline de aproximación.

1) Una diferencia notable es que el promedio móvil, en esencia, surge de una simple media aritmética, en contraste con la función spline cuya idea es más compleja.

2) El promedio móvil se puede obtener fácilmente con la ayuda de una calculadora, mientras que el algoritmo de la función spline requiere de ser programado en una computadora y su tiempo de procesamiento es mayor que el de promedio móvil, en consecuencia resulta ser más caro.

3) La técnica de promedio móvil, provee de antemano un procedimiento particular para calcular los pesos que se emplearán en la estimación de la tendencia, los cuales pueden ser positivos o negativos. A la función spline por el contrario, se le deben asignar los pesos y ser estrictamente positivos.

4) En el caso del promedio móvil, los pesos para calcular los valores de tendencia en los extremos de la serie se determinan por una tabulación de un polinomio cúbico, mientras que para el spline los valores cercanos a sus extremos se ven fuertemente afectados por la selección arbitraria de sus condiciones en la frontera.

b) El promedio móvil y mínimos cuadrados.

1) El promedio móvil puede controlar la suavidad de acuerdo a la extensión del promedio. Mínimos cuadrados puede hacerlo sólo en forma indirecta con el grado del polinomio.

2) La pérdida de precisión en los extremos de la serie con el método de mínimos cuadrados es menor que la que se obtiene con el promedio móvil.

3) Mientras que para el promedio móvil resulta conveniente el empleo de polinomios impares (sección 1.4.3) - mínimos cuadrados no presenta esta desventaja.

4) Mínimos cuadrados tiene un grado mayor de dificultad que promedio móvil tanto en cálculo como en desarrollo.

5) Mínimos cuadrados permite pronosticar.

c) Promedio móvil y spline por mínimos cuadrados.

1) El promedio móvil tiene un procedimiento particular para calcular sus pesos, que deben ser positivos o negativos. El spline por mínimos cuadrados sólo sugiere una posible elección y pueden ser enteros o fraccionarios o positivos.

d) Splíne de aproximación y mínimos cuadrados,

1) El splíne de aproximación controla la suavidad de la función que construye a través del coeficiente de suavidad S , mínimos cuadrados lo hace en forma indirecta con el grado del polinomio.

2) El desarrollo del método de splíne de aproximación es más complicado que el de mínimos cuadrados.

3) Mínimos cuadrados puede obtenerse con calculadora, por el contrario, el splíne de aproximación necesita ser calculado por computadora.

4) La pérdida de precisión en los extremos de la serie, con el método de mínimos cuadrados, es menor que la que se obtiene con el splíne de aproximación.

5) Mínimos cuadrados no presenta condiciones en la frontera a diferencia de la función splíne natural de aproximación que está sujeta a:

$$g''(t_0^-) = g''(t_n^+) = 0$$

e) Splíne de aproximación y splíne por mínimos cuadrados.

1) Ambos métodos parten de conceptos diferentes como se puede apreciar en las secciones 2.3 y 2.4.

2) El splíne de aproximación tendrá, para N puntos,

$N-1$ polinomios impares (cúbicos) en tanto que en el -- spline por mínimos cuadrados se tiene uno o más polinomios de cualquier grado (aunque todos del mismo grado) ya que se puede seleccionar el número de subintervalos.

d) Spline por mínimos cuadrados y mínimos cuadrados.

1) Los dos parten de enfoques distintos como se puede apreciar en las secciones 1.2 y 2.4.

2) El spline por mínimos cuadrados controla la suavidad con la sucesión de puntos de quiebre, mínimos cuadrados lo hace mediante el grado del polinomio.

3) El spline por mínimos cuadrados puede construirse con uno o más polinomios a diferencia de mínimos cuadrados, el cual se limita al empleo de uno.

EVALUACION

EN LAS SECCIONES SIGUIENTES PRESENTAREMOS EL ASPECTO
CUANTITATIVO DEL PROBLEMA, ES DECIR, MEDIREMOS LA EFICIEN
CIA DE CADA METODO PARA CONOCER CUAL DESCRIBE MEJOR A LA
TENDENCIA DE LA SERIE DE TIEMPO.

3.4 EVALUACION DE MINIMOS CUADRADOS

La estimación de la tendencia de una serie de tiempo por mínimos cuadrados, es un caso particular del análisis de regresión, puesto que la variable independiente x en las series de tiempo, esta limitada a representar invariablemente unidades de tiempo, mientras que en análisis de regresión, esta variable puede representar a cualquier fenómeno.

En esta rama de la estadística, se eligió a un modelo de regresión cúbico como el apropiado para estimar la tendencia de la serie simulada.

10. ¿Por qué un modelo de regresión cúbico?

Esta decisión fue de gran trascendencia para nuestro trabajo, pues un polinomio cúbico del tipo:

$$Y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

será el que sirva de fundamento para desarrollar cualquiera de las técnicas para ajuste de curvas suaves, - que intervienen en la investigación. Esta situación - nos permite tener en igualdad de circunstancias (en -- cuanto a construcción) a las técnicas elegidas para -- ser comparadas.

De acuerdo con la (1.2.35) para $k=3$, $n=120$ y t como variable independiente se tiene:

$$\sum_{k=1}^{120} (Y_k - \beta_0 - \beta_1 t - \beta_2 t^2 - \beta_3 t^3)^2$$

Usando la relación (1.2.36) tendremos un sistema de cuatro ecuaciones, que nos permitirá calcular el vec--tor b_0 :

$$120b_0 + b_1 \sum t + b_2 \sum t^2 + b_3 \sum t^3 = \sum y$$

$$b_0 \sum t + b_1 \sum t^2 + b_2 \sum t^3 + b_3 \sum t^4 = \sum Yt$$

$$b_0 \sum t^2 + b_1 \sum t^3 + b_2 \sum t^4 + b_3 \sum t^5 = \sum Yt^2$$

$$b_0 \sum t^3 + b_1 \sum t^4 + b_2 \sum t^5 + b_3 \sum t^6 = \sum Yt^3$$

Resolviendo el sistema por eliminación, se llegó a los siguientes resultados:

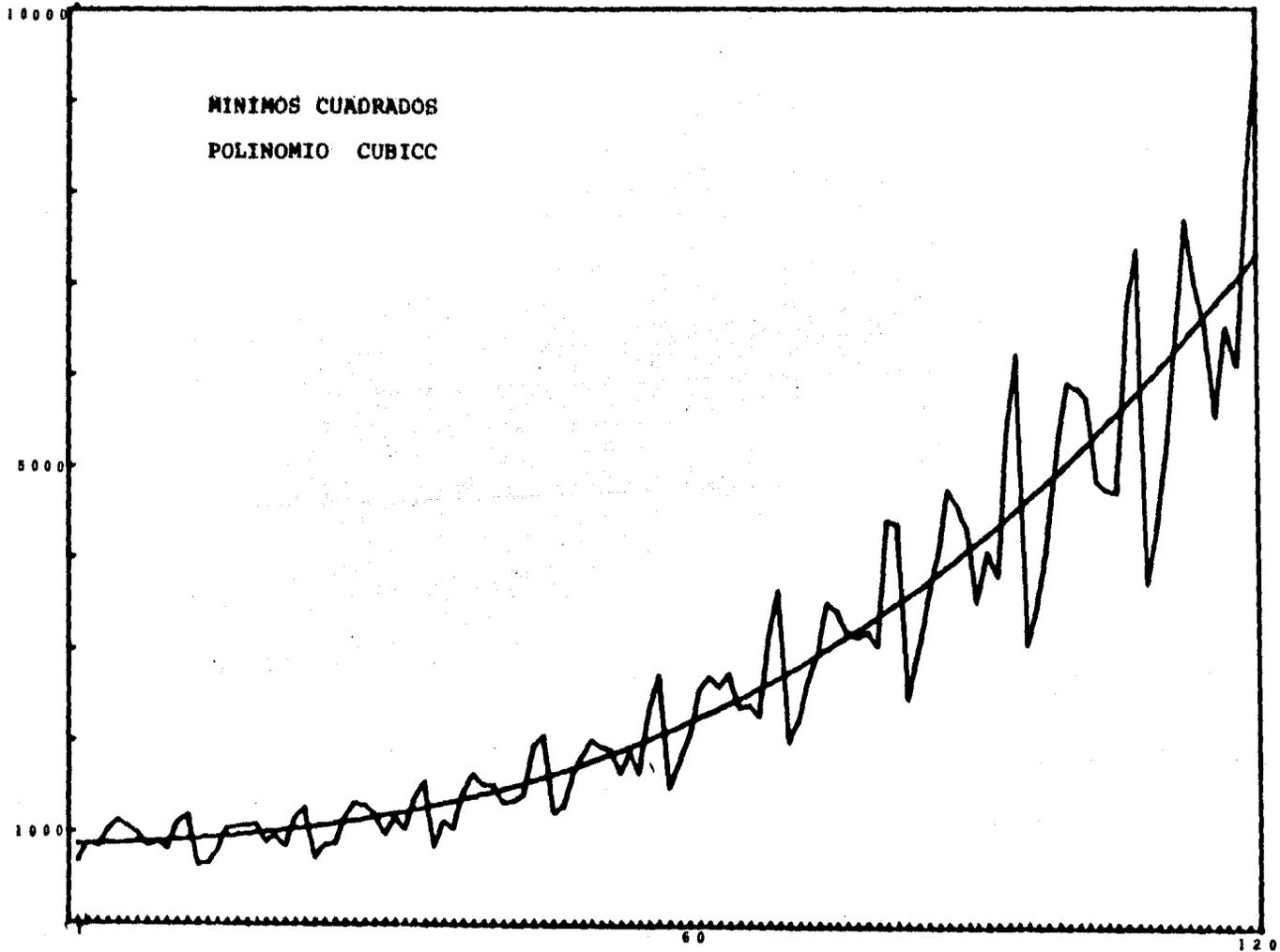
$$b_0 = 859.564\ 19850$$

$$b_1 = 3.680\ 10161$$

$$b_2 = 0.137\ 77264$$

$$b_3 = 0.002\ 3302686$$

La curva de regresión Y_R que representa a la tendencia la obtenemos al tabular (1.2.15) con los parámetros estimados (ver gráfica 7).



GRAFICA (7)

3.5 EVALUACION DEL PROMEDIO MOVIL

En base a lo explicado en la sección 1.4 procederemos a probar distintos tipos de promedio móvil en la serie simulada como son:

- a) Promedio móvil de primer orden
 - b) Promedio móvil de segundo orden
 - c) Promedio móvil mixto
 - d) Promedio móvil centrado
- a) Promedio móvil de primer orden.

A la serie de tiempo simulada se le fueron aplicando promedios móviles, con pesos obtenidos por mínimos cuadrados, en forma sistemática desde una extensión de cinco a veintiun términos. A medida que se iba incre-

mentando la extensión, se obtenía como consecuencia -- una mejor aproximación a la tendencia, es decir, una serie nueva más suave. No debemos olvidar que en este proceso se buscará un equilibrio entre la extensión apropiada y los términos de los extremos de la serie -- que estamos dispuestos a perder, pues a medida que se incrementa la extensión del promedio móvil se obtiene mayor suavidad pero pierden más valores de tendencia.

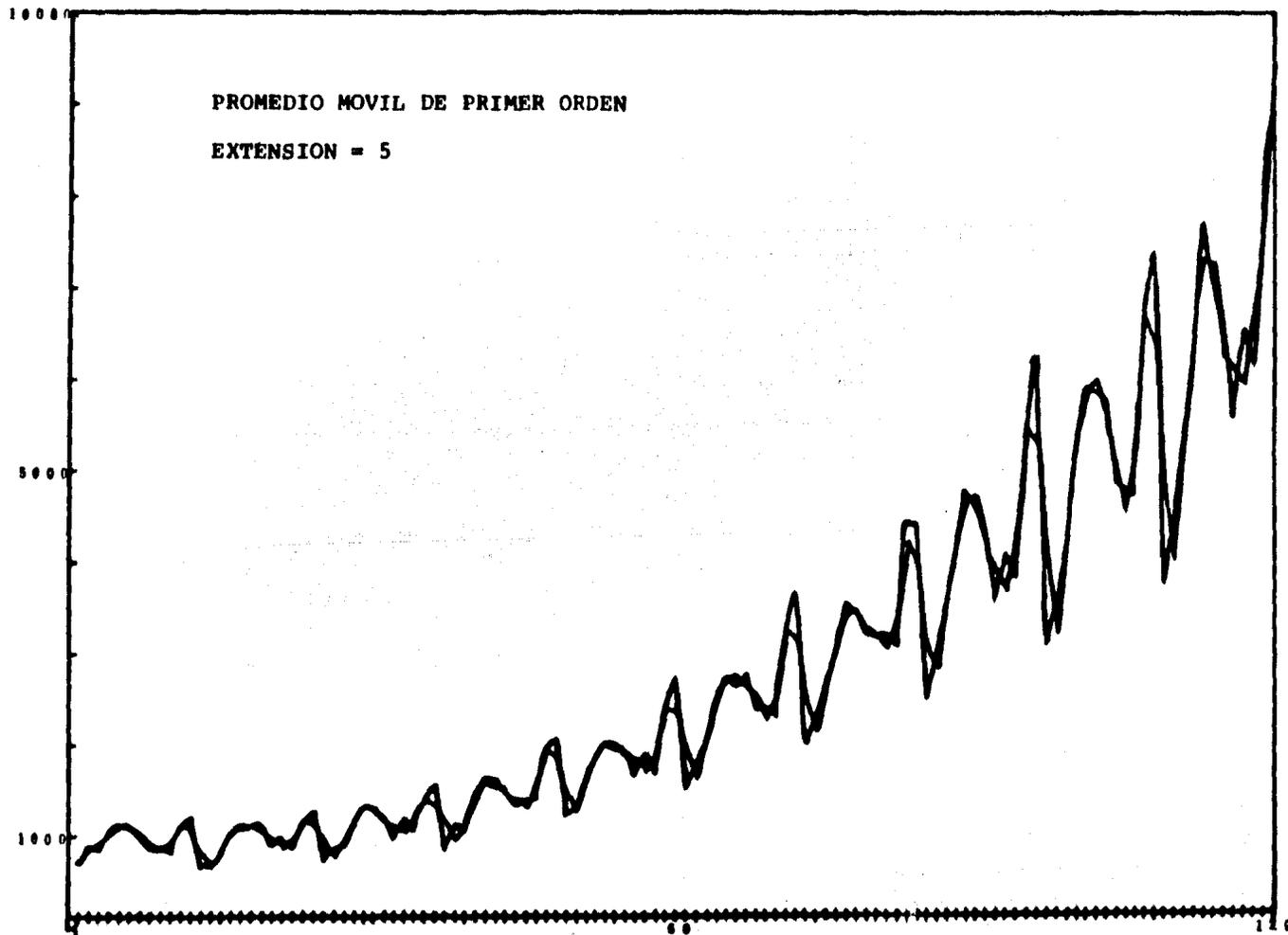
Un primer análisis del comportamiento general de cada una de las extensiones lo da la tabla 4:

TABLA (4)

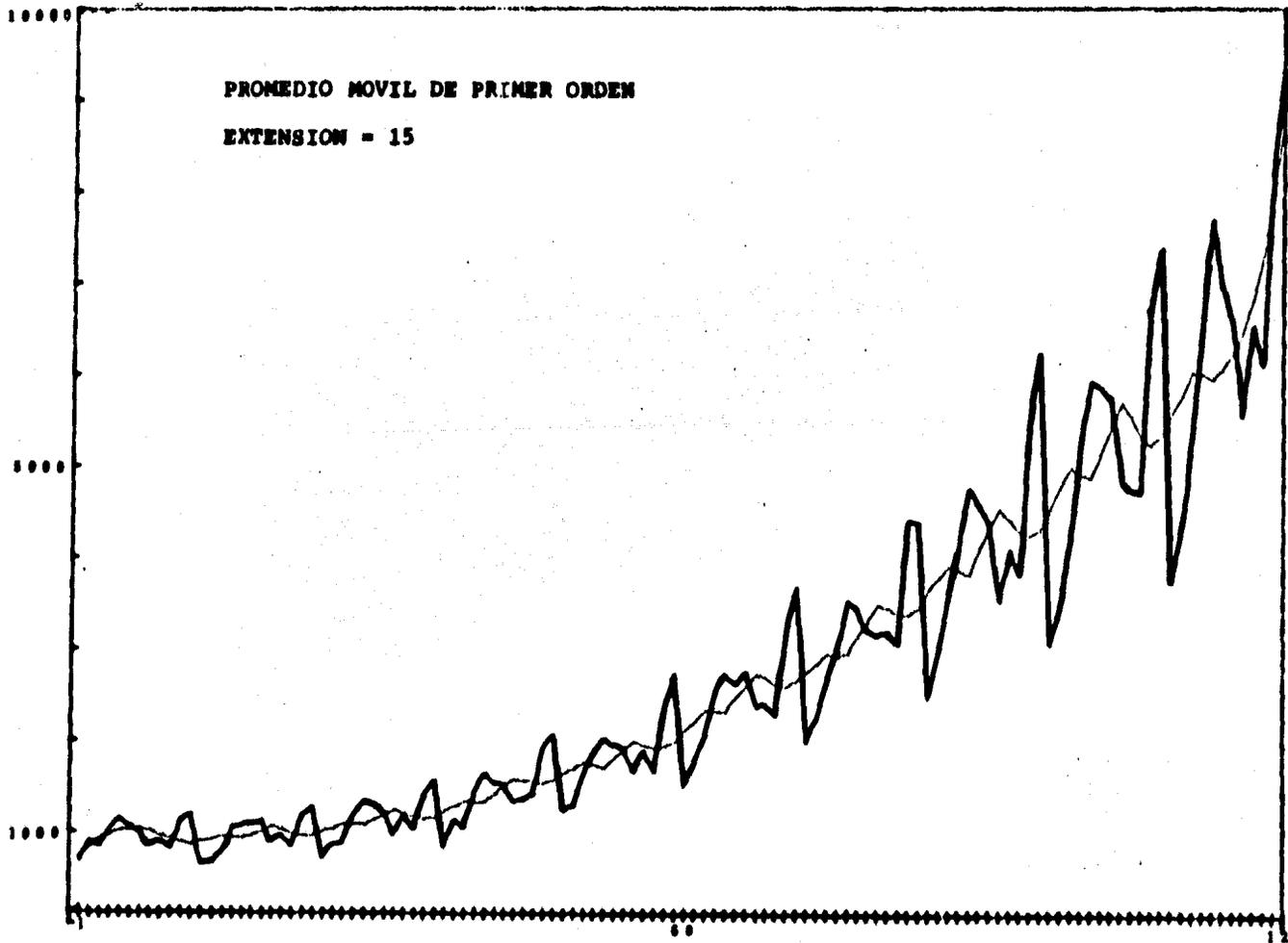
EXTENSION	ERROR CUADRATICO TOTAL	PORCENTAJE DE REDUCCION
5	523.6	-
7	425.2	18.8
9	343.4	19.2
11	302.7	11.9
13	270.9	10.5
15	257.9	4.8
17	242.4	6.0
19	236.9	2.3
21	236.2	0.3

NOTA: Las cifras de todas las tablas se redondearon a un decimal. (ver gráficas 8,9,10).

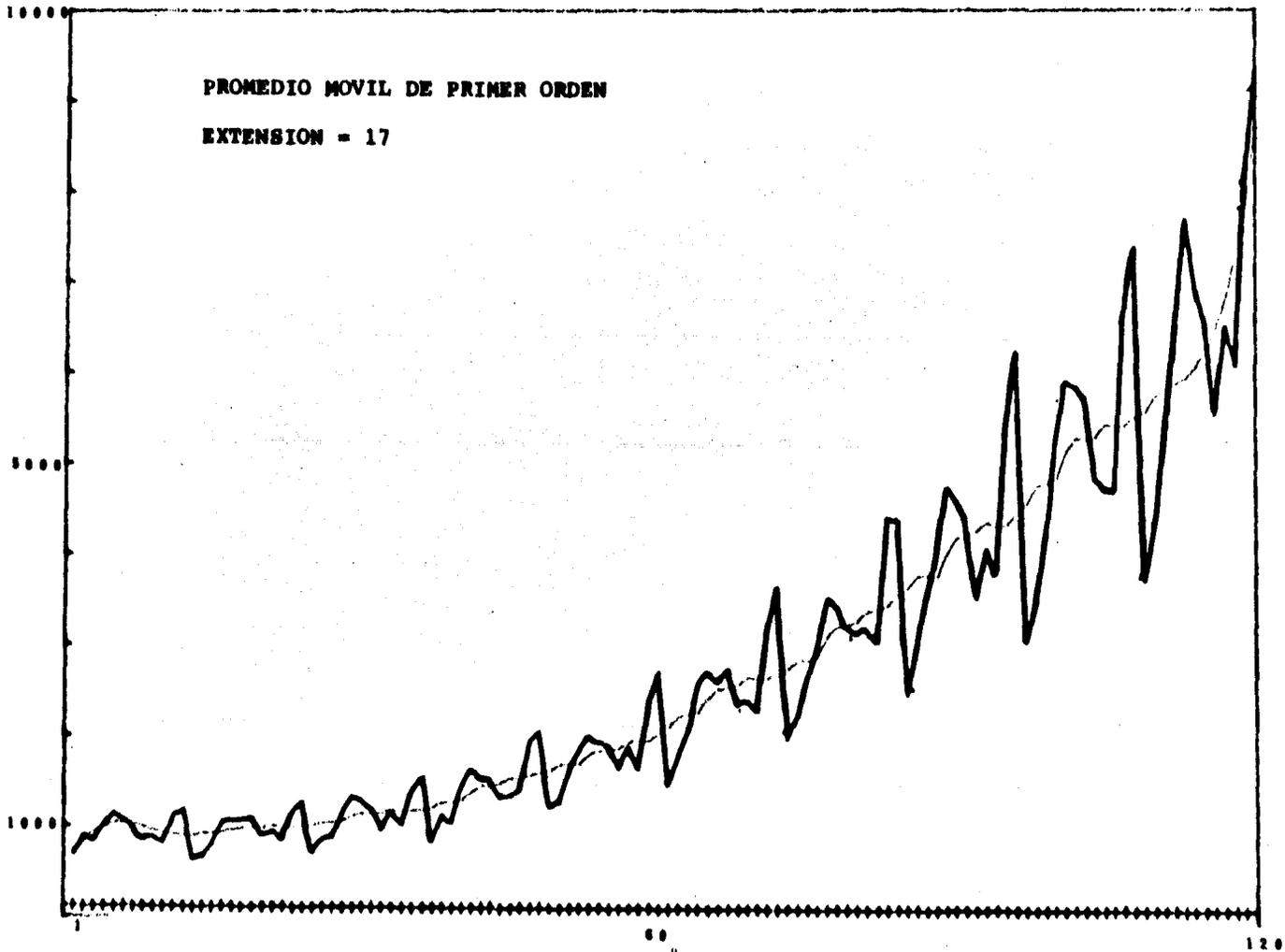
Observando la tabla anterior nos daremos cuenta -- que si bien, la extensión 17 no es la óptima en en el-



GRAFICA (B)



GRAPIC (9)



GRAFICA (10)

sentido de reducción máxima de error cuadrático total, si logra un porcentaje bueno y significativo, puesto -- que las extensiones 19 y 21 disminuyen un poco más el error cuadrático total pero en forma no relevante.

Los porcentajes que muestra la tabla se obtuvieron en forma eslabonada. Para dar una idea del porcentaje de reducción lograda en las extensiones comprendidas de 5 a 21 términos diremos que esta fue de:

$$\left[1 - \frac{236.2}{523.6} \right] * 100 = 54.9$$

Es posible que una extensión 17 sea la adecuada para nuestro trabajo. Un segundo análisis ilustrado en la tabla 5 entre las extensiones 15 y 21 nos permitirá tomar la decisión sobre esta afirmación.

Es importante recordar que al aplicar el procedimiento de promedio móvil, es distinta la forma en que se calculan las estimaciones de los valores de tendencia de la parte central de la serie al de los extremos de la misma.

Sabemos que es menos confiable la estimación de estos valores para los extremos lo cual nos lleva a considerar el error cuadrático inicial (E.C.I) para los primeros doce meses, un error cuadrático central (E.C.C.) para los siguientes noventa y seis y un error cuadrático final (E.C.F) para los últimos doce meses. La elección de doce y no

venta y seis meses es debido a que, como veremos posteriormente, en el spline promedio hay un desplazamiento que se corrige hasta el doceavo término como se puede apreciar en las gráficas 19, 20 y 21.

TABLA (5)

EXTENSION	E.C.I.	E.C.C.	E.C.F.	E.C.T.
15	100.3	124.9	728.1	257.9
17	107.5	83.3	721.6	242.4
% DE REDUCCION	-7.2	33.3	0.9	36.0
15	100.3	124.9	728.1	257.9
19	97.3	90.7	698.8	236.2
% DE REDUCCION	3.0	27.4	4.0	8.2
15	100.3	124.9	728.1	257.9
21	78.0	99.3	698.8	236.2
% DE REDUCCION	22.2	28.5	4.0	8.4

La extensión 15 se usó como base de comparación.

b) Promedio móvil de segundo orden.

De acuerdo con lo descrito en la sección 1.4, podría suponerse que un promedio móvil, de segundo orden lograría una mayor suavidad para la parte central de la serie y con ello una mejor estimación de la tendencia, --

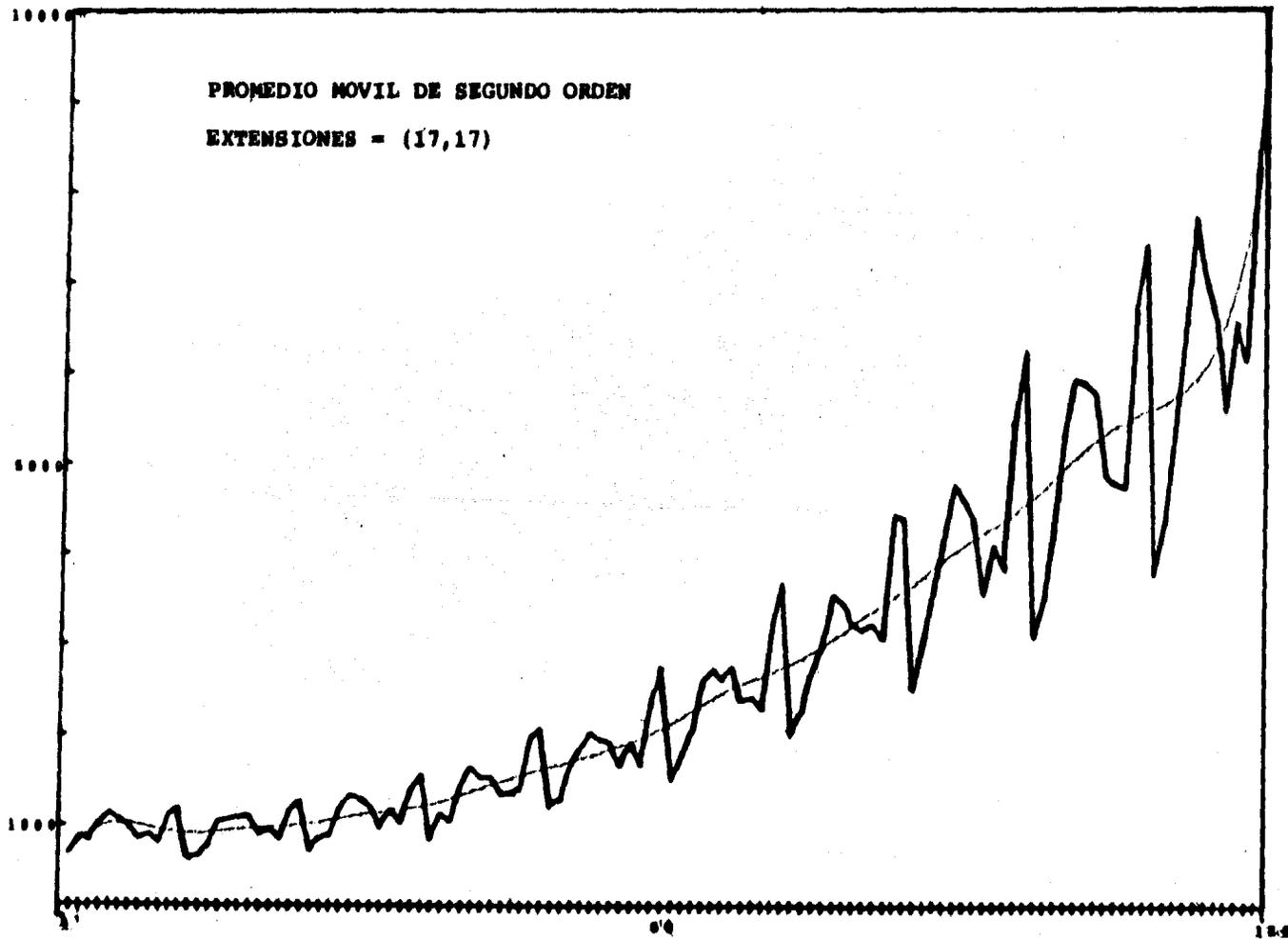
con la advertencia de que el problema de precisión en los extremos probablemente se agravaría.

Esta segunda etapa de pruebas se efectuará de la extensión 15 a la 21, para verificar si efectivamente se llega a una reducción significativa del error cuadrático medio, tanto en la parte central como en los extremos. Los resultados se resumen en la tabla 6.

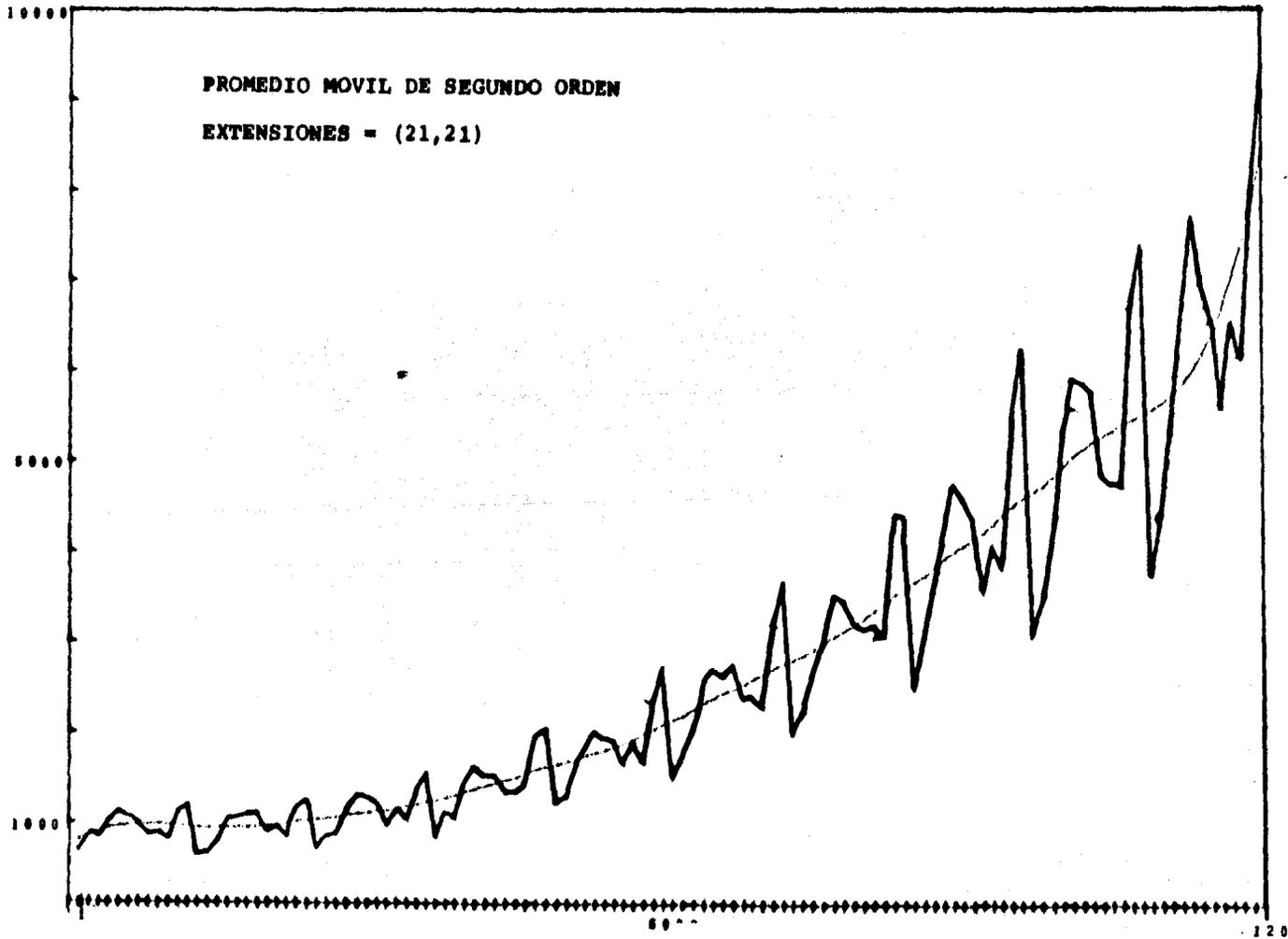
TABLA (6)

EXTENSION	E.C.I.	E.C.C.	E.C.F.	E.C.T.
15	100.3	124.9	728.1	257.9
(17,17)	102.1	65.4	716.6	236.2
% DE REDUCCION	-1.8	47.6	1.6	8.4
15	100.3	124.9	728.1	257.9
(19,19)	97.3	90.6	697.0	236.2
% DE REDUCCION	3.0	27.4	4.3	8.2
15	100.3	124.9	728.1	257.9
(21,21)	74.3	59.6	689.7	225.7
% DE REDUCCION	25.9	52.6	5.3	12.5

(Ver gráfica 11 y 12)



GRAFICA (11)



GRAFICA (12)

En la tabla 6 la extensión 15 de primer orden es la base de comparación y la extensión 21 fue la mejor. Al comparar los resultados de las tablas 5 y 6 se aprecia claramente que la extensión 21 para un promedio móvil de segundo orden es la mejor.

c) Promedio Móvil Mixto.

Le llamaremos así a un promedio móvil doble en donde se han elegido extensiones diferentes, es decir, a la serie original se le aplicó un promedio móvil con extensiones 17 y luego a la serie resultante se le aplicó uno de 15. A continuación se invirtió el orden de las extensiones para ver qué efectos causaba este cambio.

TABLA (7)

EXTENSION	E.C.I.	E.C.C.	E.C.F.	E.C.T.
(15,15)	98.3	74.2	720.6	239.4
(15,17)	93.6	68.0	711.2	234.8
% DE REDUCCION	4.9	8.4	1.3	1.5
(17,15)	105.9	67.9	720.3	238.0
% DE REDUCCION	-7.7	8.5	0.1	0.6
(17,17)	102.1	65.4	716.6	236.2
(17,15)	105.9	67.9	720.0	238.6
% DE REDUCCION	-3.7	-3.8	-0.5	-0.8
(15,17)	93.5	68.0	711.2	234.8
% DE REDUCCION	8.4	-3.9	0.8	0.6

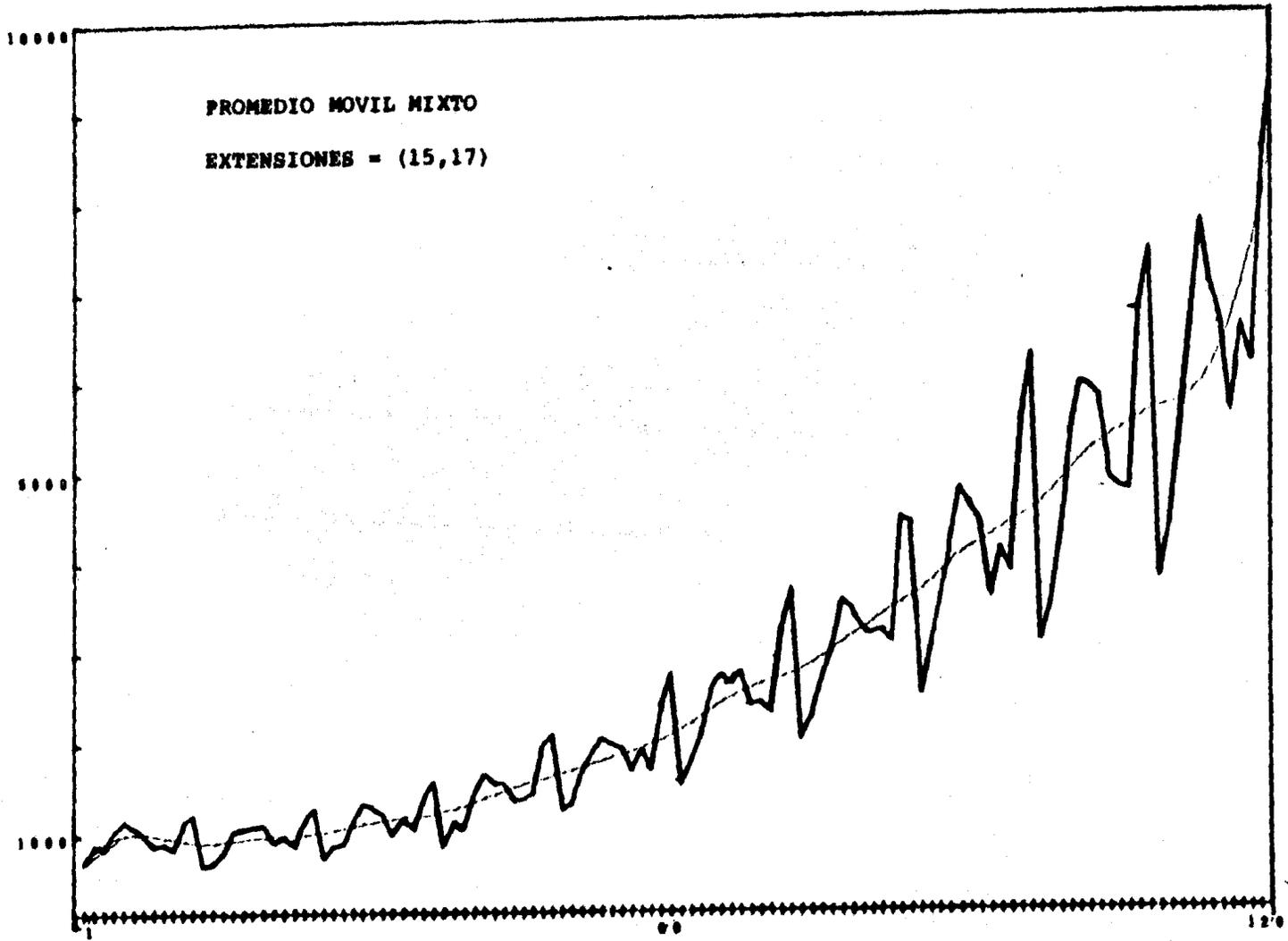
(Ver gráfica 13)

Note que en esta última tabla, se tomaron dos bases - de comparación, la extensión (15,15) y (17,17), para evi- tar preferencias con los promedios en discusión.

De la tabla 7, se desprende la recomendación de que - en un promedio móvil de segundo orden, con extensiones - distintas, es preferible aplicar primero la extensión me- nor, pues como se puede observar en dicha tabla, el E.C. C. no experimenta una reducción significativa, pero en - cambio, esta reducción se hace más palpable en el error- cuadrático para ambos extremos de la serie. Esto podría explicarse en la práctica diciendo que en algunas series conviene aplicar extensiones diferentes, dado que una ex- tensión modelará cierto tipo de variaciones y la otra -- las restantes.

d) Promedio Móvil Centrado.

J. Durbin considera innecesario emplear un promedio - móvil centrado, sin embargo, M. Kendall dice que es reco- mendable "limpiar" la serie antes de trabajar con ella, - es decir, lo usa como filtro. En particular nosotros se- guiremos el argumento de Kendall, para información men-- sual, aunque esto es completamente generalizable.



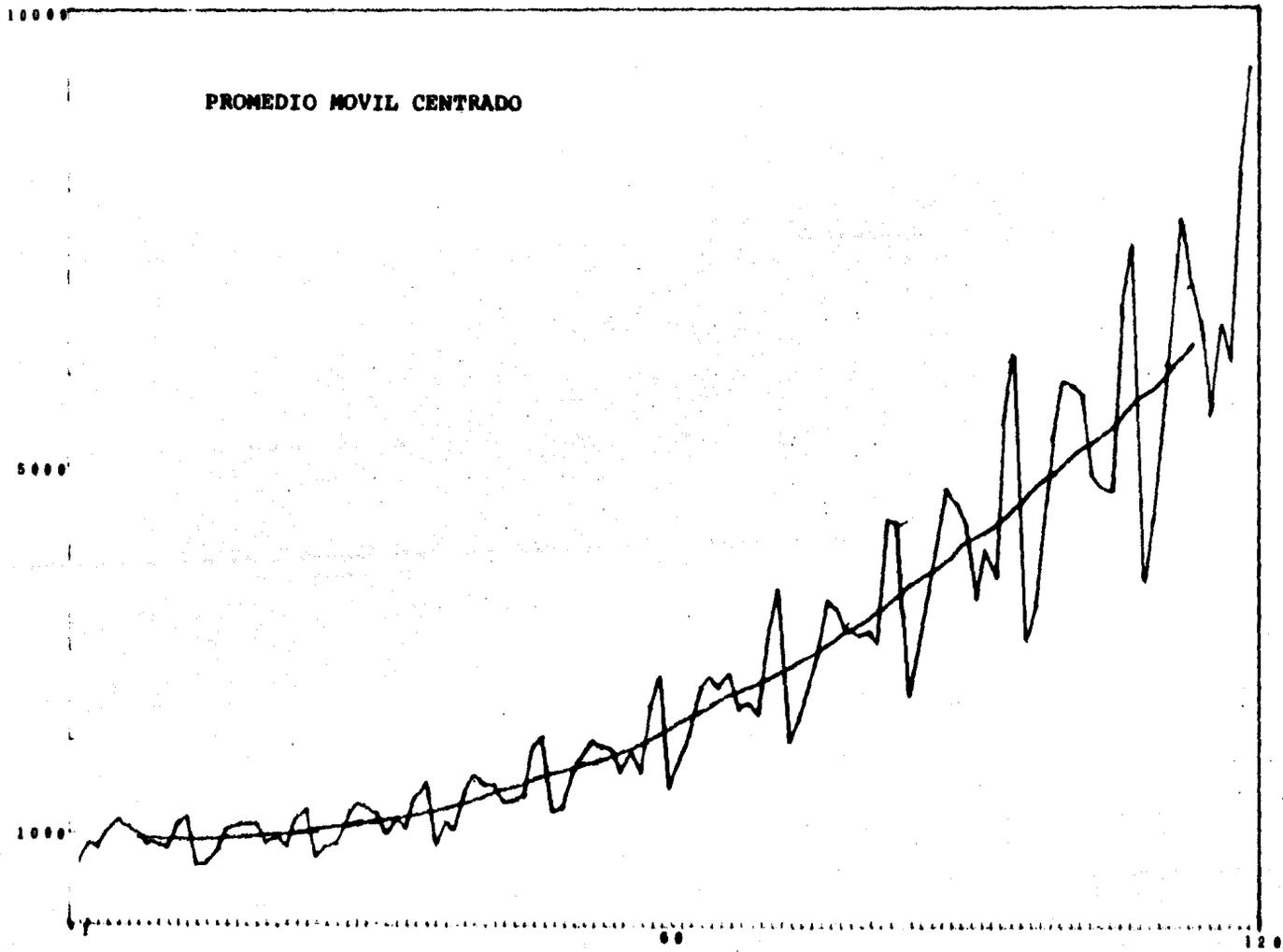
GRAFICA (13)

El promedio móvil centrado es un primer paso para empezar a desestacionalizar una serie de tiempo. Al aplicar este promedio se perdían los seis primeros y seis - últimos valores, por lo que se recurrió a completar la serie mediante el procedimiento de los valores en los - extremos con una extensión igual a 13. Posteriormente a la serie resultante se le aplicó una extensión de 17 y 21. Los resultados se muestran en la tabla 8.

TABLA (8)

	E.C.I.	E.C.C.	E.C.F.	E.C.T.
PROMEDIO MOVIL CENTRADO	92.1	66.6	724.0	238.4
EXTENSION 17	86.2	64.8	711.3	233.9
EXTENSION 21	80.4	63.6	698.9	229.6

(Ver gráfica 14)



GRAFICA (14)

3.6 EVALUACION DEL SPLINE DE APROXIMACION

La función spline presentado en el artículo "Smoothing by Spline Functions" por C.H. Reinsch, es un algoritmo que se trabaja a base de prueba y error, motivo por el cual tuvimos que realizar distintas pruebas con el propósito de obtener una buena estimación de la tendencia de la serie de tiempo.

En esta sección nos ocuparemos de efectuar la estimación de la tendencia, mediante el algoritmo de Reinsch en el cual emplearemos los siguientes parámetros:

- a) El coeficiente de control de suavidad, que satis-

face la restricción (2.3.2). En el caso particular de $S=0$ la función spline llevará a cabo una interpolación cúbica de los datos (120 en nuestro problema) cuando a S se le asigna un valor muy grande, el spline describe una línea recta. Como ya se explicó en 2.3.3, el intervalo de confianza en el que está comprendido el valor de S es:

$$104.5 \leq S \leq 135.5$$

si las ΔY_k para $k= 1,2,\dots,120$ se toman como las estimaciones de los errores de las desviaciones estándar de los datos Y_k .

b) Los otros parámetros que hay que dar al algoritmo son los $\Delta Y_k > 0$ ($k= 1,2,\dots,120$), que representan los pesos que afectan a cada dato de la serie. Los pesos influyen fuertemente en el spline y hay que tener mucho cuidado al asignarles un cierto valor. Un primer valor que podemos dar si no conocemos bien la información es la unidad, en el caso en que se conozca qué datos son más importantes que otros, entonces sería conveniente dar a éstos un peso mayor a la unidad, y si estos datos no son tan significativos o se sospecha que están influidos por un error grande, se les dará un peso menor a uno. La asignación de los pesos ΔY_k ($k= 1,2,\dots,120$) si no se cuenta con la desviación estándar de cada dato, dependerá mucho del criterio personal y experiencia, así como del problema que se tenga.

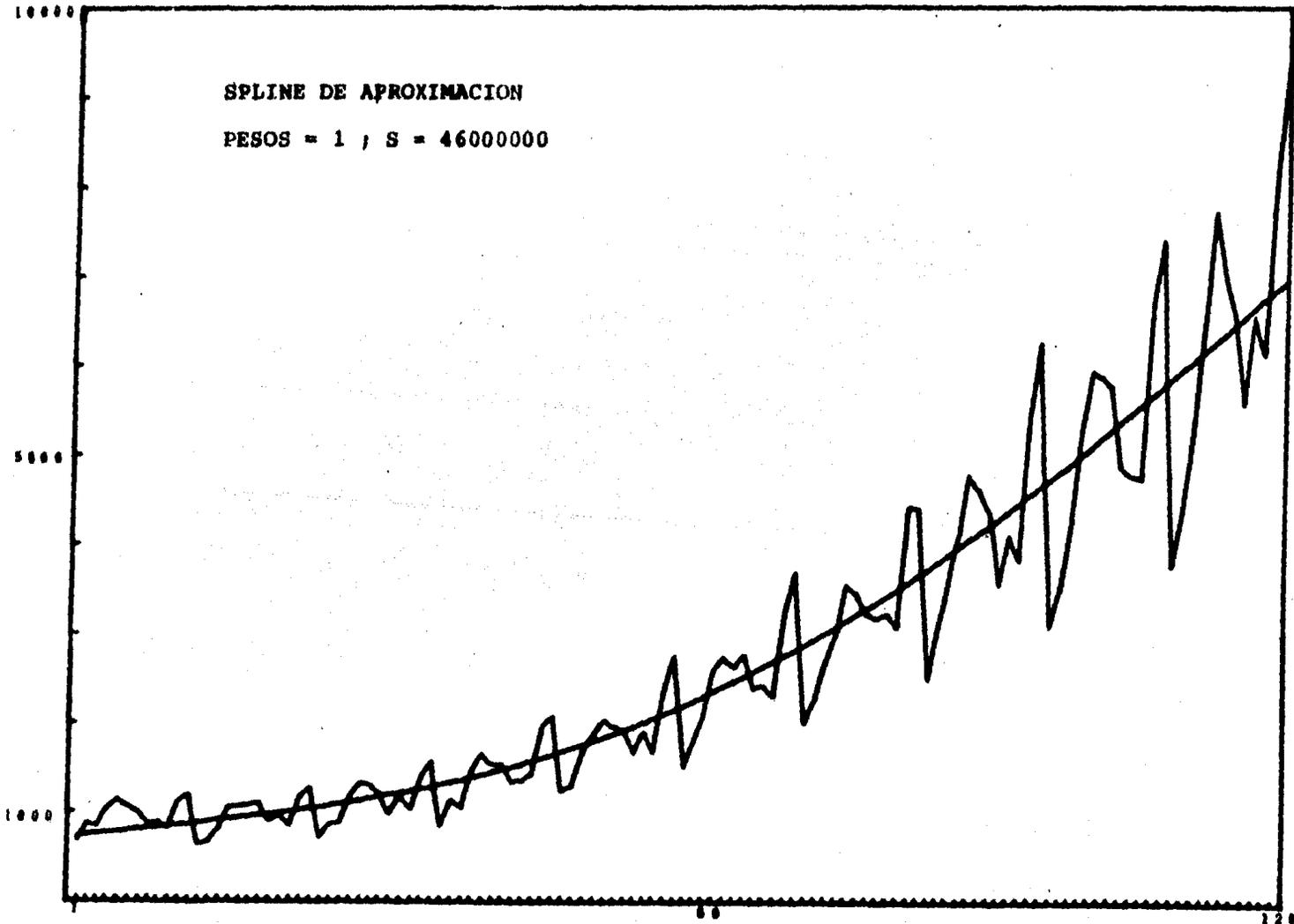
En las pruebas que se llevaron a cabo cuando se dieron pesos $\Delta Y_k = 1$ ($k = 1, 2, \dots, 120$) y valores de $S = 120$ y $S = 1000$, se observó que se obtenía una función spline muy parecida a la serie de tiempo; sin embargo, se logró una buena aproximación a la tendencia cuando se dió valores muy grandes para S . En la tabla 9 se presentan los resultados obtenidos.

TABLA (9)

SUAVIDAD	PESOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
1 000 000	1	145.8	489.4	1211.1	583.5
40 000 000	1	58.3	64.0	606.0	200.9
60 000 000	1	588.0	263.9	270.9	307.0
50 000 000	1	326.0	160.1	57.1	177.4
45 000 000	1	100.0	87.9	308.6	161.9
47 000 000	1	213.2	119.1	138.0	133.4
46 000 000	1	161.4	104.4	205.9	124.8

(ver gráfica 15)

Cabe mencionar que al hacer las pruebas se seleccionó una S de doscientos millones, aquí se obtuvo una función lineal que representa la suavidad máxima, entonces se redujeron S a cuarenta millones y se prosiguió en el orden que muestra la tabla 9, en el procedimiento des--



GRAFICA (15)

crítico se hicieron pruebas intermedias para asegurar que el camino que se estaba siguiendo fuera el correcto, no obstante, no se mencionaron para evitar redundancia. - Un inconveniente que se tiene al seleccionar una S muy grande es que el algoritmo requiere de mucho tiempo de procesamiento. Obsérvese que el mejor resultado para el E.C.T. se obtuvo en $S = 46$ millones. En la tabla 10 se presenta un ejemplo de la función spline mientras que para el E.C.C. fue en $S = 43$ millones.

De acuerdo a la sugerencia que da Reinsch de tomar para cada ΔY_k la desviación estándar del error del punto Y_k , y como en las series de tiempo por lo general la componente que más dificulta la estimación de la tendencia es la estacionalidad*, entonces se tomó ésta como el error y se restó de su media para dar el peso del punto respectivo.

La primera prueba se hizo exactamente como lo indica el artículo presentado por Reinsch, tomando:

$$\Delta Y_k^{-2} = (\eta_k - 1)\epsilon_k - \text{media}$$

$$\text{media} = \frac{\sum_{k=1}^{120} (\eta_k - 1)\epsilon_k}{120}$$

con $K = 1, 2, \dots, 120$ y $S = 135$

*Hay que recordar que como simulamos la serie de tiempo conocemos el valor de la componente estacional en cada punto.

ECM INICIAL 161.4450369
 ECM CENTRAL 104.3813567
 ECM FINAL 205.9403412
 ECM TOTAL 124.7556271

TABLA (10)

A	B	C	D
732.6228564	12.16970023	0	-1.11172E-04
744.7919455	12.16936671	-3.33515E-04	2.52571E-04
756.9612313	12.16945739	4.24197E-04	5.01492E-04
769.1316143	12.17181026	1.92867E-03	1.21901E-03
781.3065723	12.17932464	5.58572E-03	2.19732E-03
793.4936800	12.19708804	0.012177683	2.96868E-03
805.7059144	12.2303494	0.021003716	3.49840E-03
817.9608459	12.28301203	0.031578925	3.62382E-03
830.2790607	12.35704124	0.042450384	3.79587E-03
842.6823482	12.45332962	0.053837994	3.75054E-03
855.1932663	12.5722571	0.065089603	4.46196E-03
867.8350750	12.71582188	0.078475477	5.37508E-03
880.6347474	12.88889837	0.094600720	4.74115E-03
893.6229877	13.09232304	0.108824182	4.10356E-03
906.8282384	13.32228168	0.121134866	3.80430E-03
920.2754593	13.57596469	0.132547779	4.14950E-03
933.9881213	13.8535082	0.144996292	4.52010E-03
947.9911459	14.15706113	0.158556594	4.89301E-03
962.3116566	14.48885321	0.173235613	5.27772E-03
976.9790231	14.85115698	0.189068765	5.07570E-03
992.0243246	15.24452159	0.204295853	4.97476E-03
1007.478117	15.66803692	0.219220136	4.55578E-03
1023.369930	16.12014431	0.232887490	4.94889E-03
1039.72791	16.60076563	0.247734158	5.59609E-03
1056.582006	17.11302167	0.264522440	4.68613E-03
1073.964236	17.6561251	0.278580822	4.09956E-03
1091.903042	18.22558496	0.290879510	3.52345E-03
1110.423030	18.81791412	0.301449871	3.69924E-03
1129.546093	19.43191122	0.312547603	4.24411E-03
1149.294796	20.06973846	0.325279930	4.66231E-03
1169.694477	20.73428447	0.339266866	4.78398E-03
1190.772812	21.4271699	0.353618806	4.21105E-03
1212.557812	22.14703984	0.366251953	4.09244E-03
1235.075196	22.89181997	0.378529286	3.55489E-03
1258.3491	23.65954246	0.389193951	3.89800E-03
1282.40735	24.4496242	0.400887963	4.67767E-03
1307.256924	25.26543317	0.414920978	3.40845E-03
1332.940687	26.10549985	0.43146423	2.84742E-03
1359.47418	26.96433427	0.433688694	2.00462E-03
1386.874208	27.83772494	0.439702566	2.18150E-03
1415.153817	28.72367347	0.446247066	2.81021E-03

1505.419405	31.48086363	0.473666358	2.79078E-03
1537.376725	32.43656827	0.482033686	2.29505E-03
1570.297627	33.40753004	0.489233836	1.87700E-03
1604.195158	34.39100753	0.494554826	2.86634E-03
1639.084387	35.38871632	0.5003153832	4.05708E-03
1674.980314	36.4071944	0.515325081	2.83203E-03
1711.905672	37.44635762	0.523889165	1.66534E-03
1749.877534	38.49903184	0.528835175	1.41853E-03
1788.906819	39.56095634	0.533090757	1.61251E-03
1829.002479	40.63197669	0.537928287	2.14982E-03
1870.174534	41.71427834	0.544377759	2.38676E-03
1912.435577	42.81019520	0.551538030	2.40539E-03
1955.799715	43.92048792	0.558754202	1.60992E-03
2000.280567	45.04282542	0.563583972	1.31383E-03
2045.88829	46.17393353	0.567525471	2.73767E-04
2092.630023	47.30980533	0.568346772	9.60637E-04
2140.509136	48.4493805	0.571228684	2.61947E-03
2189.532365	49.59969549	0.579087102	7.32351E-04
2239.711880	50.76006535	0.581284154	-5.38614E-04
2291.05269	51.92101658	0.579668312	-1.17625E-03
2343.552199	53.07682458	0.576139554	-5.44855E-04
2397.204618	54.22746732	0.574504987	3.58896E-04
2452.006950	55.37755282	0.575581676	0.38231E-04
2507.960922	56.53123415	0.578096370	1.51620E-03
2565.071769	57.69197294	0.582644959	1.00973E-03
2623.347397	58.8602903	0.585674138	4.09750E-04
2682.79377	60.03286900	0.586903389	-6.68166E-04
2743.412875	61.20466939	0.584898892	5.79202E-04
2805.203023	62.37620473	0.586636499	2.99198E-03
2868.168856	63.55845196	0.595612427	5.90859E-04
2932.323511	64.75144744	0.597385005	-1.31987E-03
2997.671024	65.94226137	0.593425408	-2.20399E-03
3064.204506	67.12249638	0.586813445	-2.43142E-03
3131.911385	68.28882832	0.579519171	-1.30914E-03
3200.778423	69.44393770	0.575591745	-6.15352E-04
3270.797337	70.59327689	0.573745689	-7.71911E-04
3341.963588	71.73844758	0.571429957	-1.26118E-03
3414.272204	72.87752444	0.567646423	-1.83174E-03
3487.715543	74.00731933	0.562151214	-2.99899E-03
3562.282015	75.12262597	0.553154245	-6.24242E-04
3637.95717	76.22705838	0.551281520	1.47968E-03
3714.73699	77.33406078	0.555720565	-1.94657E-03
3792.624825	78.43966128	0.549880868	-4.08545E-03

4032.775794	81.61922734	0.508950380	-2.57550E-03
4114.901396	82.62940060	0.501203682	-1.21357E-03
4198.030807	83.62820650	0.497583180	-7.72638E-04
4282.155825	84.62105460	0.495265266	-2.81083E-03
4367.269334	85.60315253	0.486832800	-3.59486E-03
4453.355724	86.56603286	0.476048212	-5.35761E-03
4540.392448	87.50205605	0.459975373	-2.87023E-03
4628.351609	88.41399585	0.451364681	-1.62151E-03
4717.21739	89.32090935	0.456229207	-2.46400E-03
4806.992245	90.22455567	0.447082217	-6.63864E-03
4897.657500	91.09931394	0.427421300	-8.50496E-03
4989.175730	91.92864039	0.401296434	-7.64294E-03
5081.499634	92.70952434	0.378977607	-5.30943E-03
5174.581826	93.45155090	0.363049314	-3.38216E-03
5268.393044	94.1675021	0.352902632	-2.00379E-03
5362.911445	94.86729569	0.346391454	-3.41985E-03
5458.122213	95.55081816	0.336631910	-5.37472E-03
5554.004298	96.20785783	0.320507752	-7.68256E-03
5650.525071	96.82592437	0.317460062	-4.77446E-03
5747.643681	97.40652118	0.283196679	-1.95844E-04
5845.333143	97.97220669	0.282549147	-6.06887E-03
5943.58183	98.51909822	0.264742545	-0.010516139
6042.354755	99.01623502	0.232794127	-0.012640762
6141.591143	99.44390035	0.194871840	-0.011824292
6241.21809	99.79817122	0.159398963	-7.74126E-03
6341.167920	100.0937448	0.136175187	-5.94321E-03
6441.391897	100.3482655	0.118345551	-5.57317E-03
6541.852935	100.5682368	0.101626638	-8.29093E-03
6642.514507	100.7466165	0.076753560	-8.58434E-03
6743.322292	100.8743705	0.051000545	-0.010326520
6844.244337	100.9453920	0.020020986	-6.67366E-03
6945.203076	100.9453920	0	-6.67366E-03

el límite superior del intervalo de confianza, lo cual nos dió como resultado casi la interpolación. Posteriormente se probó con los mismos valores para ΔY_k , pero con $S = 420$ (3.5 veces más el punto medio (120) del intervalo de confianza) y siguió dando casi la interpolación, no se continuó puesto que esta última S ya estaba muy alejada del intervalo propuesto.

Después se tomó:

$$(\Delta Y_k)^2 = ((n_k - 1)\xi_k - \text{media})^2.$$

y con una $S = 420$ dando como resultado una recta. Como la recta representa el suavizamiento máximo entonces se disminuyó la S consecutivamente en 200 y 120 volviendo a dar una función spline lineal.

Finalmente se probó:

$$\Delta Y_k = \text{abs}((n_k - 1)\xi_k - \text{media})$$

y $S = 135$, fué aquí donde se empezó a lograr una buena aproximación a la tendencia. En la tabla 11 están los resultados de las pruebas realizadas en el orden efectuado.

TABLA (11)

SUAVIDAD	E.C.I.	E.C.C.	E.C.F.	E.C.T.
135	57.4	84.9	329.6	130.6
120	52.1	61.8	83.2	63.4
105	57.7	70.4	133.5	78.0

SUAVIDAD	E.C.I.	E.C.C.	E.C.F.	E.C.T.
115	48.0	63.6	107.0	67.9
122	54.8	61.4	68.8	61.6
118	49.9	62.4	94.2	65.2

(Ver gráficas 16, 17, 18)

El mejor resultado total se obtiene en $S = 122$ que es un valor cercano al punto medio del intervalo de confianza. En la tabla 12 se muestran los valores de los parámetros del spline con $S = 122$.

Otra prueba que se hizo, fue aplicar a la serie de tiempo un promedio móvil de extensión 17 y a éste resultado un spline, no condujo a resultados favorables (el E.C.T. = 242.3).

Spline Promedio.

El problema que se presenta al estimar la tendencia de una serie de tiempo, es el de la estacionalidad, ya que ésta es la componente que impide por lo regular, obtener una buena aproximación de la tendencia.

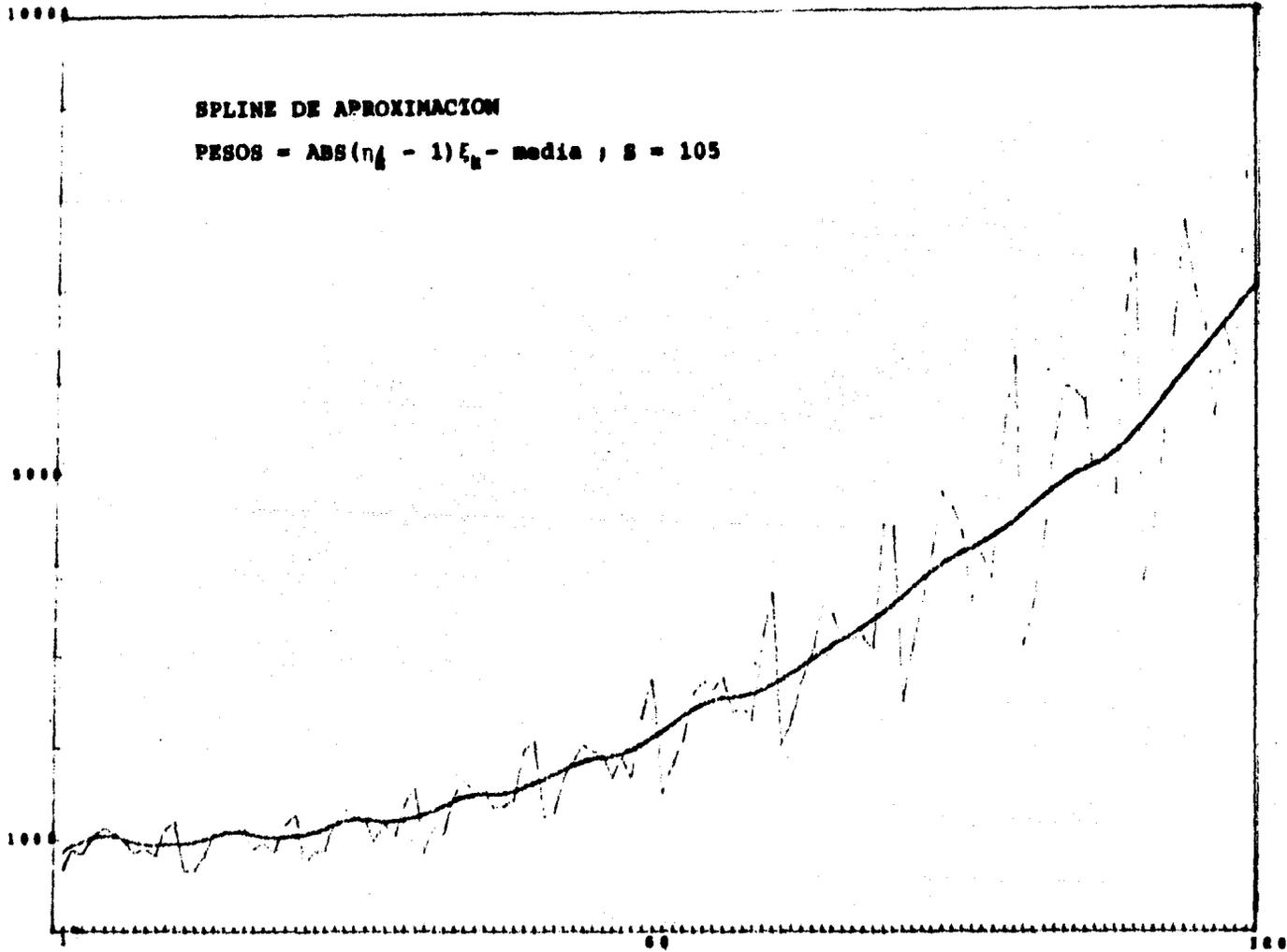
Tal dificultad se hizo patente al quererla estimar con la función spline de aproximación con pesos uno, -- pues si se recuerda cuando se usó un coeficiente de suavidad pequeño se obtenía una aproximación muy parecida a la serie de tiempo original, por lo cual había que au

A	B	C	D
1008.100750	3.62095843	0	-0.069704771
1011.652004	3.411844109	-0.209114313	-0.128782716
1014.72595	2.607267349	-0.595462461	-0.323050102
1016.414705	0.447192093	-1.564612785	-0.137440898
1015.159544	-3.095256166	-1.977835479	0.049765154
1010.136217	-6.901631659	-1.828540015	0.607059405
1002.013105	-8.737533469	-7.36180E-03	0.443756940
993.7119668	-7.420986361	1.333909018	0.294044757
987.9089343	-3.891033958	2.206043287	-0.060728934
986.1632147	0.338865795	2.033856485	-0.253842507
988.3720945	0.625051217	1.263333963	-0.042315090
993.1171596	6.02276387	1.135383692	0.074652867
1000.349960	0.517489037	1.359342294	-0.010274591
1010.216518	11.20535066	1.2251852	-0.170677153
1022.579710	13.35035627	0.816487062	-0.439058563
1036.307494	13.66615466	-0.500888643	-0.267824721
1049.305136	11.86130322	-1.304162805	-0.338957685
1059.523318	8.536104556	-2.021035859	-0.033903357
1066.004404	4.392322754	-2.12274593	0.690056347
1068.964117	2.216999954	-0.052576890	0.508253527
1071.636793	3.630606744	1.472183690	0.091996076
1076.837580	6.856962330	1.748171918	-0.156920808
1085.285793	9.882543745	1.277409494	-0.042464850
1096.403282	12.30996819	1.150014945	0.041738505
1109.905003	14.73521356	1.275230461	-0.038644204
1125.876803	17.16974186	1.13229785	-0.143376525
1144.062466	19.05820796	0.729168276	-0.41450768
1163.435335	19.27302149	-0.514354764	-0.287720202
1181.906282	17.38115137	-1.377515369	-0.122723773
1197.787194	14.2579493	-1.745686689	0.427279290
1210.726736	12.0484138	-0.463848820	0.452655523
1222.763956	12.47868273	0.894117750	0.239902508
1236.376659	14.98662575	1.613825273	0.028416574
1253.005527	18.29952603	1.299074996	-0.161003133
1272.843125	21.21466664	1.216065597	-0.066965504
1295.206891	23.4459013	1.015169085	0.012642792
1319.680605	25.51416783	1.05309746	-0.057411310
1346.190459	27.4481288	0.880863531	-0.118718119
1374.400733	28.85370151	0.524709174	-0.367931307
1403.411212	28.79932593	-0.579084748	-0.121959776
1431.509494	27.27527712	-0.944964077	0.011540606
1457.851347	25.41997078	-0.910342260	0.206205275
1482.56718	24.21790209	-0.291726436	0.422313723
1506.91567	24.9013904	0.975214734	0.317545973
1533.109821	27.60445779	1.927852653	-0.135613382
1562.706519	31.25332299	1.521012508	-0.225802535
1595.255052	33.61794037	0.843604903	-0.062154628
1629.654442	35.11868627	0.657141019	0.97075E-03
1665.43924	36.45988040	0.684053264	0.031546450
1702.551717	37.73361763	0.589683915	-0.107159868
1740.767859	38.59150586	0.268204311	-0.149729897
1779.477839	38.67872478	-0.180985379	0.020857726
1817.996436	38.3793272	-0.118412202	0.086331929
1856.343683	38.40149858	0.140563584	0.196664694
1895.08243	39.27265981	0.730577667	0.238929708
1935.324597	41.45060426	1.447366793	0.126596262
1978.349165	44.73512661	1.827155580	-0.043763894
2024.857663	48.24808609	1.695803098	-0.191468472
2074.610885	51.06528854	1.12139848	-0.130364810
2126.666407	52.91699086	0.730304050	-0.077164099
2180.238538	54.14610658	0.498811770	-0.115370360
2234.766086	54.79761906	0.152700689	-0.166724531

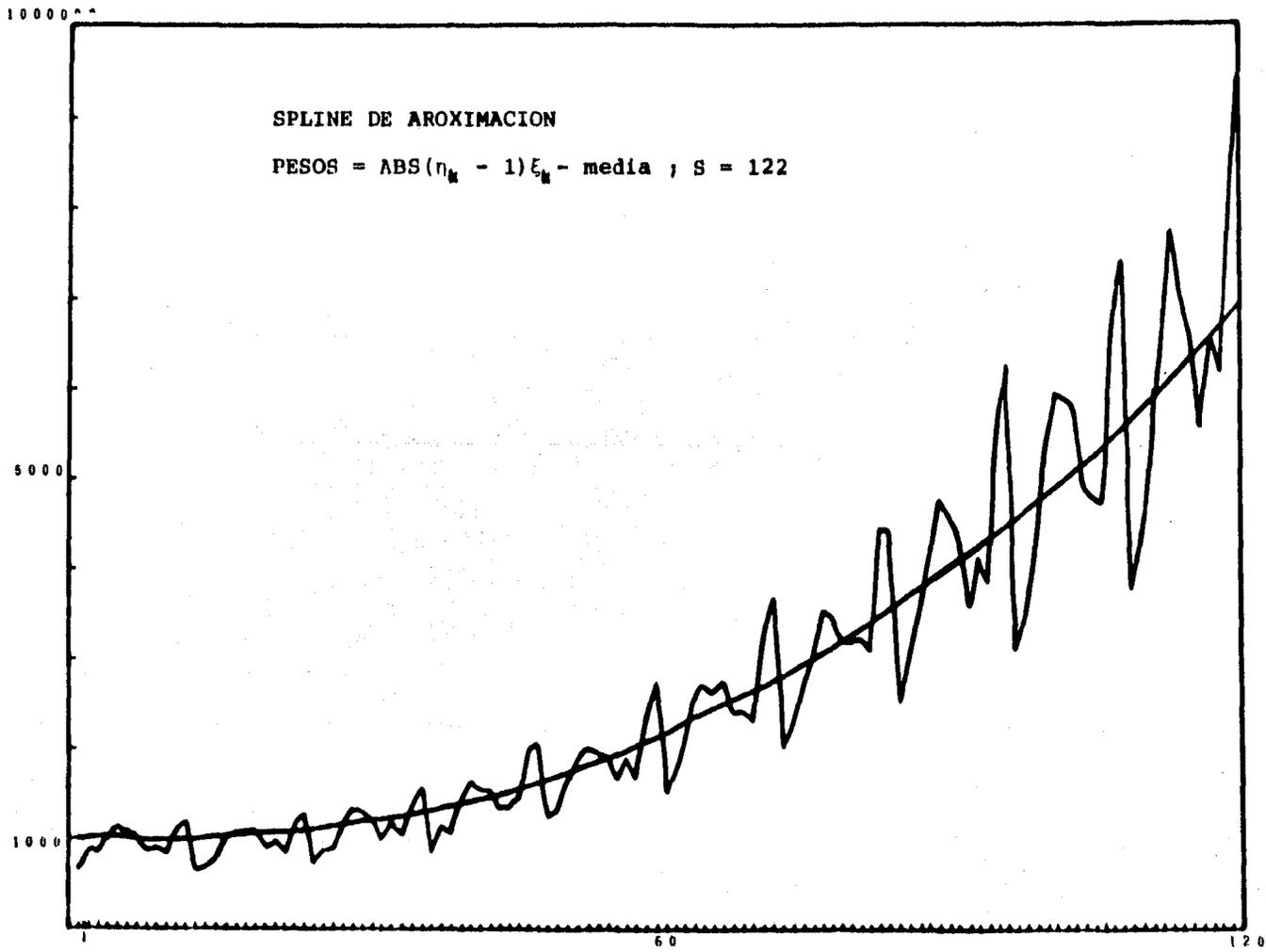
TABLA (12)

Sumidad: 122

2359.54468	54.60284682	-0.347472903	-0.246373681
2343.558681	53.16877395	-1.086593946	-0.024954564
2395.615913	50.92072837	-1.161457636	0.029771289
2445.404955	48.68712688	-1.07214383	0.126221360
2493.146160	46.92150517	-0.693477868	0.336625429
2539.710312	46.54442573	0.316398419	0.293763341
2586.865400	48.05851258	1.19708834	0.023513302
2636.145114	50.52442933	1.268238347	-0.068905873
2687.868866	52.85416855	1.061510727	-6.61774E-03
2741.777927	54.95733645	1.041657518	0.038312851
2797.815234	57.15538977	1.156596072	0.010858569
2856.138279	59.50135748	1.199171760	-0.032020967
2916.796787	61.78363818	1.093108677	-0.085584452
2979.587950	63.71310260	0.836355523	-0.083918390
3044.053489	65.13405838	0.584608351	-0.038518397
3109.733630	66.18770387	0.469045159	0.057953700
3176.448332	67.29965529	0.642966261	0.074103970
3244.464998	68.80777373	0.865218171	0.060319920
3314.193316	70.71917582	1.046177932	-0.055400400
3385.908269	72.64533042	0.879976734	-0.111385887
3459.32219	74.07112633	0.545819074	-0.044414981
3533.89472	75.02951946	0.412574131	-0.020313089
3609.316501	75.7937285	0.351624863	-0.044520126
3685.417344	76.36943774	0.218074484	-0.073805354
3761.925051	76.57817068	-3.34153E-03	-0.104048671
3838.395832	76.25934150	-0.315487590	-0.038717629
3914.300968	75.51221354	0.431140477	6.77088E-03
3989.389312	74.66924518	0.411137852	0.105171769
4063.751401	74.16210478	-0.095312544	0.173516894
4137.99121	74.49103029	0.424728137	0.105014094
4213.011993	75.6555488	0.739780420	2.57332E-03
4289.403895	77.14282957	0.747500375	-0.052135007
4367.24809	78.48142522	0.591095354	-7.43201E-03
4446.313179	79.64131958	0.568793334	0.023203481
4526.546501	80.84852765	0.638409777	3.16115E-03
4608.036600	82.13483049	0.647893231	-0.029011558
4690.790312	83.34358212	0.560858556	-0.074760894
4774.619992	84.24101652	0.336575874	7.81558E-03
4859.205400	84.93761510	0.360022625	0.042845336
4944.545883	85.78619633	0.488558633	0.131602612
5030.95224	87.1581216	0.883366468	0.233526407
5119.227255	89.62543336	1.583945689	0.212855896
5210.649490	93.43189272	2.222513377	-0.055017742
5306.248078	97.70946604	2.055060152	-0.101601581
5405.911003	101.5147822	1.750255409	-0.067668701
5509.108372	104.8122874	1.547249306	-0.047218911
5615.420689	107.7651283	1.405592573	-0.064303250
5724.527107	110.383402	1.212682823	-0.090698764
5836.032493	112.5366711	0.94058653	-0.128228678
5949.301522	114.093158	0.555900495	-0.080667252
6063.809913	114.9029574	0.313893738	-0.045834378
6179.060935	115.3932517	0.176395604	3.37329E-03
6294.633956	115.7561628	0.186515482	0.027912134
6410.604546	116.21293	0.270251883	-0.017660830
6527.070067	116.7004513	0.217269393	-0.012995305
6643.974792	117.0960023	0.178281679	-0.041731422
6761.207344	117.3273714	0.053087413	-0.017695804
6878.570197		0	



GRAFICA (16)



GRAFICA (17)

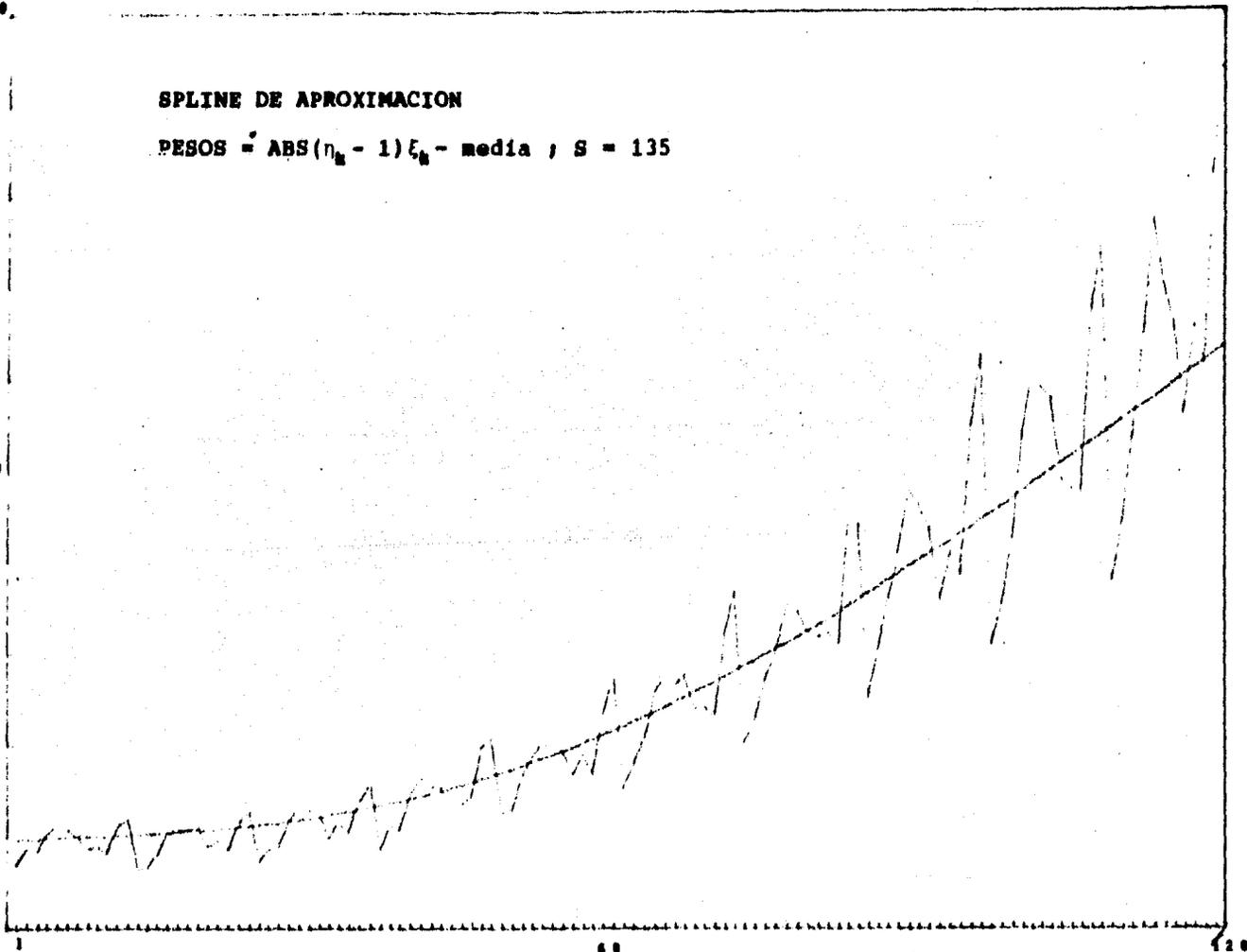
10000

SPLINE DE APROXIMACION

PESOS = $\sum \text{ABS}(\eta_k - 1)\xi_k - \text{media}$; S = 135

5000

1000



202

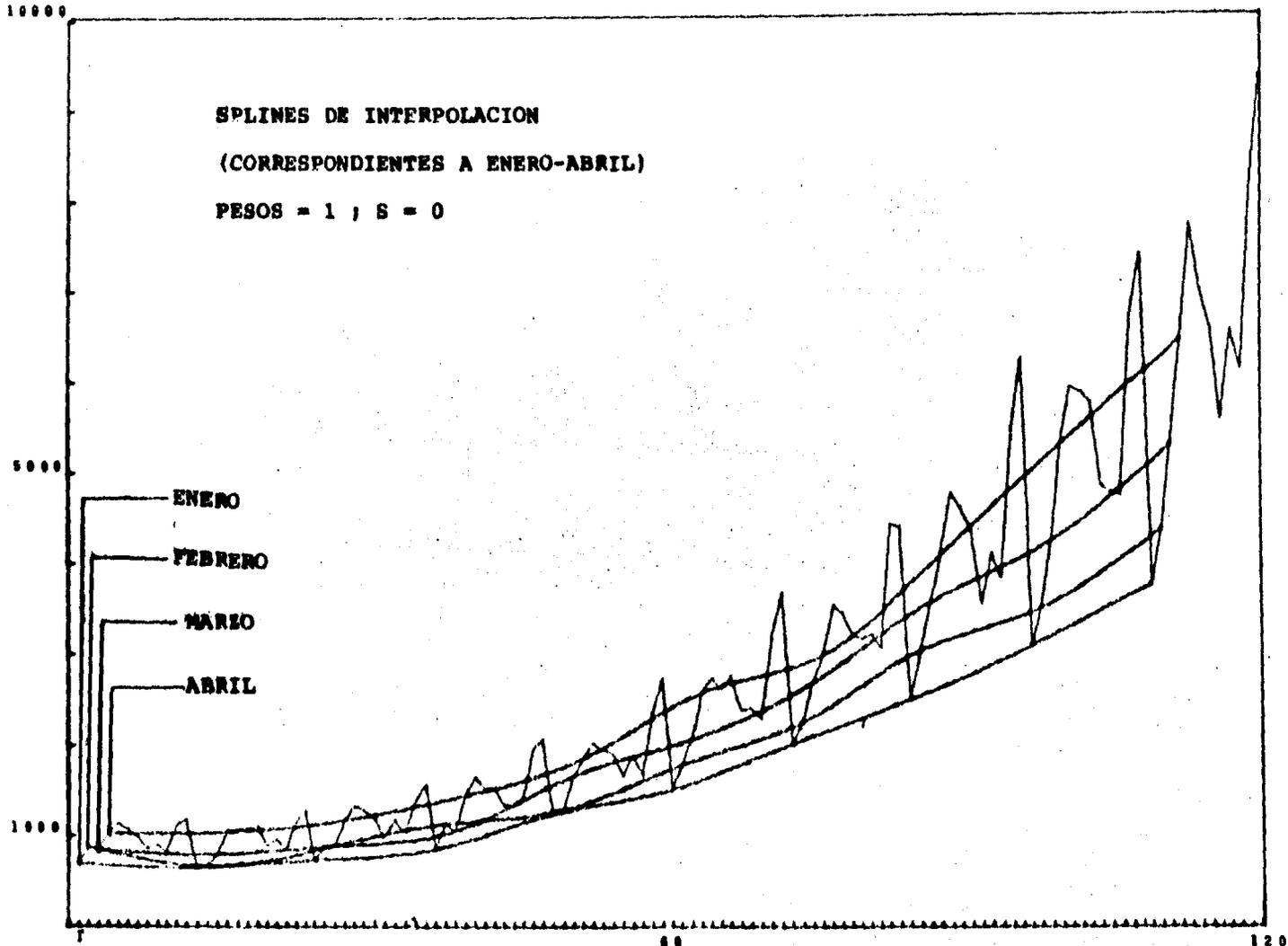
GRAFICA (10)

mentar fuertemente dicho coeficiente para lograr anular hasta cierto punto la estacionalidad. En cambio, en el promedio móvil se obtuvo mayor éxito en este aspecto, - pues modelaba a la serie con mayor facilidad.

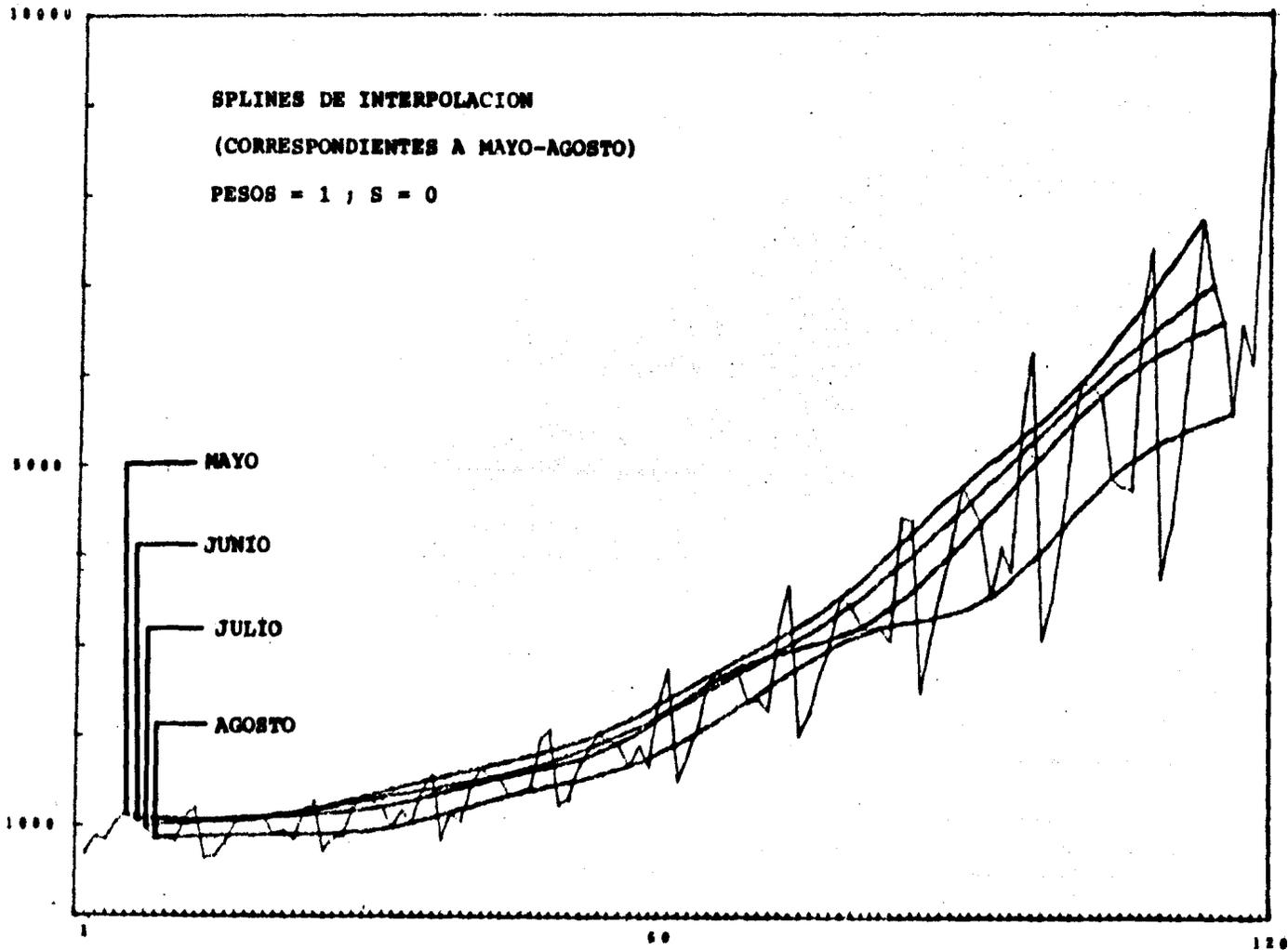
Como no se cuenta con un spline particular, que elimi-
ne la estacionalidad de una serie de tiempo enfrentare-
mos el problema en forma indirecta, es decir, modifican-
do el procedimiento del manejo de los datos en la forma
siguiente:

Se obtienen doce funciones spline de interpolación--
para cada conjunto de meses: enero, febrero,..., diciem-
bre. Con pesos igual a uno, para cada conjunto de me-
ses.

A lo largo de los diez años de información, se ob-
tendrá un spline en la subserie formada por los diez e-
neros, otros con los diez febreros y así sucesivamente-
hasta tener los doce splines como se muestra en las grá-
ficas 19, 20 y 21. Los doce splines se graficaron en-
grupos de 4 sólo por claridad. Como se puede apreciar-
en las gráficas, los splines sufren un desplazamiento -
hacia la derecha de acuerdo con el mes con que están in-
terpolando, lo que ocasiona nuevamente una pérdida de -
precisión en ambos extremos de la serie. Por el proble-
ma anterior, se consideró que se deberían separar los -



GRAFICA (19)



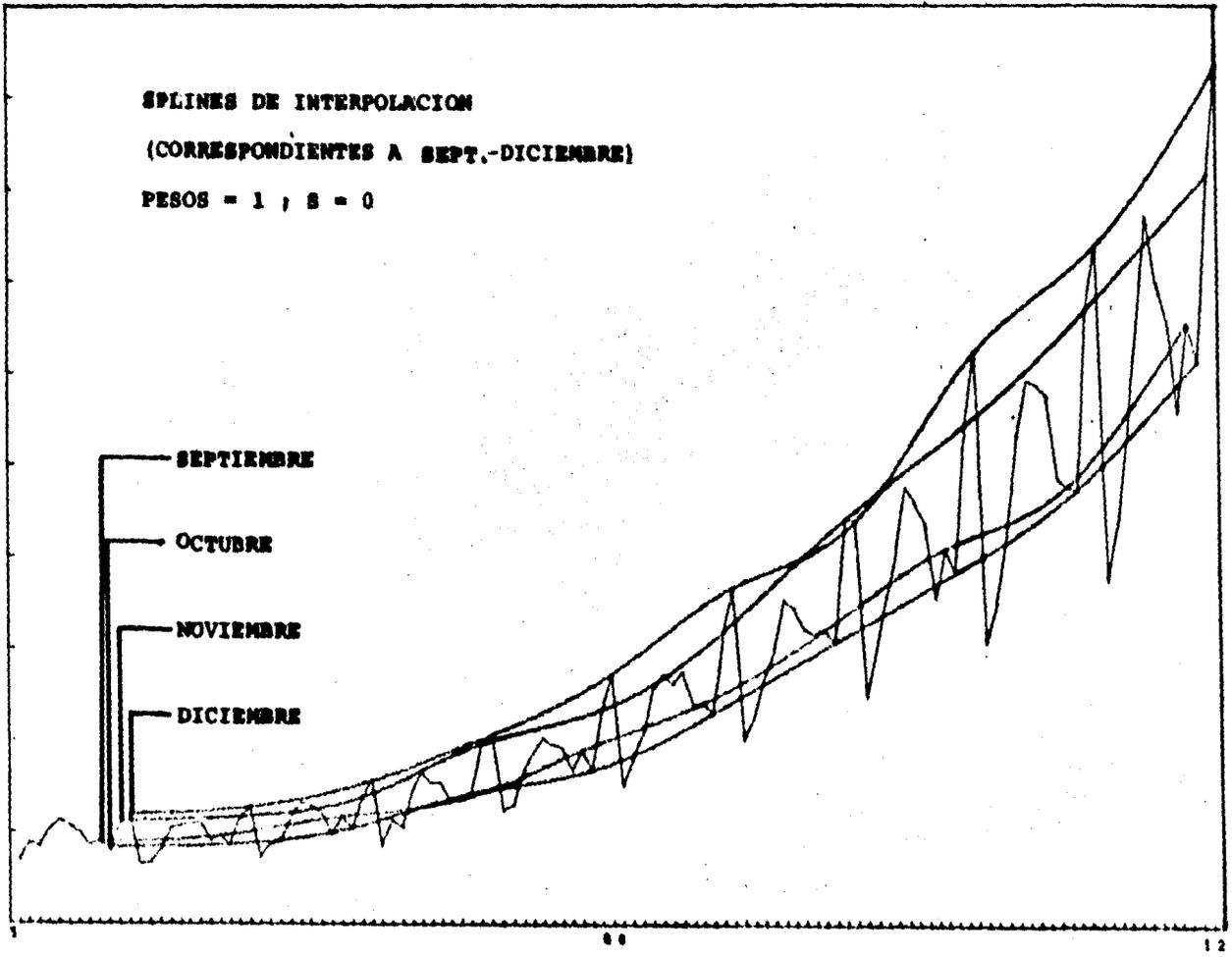
GRAFICA (20)

10000

SPLINES DE INTERPOLACION
(CORRESPONDIENTES A SEPT.-DICIEMBRE)
PESOS = 1 ; S = 0

5000

1000



00

120

GRAFICA (21)

96 términos centrales de la serie, de los 12 iniciales y 12 finales.

Los doce splines que se usaron son de interpolación, dado que si quisiéramos emplear splines de aproximación no contaríamos con un criterio para la mejor elección del coeficiente de suavidad S . Además por el teorema de Holladay (apéndice B) sabemos que la función spline de interpolación es la más suave, lo cual nos garantiza que el método propuesto (spline promedio) reflejará desde su inicio un comportamiento suave y posteriormente central cuando se tomen los promedios.

Para el spline de los eneros se obtuvieron 109 valores al tabular la función desde t_1 hasta t_{109} , para el spline de los febreros se tabuló desde t_2 hasta t_{110} y así sucesivamente hasta llegar al spline de los diciembres donde se obtuvieron los valores de t_{12} hasta t_{120} .

Una vez tabulados los doce splines, el siguiente paso consistió en promediar los valores de las funciones en cada mes, así por ejemplo para la fecha t_k los promedios se calcularon de:

$$\bar{Y}_k = \left(\frac{\sum_{j=1}^{12} f_j(t_k)}{12} \right)$$

donde $k = 12, 13, \dots, 109$

- f_1 - Es el spline de los eneros
 f_2 - Es el spline de los febreros
 .
 .
 .
 f_{12} - Es el spline de los diciembres

Para los extremos de la serie se tiene un desplazamiento de los splines hacia la derecha como ya se explicó, en este caso el promedio se obtuvo de:

$$\bar{Y}_k = \left(\frac{\sum_{i=1}^n f_i(t_k)}{i} \right)$$

donde $n = k$

$k = 1, 2, \dots, 11$ para el extremo izquierdo

$k = 110, 111, \dots, 120$ para el extremo derecho

De este proceso llegaremos a una serie de promedios que denotaremos así: $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_{120}$ y que representan a la tendencia de la serie en forma discreta, pues suponemos que los puntos \bar{Y}_k carecen en buena parte de la influencia estacional, $\bar{Y}_1 = f_1(t_1)$ y $\bar{Y}_{120} = f_{12}(t_{120})$ pues no tienen con quien promediarse.

Cada uno de los puntos \bar{Y}_k calculados, al compararse contra los correspondientes valores de la tendencia verdadera ξ_k nos darán idea de qué tanto nos estamos acercando a la estimación de la tendencia, como se puede apreciar en la tabla 13.

TABLA 13 TENDENCIA

DESVIACION

	TENDENCIA	\bar{Y}	DESVIACION
1	988.5	692.9155965	295.5844035
2	978	720.3605812	197.1994182
3	968.5	791.9570634	176.5429366
4	960	840.8078419	119.1921581
5	952.5	891.3965708	61.10342916
6	946	912.1255405	32.87445945
7	940.5	917.9466801	22.55331905
8	936	906.7773903	29.22260969
9	932.5	902.364794	30.13520596
10	930	893.0505238	36.84947617
11	928.5	911.8660426	16.63395744
12	928	935.6157172	7.615717216
13	928.5	936.9267902	8.426790166
14	930	939.1894324	9.189432383
15	932.5	942.498068	9.998068091
16	936	946.3007289	10.90072887
17	940.5	952.4164713	11.91647132
18	946	959.0502966	13.05029658
19	952.5	966.7685568	14.26855678
20	960	975.5091025	15.5091025
21	968.5	985.2270397	16.72703973
22	978	995.8960045	17.89600453
23	988.5	1007.467795	18.96779532
24	1000	1019.894295	19.89429475
25	1012.5	1033.143245	20.64324451
26	1026	1047.186025	21.18602487
27	1040.5	1062.024244	21.52424355
28	1056	1077.68371	21.68371015
29	1072.5	1094.232184	21.73218414
30	1090	1111.74231	21.74231086
31	1108.5	1130.306314	21.80631433
32	1128	1150.03615	22.03615086
33	1148.5	1171.002162	22.50216179
34	1170	1193.232242	23.23224196
35	1192.5	1216.783105	24.28310503
36	1216	1241.67494	25.67494074
37	1240.5	1267.855303	27.35530332
38	1266	1295.235967	29.2359674
39	1292.5	1323.687189	31.18718831
40	1320	1353.092616	33.09261567
41	1348.5	1383.277727	34.7777273
42	1378	1414.083916	36.08391607
43	1408.5	1445.377149	36.87714947
44	1440	1477.035684	37.03568407
45	1472.5	1509.011457	36.51145722
46	1506	1541.312671	35.31267104
47	1540.5	1573.915372	33.4153716
48	1576	1606.862816	30.862816
49	1612.5	1640.308233	27.80823273
50	1650	1674.470095	24.47009474
51	1688.5	1709.616076	21.11607617
52	1728	1745.979418	17.979418
53	1768.5	1783.834326	15.33432648
54	1810	1823.38817	13.38817094
55	1852.5	1864.785908	12.28590758
56	1896	1906.123976	12.12397598
57	1940.5	1953.377338	12.87733818
58	1986	2000.467139	14.46713877
59	2032.5	2049.346961	16.84696135
60	2080	2099.918285	19.91828473

61	2138.5	2151.99489	23.49748879
62	2178	2205.297238	27.28723818
63	2238.5	2259.478843	30.97884334
64	2280	2314.258065	34.25806529
65	2332.5	2369.571983	38.57198282
66	2386	2424.688424	38.68842437
67	2440.5	2480.164771	39.66477104
68	2496	2535.828707	39.82870711
69	2552.5	2591.846869	39.84686896
70	2610	2648.379270	38.3792697
71	2668.5	2705.548071	37.04807107
72	2728	2763.47411	35.47411068
73	2788.5	2822.298423	33.7984229
74	2850	2882.336214	32.33621418
75	2912.5	2943.956512	31.45651239
76	2976	3007.435486	31.43548629
77	3040.5	3072.960546	32.46054595
78	3106	3140.573978	34.57397835
79	3172.5	3210.203306	37.70330507
80	3240	3281.727864	41.7278638
81	3308.5	3354.925705	46.42570505
82	3378	3429.710166	51.71016641
83	3448.5	3506.003672	57.50367235
84	3520	3583.747982	63.74798211
85	3592.5	3662.887476	70.88747505
86	3666	3743.116903	77.11690274
87	3740.5	3824.676252	83.5762521
88	3816	3905.527988	89.52798806
89	3892.5	3987.349259	94.8492586
90	3970	4069.507032	99.50703197
91	4048.5	4152.076243	103.5762434
92	4128	4235.152102	107.1521017
93	4208.5	4318.762649	110.2626487
94	4290	4402.751154	112.7511543
95	4372.5	4487.031617	114.5316170
96	4456	4571.554010	115.5540099
97	4540.5	4656.299006	115.7990059
98	4626	4741.506620	115.5066196
99	4712.5	4827.465674	114.9656736
100	4800	4914.368925	114.3689253
101	4888.5	5002.323903	113.8239032
102	4978	5091.4016	113.4016003
103	5069.5	5181.590808	113.0908085
104	5160	5272.895769	112.8957692
105	5252.5	5365.437364	112.9373637
106	5346	5459.472005	113.4720045
107	5440.5	5555.137377	114.6373768
108	5536	5652.491738	116.4917385
109	5632.5	5751.540726	119.0407258
110	5730	6038.271877	308.2718767
111	5828.5	6314.886455	486.3864555
112	5928	6540.489984	612.4899845
113	6028.5	6668.601062	640.1010618
114	6130	6630.51271	500.5127108
115	6232.5	6685.612862	453.1128625
116	6336	6835.448704	499.448704
117	6440.5	7302.243391	861.7433913
118	6546	7701.325215	1155.325215
119	6652.5	8645.953395	1993.459395
120	6760	9328.808778	2568.808778

Si ahora hacemos pasar por estos puntos \bar{Y}_k un spline de interpolación, con un valor para el coeficiente de suavidad S igual a cero y con pesos ΔY_k igual a uno se ganará una función continua a través de la cual tendremos valores intermedios.

El resultado de esta prueba se muestra en la tabla 14.

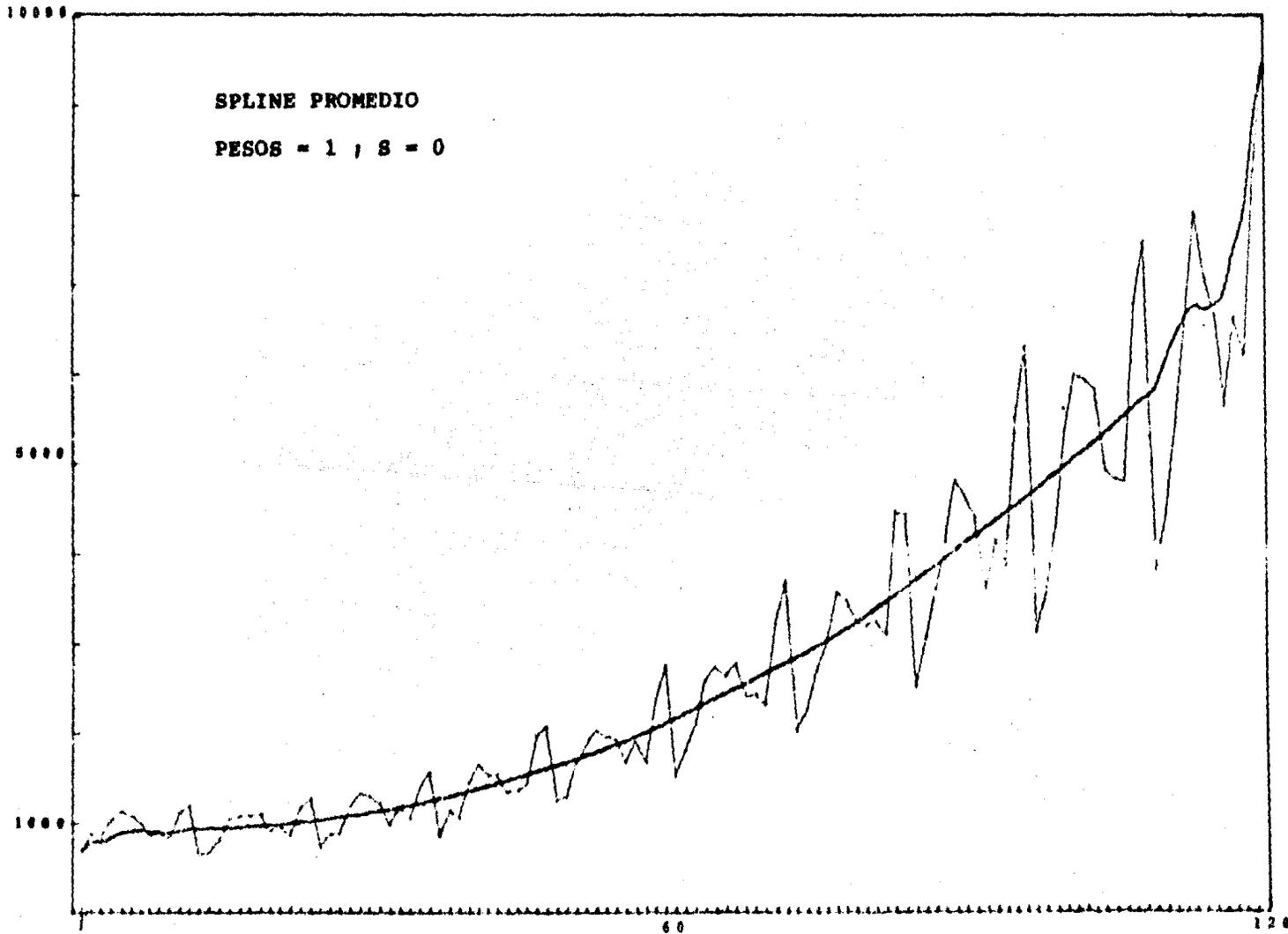
TABLA (14)

SUAVIDAD	PESOS	E.C.I.	E.C.C.	E.C.F.	F.C.T.
0	1	122.6	59.3	1098.7	353.6

(Ver gráfica 22)

Se debe señalar que el error cuadrático total anterior se incrementó fuertemente debido a la influencia de éste en su parte final, sin embargo, cabe mencionar que el comportamiento del spline en la parte central de esta serie dió el error cuadrático mínimo, de entre todos los métodos expuestos. No se usó el spline de aproximación pues no se cuenta con la desviación estándar de los errores de los puntos \bar{Y}_k .

Desafortunadamente el spline promedio falla bastante en los extremos, sería bueno continuar mejorando este procedimiento.



GRAFICA (22)

3.7 EVALUACION DEL SPLINE POR MINIMOS CUADRADOS

Este algoritmo lo presentó Carl de Boor en su libro "A practical guide to splines". El algoritmo, como ya se explicó en la sección 2.3.4, selecciona la mejor función, que en nuestro problema, se trata de un polinomio cúbico en un espacio lineal de funciones spline S_3 , usando el principio de mínimos cuadrados (desde un punto de vista geométrico).

Este método nos permite elegir de uno a $(N-1)$ divisiones del intervalo $[a, b]$ donde "N" es el número de observaciones. Se deben asignar pesos positivos y darle además, una sucesión de "puntos de quiebre", que de

acuerdo con la teoría de los splines (Carl de Boor, pág. 73) se puede obtener con ellos una función más suave.

La elección del número de subintervalos, pesos y "puntos de quiebre" hacen que este método, al igual que el spline de aproximación tenga que ser probado para distintos valores.

En la elección del número de subintervalos se observó que a mayor número de estos, la función resultante se acercaba más a la interpolación, de esta manera los mejores resultados en nuestro trabajo fueron de uno a cinco subintervalos.

Los pesos que se asignaron son los que propone Carl de Boor (fórmulas 2.4.3), unitarios y en base a la estacionalidad en cada punto t_k ($k = 1, 2, \dots, 120$).

Los puntos de quiebre que se usaron fueron los que propone Carl de Boor y las propias abscisas t_k ($k = 1, 2, \dots, 120$).

En este método es posible iterar varias veces, pero en la práctica nos encontramos con el inconveniente de que por lo general en la segunda iteración se obtienen valores del error cuadrático mayores a los de la primera, o bien, se obtenían las mismas cantidades, por es--

ta razón no consideramos útil su empleo.

En las tablas 15,16 y 17 se dan los resultados alcanzados en las distintas pruebas.

Puntos de quiebre = las abscisas (t_k)

TABLA (15)-

PESOS	INTERVALOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
1	1	921.03	386.65	187.86	456.03
1	120	627.03	4145.35	78804.00	25195.10
1	60	155.02	1010.40	536.50	920.78
1	30	155.02	645.81	235.37	584.46
1	10	502.70	498.20	180.02	476.50
1	5	763.50	438.90	172.80	464.10
1	3	850.90	417.60	184.00	464.00
1	2	889.00	402.50	194.60	460.88

TABLA (16)

PESOS	INTERVALOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
*	1	925.70	386.60	110.10	454.40
*	120	627.00	4145.30	78801.40	25194.30
*	10	502.70	490.90	117.50	468.48
*	5	763.20	430.40	104.30	455.60
*	3	851.60	411.70	103.30	457.30
*	2	886.40	388.50	103.10	447.90

* ($t_{(i+1)} - t_{(i-1)}$)/2

TABLA (17)

PESOS	ITERACIONES	E.C.I.	E.C.C.	E.C.F.	E.C.T.
EN BASE A LA ES TACIONA LIDAD.	1	917.18	430.37	613.32	519.53
	10	542.27	448.34	367.88	451.38
	5	759.53	432.14	432.00	475.35
	3	843.15	406.10	499.14	367.93
	2	883.26	404.89	441.20	478.15

$$\text{Puntos de quiebre} = (t(i) - t(i-1)) / 2$$

TABLA (18)

PESOS	ITERACIONES	E.C.I.	E.C.C.	E.C.F.	E.C.T.
1	1	924.40	397.70	211.60	461.70
1	10	543.60	490.20	196.90	475.30
1	5	754.90	414.90	178.10	444.80
1	3	847.00	408.90	196.60	457.60
1	1	888.30	401.60	160.20	458.80

TABLA (19)

PESOS	INTERVALOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
*	1	921.80	383.12	92.60	450.85
*	10	543.60	482.70	118.80	466.22
*	5	759.30	418.50	118.60	446.30
*	3	844.50	395.70	103.70	444.50
*	2	888.13	392.80	102.70	450.90

$$* (t(i+1) - t(i-1)) / 2$$

Otra prueba que se realizó fue aplicar este tipo de - spline a la serie obtenida con promedio móvil centrado, el cual tiene la característica de reducir el efecto estacional, los resultados se dan en las tablas 20, 21, 22 y 23.

Puntos de quiebre = a las abscisas (t_k)

TABLA (20)

PESOS	INTERVALOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
1	1	918.90	395.70	202.00	462.50
1	10	496.20	508.10	194.50	482.20
1	5	761.80	447.70	186.70	471.00
1	3	849.90	425.80	198.30	470.30
1	2	887.40	410.20	208.80	466.60

TABLA (21)

PESOS	INTERVALOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
*	1	925.00	397.90	143.50	462.90
*	10	496.20	500.40	143.20	476.40
*	5	761.60	442.20	130.60	464.90
*	3	850.70	423.20	132.90	466.20
*	2	884.90	399.30	147.10	556.10

* ($t(k+1) - t(k-1)$) /2

$$\text{Puntos de quiebre} = (t(k) - t(k-1)) / 2$$

TABLA (22)

PESOS	INTERVALOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
1	1	923.30	399.20	226.00	466.70
1	10	538.10	495.80	210.70	479.80
1	5	752.90	423.90	192.00	451.80
1	3	845.90	416.30	210.80	463.30
1	2	886.60	412.80	174.10	466.90

TABLA (23)

PESOS	INTERVALOS	E.C.I.	E.C.C.	E.C.F.	E.C.T.
*	1	920.90	398.80	114.70	461.90
*	10	538.10	491.40	147.80	473.70
*	5	757.80	427.80	152.50	454.00
*	3	843.30	407.60	135.00	453.70
*	2	886.60	405.00	133.80	460.00

$$* (t(k+1) - t(k-1)) / 2$$

CONCLUSIONES

CONCLUSIONES PARTICULARES

Para formular las conclusiones, primero mencionaremos el comportamiento que presentaron cada una de las técnicas al aplicarlas a nuestra serie de tiempo. Posteriormente se eligieron las mejores en base a los resultados.

Mínimos Cuadrados.

Aunque los resultados alcanzados fueron buenos, en este procedimiento nos encontramos con el problema de resolver el sistema de ecuaciones lineales, dado que su solución puede ser imprecisa cuando los coeficientes -- del sistema son muy desproporcionados como sucedió en --

nuestro trabajo.

Un rasgo notorio fue que en los extremos de la serie dió una muy buena aproximación a la tendencia, superando a todas las técnicas del promedio móvil como se puede ver en las tablas correspondientes.

Promedio Móvil.

Comunmente el promedio móvil es la técnica empleada para obtener la tendencia de una serie de tiempo, por este motivo aplicamos varias clases de este método. A continuación mencionaremos las características que observamos en la aplicación práctica de cada una de ellas.

a) Promedio móvil de primer orden.- Aquí resaltaba la facilidad con que se podía manejar la técnica, dado que se podían calcular todas las extensiones en forma casi inmediata sin la necesidad de dar al programa ningún parámetro.

b) Promedio móvil de segundo orden.- Para aplicar este método fue necesario calcular inicialmente un promedio móvil de primer orden y dejar los resultados en archivos (de programación) para aplicarles después un promedio móvil de la misma extensión y así obtener la tendencia.

Aunque hubo que correr dos veces el mismo programa siguió siendo de manejo fácil.

Nota: En promedios móviles de orden superior (tercero y cuarto) se observó que en la práctica de nuestro trabajo ya no se logró una disminución significativa -- del error cuadrático total, pues dieron respectivamente errores de 224.8 y 224.3 para una extensión 21.

c) Promedio móvil mixto.- En este caso se corrió dos veces el mismo programa pero con extensiones diferentes, los resultados no fueron mejores a los del promedio móvil de segundo orden.

d) Promedio móvil centrado.- Esta técnica fue empleada primero como un filtro para aplicar después un sólo-promedio móvil por mínimos cuadrados de primer orden -- con extensiones 17 y 21. Aquí se notó que al seguir esta secuencia se logra una buena estimación de la tendencia sobre todo cuando se trabaja con series que tengan estacionalidad. En general se apreció que este procedimiento dió resultados cercanos al promedio móvil de segundo orden.

Al aplicar el promedio móvil centrado a la serie de tiempo se pierden los seis primeros y últimos valores de la serie, por lo que fue necesario completar con las tablas del apéndice A, obteniendo resultados buenos.

De los procedimientos anteriores, podríamos escoger al promedio móvil de segundo orden como el mejor para representar a la tendencia, puesto que en esta serie, además de ser el más exacto es fácil de usar.

Spline.

La función spline a diferencia de mínimos cuadrados y de promedio móvil, no se había usado para estimar la tendencia de series de tiempo, por lo que hicimos varios intentos para lograrlo, con diferentes modificaciones.

Spline de aproximación.- En este método se redujo significativamente el error cuadrático bajo las siguientes condiciones:

a) Para el spline con pesos unitarios se tuvo que buscar el coeficiente de suavidad S que diera el menor error cuadrático, esto es posible sólo si se conoce la tendencia verdadera, pues el error cuadrático está en términos de esta componente.

b) Para el spline con pesos asignados (en base a la estacionalidad en cada punto t_n), se obtuvieron los mejores resultados de entre todos los métodos.

No debe olvidarse que la precisión dependiera directa--

mente de la eficiencia con que se estime la estacionalidad de la serie.

Es posible que si empleamos métodos espectrales (Kendall - 1976) para estimar la estacionalidad, esto nos permita obtener los pesos del spline de una manera más precisa, para la estimación de la tendencia.

Spline promedio.- En este caso obtuvimos un error -- cuadrático total aceptable, mientras que el error cuadrático central dió el mejor resultado de todos los métodos, con la ventaja de que los pesos son unitarios y el coeficiente de suavidad es igual a cero.

Este hecho coloca en un lugar preponderante al spline sobre los demás métodos. Por otra parte resulta ser un método muy laborioso y que pierde precisión en los extremos de la serie.

Spline por mínimos cuadrados.- Este algoritmo se usó con el fin de probar otra técnica de splines para estimar la tendencia de la serie de tiempo.

Al igual que los métodos anteriores se efectuaron varias pruebas con las distintas alternativas que esta técnica ofrecía:

a) Primero se usaron como puntos de quiebre las abscis

sas t_k ($k = 1, 2, \dots, 120$) con tres distintos tipos de pesos:

1) Unitarios

2) $\Delta Y_k = \text{ABS}((n_k - 1)\xi_k - \text{media})$

$$\text{media} = \frac{\sum_{k=1}^{120} (n_k - 1)\xi_k}{120} \quad k = 1, 2, \dots, 120$$

3) los que sugiere Carl de Boor.

Cada uno de estos pesos con diferente número de intervalos.

b) En la segunda prueba sólo se cambió el valor de los puntos de quiebre por los que indica Carl de Boor; fue aquí donde se consiguieron los mejores resultados de esta técnica.

Se usaron los pesos indicados en la fórmula (inciso anterior), ya que en el spline de Reinsch se lograron muy buenos resultados, sin embargo, éstos fueron muy similares a las pruebas con los otros pesos.

Desafortunadamente este método no dió buenos resultados para la serie que simulamos.

De entre todas las técnicas de spline consideramos - al spline de aproximación con pesos calculados a partir de los valores estacionales como el mejor, puesto que - los errores cuadráticos tanto en su parte central y en los extremos fue uno de los menores pero en el total sí llegó a ser el menor.

CONCLUSIONES GENERALES

Las aplicaciones prácticas de la tendencia de una serie de tiempo tienen gran importancia dentro de la planeación social, económica y administrativa pues permite analizar sus fenómenos y aún prever comportamientos futuros. En las empresas comerciales, algunas veces se requiere de la estimación de la estacionalidad, para lo cual se necesita primero, estimar la tendencia.

Una vez comentado el comportamiento de cada técnica, resta por decidir cuáles de ellas resultan las más eficientes en precisión y manejo para representar a la tendencia de nuestra serie de tiempo teórica. Para medir la precisión requerimos de un criterio que nos indique a partir de qué valor, para el error cuadrático, aceptaremos o rechazaremos la técnica.

El criterio de elección que usaremos consiste en tomar la idea de la media aritmética entre un error cuadrático igual a cero (máxima precisión) y un error cuadrático, representado por las diferencias entre la ten

dencia verdadera y la serie de tiempo teórica. De esta manera el valor de la media que tenemos lo podemos comparar contra el mejor error cuadrático total de cada técnica y se éste es menor o igual al valor de la media entonces podemos decir, de acuerdo a este criterio, que la aproximación a la tendencia de la serie es buena. Por el contrario se el error cuadrático es mayor entonces el método es rechazado.

El valor de la media aritmética es de 313.2 y en forma análoga obtenemos los valores para la parte inicial, central y final de la serie que son de 77.5, 268.7 y 633.2, pero definitivamente la comparación más importante será la del error cuadrático total.

En la tabla 24 se presentan los mejores resultados de cada técnica, y si los comparamos con los valores dados por el criterio de elección, podemos notar que todos los errores cuadráticos son menores excepto el error cuadrático final del promedio móvil. Este defecto del promedio móvil se compensa por su gran facilidad de manejo a diferencia del spline, que es más preciso pero a la vez más complicado y costoso. El caso intermedio lo presenta mínimos cuadrados tanto en precisión, dificultad y costo.

	TABLA (24)			
	E.C.I.	E.C.C.	E.C.F.	E.C.T.
VALORES DE ELECCION	77.5	267.7	633.2	313.2
TECNICAS:				
MINIMOS CUADRADOS.	69.1	78.4	431.4	154.9
PROMEDIO MOVIL 2o. ORDEN	74.3	59.6	689.6	225.7
SPLINE DE APROXIMACION CON PESOS EN BASE A LA ESTACIONALIDAD.	54.8	61.4	68.9	61.6

Conviene considerar por separado el caso muy especial del spline promedio, punto que si recordamos su E.C.C. = 59.3, - fue el mejor de todos, con la limitante que cada vez que se aplique este método, sus once primeros y últimos valores de tendencia darán un error cuadrático muy elevado (razón por la que el criterio de selección lo eliminó). Por otra parte, es importante mencionar que este procedimiento sólo se puede aplicar a información mensual y trimestral, y con una buena cantidad de información.

En general, podemos establecer que el uso de cualquier tipo de promedio móvil o spline, sólo estimará a la tendencia, dando lugar a diferentes grados de error, puesto que la tendencia calculada también ha captado parte de las variaciones estacionales e irregulares, que en un análisis -

nos pueden conducir a interpretaciones equivocadas. Esto dependerá directamente del procedimiento elegido y de que las consideraciones previas estén bien establecidas. En consecuencia, muy probablemente, la estimación posterior de las componentes estacional e irregular, también se verán distorsionadas en función de la técnica particular empleada. Esta dificultad, parece inevitable, al menos a lo que concierne a la estimación de la tendencia por los métodos aquí expuestos.

Finalmente podemos decir que el presente trabajo no está en posición de inferir conclusiones definitivas debido a que se trabajó con una sola serie, pero sí es posible -- considerarlo como la ilustración comparativa de técnicas -- donde se muestran sus ventajas y desventajas, y que pueden utilizarse como punto de partida para investigaciones posteriores, donde se tenga interés particular acerca de algunos de los métodos aquí expuestos. Sin embargo, podemos afirmar que es posible aplicar métodos del análisis numérico, con buenos resultados, en la estimación de la tendencia de una serie de tiempo.

OBJETIVOS ALCANZADOS

Consideramos que los objetivos propuestos en la pág. seis si se cubrieron, pues:

- 1) Se incluyó la información suficiente para la comprensión de todos los temas.
- 2) El método propuesto (spline) si dió buenos resultados, aunque con las condiciones ya mencionadas.
- 3) El empleo de la serie simulada nos permitió manejar el problema con sencillez y objetividad para ilustrar el comportamiento de cada uno de los métodos empleados.
- 4) a) El criterio de evaluación (pág. 146) nos permitió medir la eficiencia con que cada método estimó a la tendencia.

b) El criterio de elección (pág. 225) nos permitió seleccionar los métodos que dieron los mejores resultados para la serie simulada.

APENDICE A

El algoritmo de obtención de los pesos para un promedio móvil por mínimos cuadrados y tomando como punto de partida a un polinomio cúbico, es relativamente sencillo. Tales pesos se utilizan para estimar los valores de tendencia de una serie de tiempo.

Los coeficientes del sistema de ecuaciones lineales, que nos permite calcular dichos pesos, sólo dependen de la extensión del promedio elegida. Así que conforme la extensión del promedio se incrementa, también será mayor el número de operaciones para determinar a los pesos. Pero tales operaciones originan un problema, pues los resultados parciales son tan grandes, que la máquina los expresa en forma

exponencial, lo que implica una pérdida de precisión en el cálculo de los pesos que se muestran en la tabla (1).

Debido a este inconveniente, se le incorporaron al programa, rutinas que realizarán los cálculos utilizando el principio de mínimo común múltiplo, para hacer simplificaciones a lo largo del procedimiento. Esta alternativa finalmente nos dió los resultados con toda precisión.

En cuanto a los pesos que consideramos que se debían emplear para estimar los valores de tendencia para los extremos de la serie de tiempo, se tienen que calcular de una manera particular, tal y como se explicó anteriormente (sec. 1.4.5), pero nuevamente se presentó el problema de tener números grandes en los resultados parciales, y ahora con mayor intensidad. Aquí también se procedió a agregar al programa rutinas de simplificaciones parciales, para no dejar crecer a los números en la misma forma como se hizo antes.

Para calcular los pesos correspondientes para los extremos de la serie, se utilizó en cada extensión del promedio, una fórmula similar a la (1.4.12).

El empleo de las rutinas de simplificación, nos condujo a obtener vectores de pesos más simplificados que los presentados por Kendall (1976), y para verificarlo pueden compararse con las tablas aquí incluidas.

Los programas se corrieron en la microcomputadora HP-9830,

EXTENSION= 7

EXTENSION= 5

69	2
4	27
-6	12
4	-8
-1	2
70	35

39	8	-4
8	19	16
-4	16	19
-4	6	12
1	-4	2
4	-7	-4
-2	4	1
42	42	42

EXTENSION= 9

85	56	-28	-56
28	65	392	84
-2	56	515	144
-12	36	432	145
-9	12	234	108
0	-9	12	54
8	-20	-143	4
8	-14	-140	-21
-7	16	112	0
99	198	1386	462

EXTENSION= 11

113	48	24	-72	-51
48	41	96	132	36
8	32	123	232	86
-12	22	116	251	106
-17	12	86	212	103
-12	3	44	138	84
-2	-4	1	52	56
8	-8	-32	-23	26
13	-8	-44	-64	1
8	-3	-24	-48	-12
-12	8	39	48	-6
143	143	429	858	429

EXTENSION= 13

232

265	33	231	-33	-407	-116
132	25	396	132	308	33
42	18	460	222	738	128
-12	12	444	251	932	182
-37	7	369	233	939	202
-40	3	256	182	808	195
-28	0	126	112	588	168
-8	-2	0	37	328	128
13	-3	-101	-29	77	32
28	-3	-156	-72	-116	37
30	-2	-144	-78	-202	0
12	0	-44	-33	-132	-22
-33	3	165	77	143	-32
564	91	2002	1001	4004	1001

EXTENSION= 15

2059	8008	44044	2002	-19201	-28756	-3094
1144	5833	52624	17017	19604	8879	156
484	4048	54709	25762	44294	34244	2511
44	2618	51524	29252	57004	49054	4076
-211	1508	44294	38502	59869	55024	4956
-316	683	34244	24527	55024	53869	5256
-306	108	22599	18342	44604	47304	5081
-216	-252	10584	10962	30744	37044	4536
-81	-432	-576	3402	15579	24804	3726
64	-467	-9656	2222	1244	12299	2756
184	-392	-15431	-8198	-10126	1244	1791
244	-242	-16676	-10208	-16396	-6646	756
209	-52	-12166	-8338	-15431	-9656	-64
44	143	-676	-1573	-5096	-6071	-624
-286	308	19019	11102	16744	5824	-819
3060	21420	278460	139230	278460	278460	30940

EXTENSION= 17

605	1456	728	208	-66	-320	-384	-352
364	1055	728	468	134	121	20	-42
182	728	691	624	267	428	313	196
52	468	624	691	342	618	508	369
-33	268	534	684	368	708	618	484
-80	121	428	618	354	715	656	548
-96	20	313	508	309	656	635	568
-88	-42	196	369	242	548	568	551
-63	-72	84	216	162	408	468	504
-28	-77	-16	64	78	253	348	434
10	-64	-97	-72	-1	100	221	348
44	-40	-152	-177	-66	-34	100	253
67	-12	-174	-236	-108	-132	-2	156
72	13	-156	-234	-118	-177	-72	64
52	28	-91	-156	-87	-152	-97	-16
0	26	28	13	-6	-48	-64	-77
-91	0	208	288	134	176	48	-112
969	3876	3876	3876	1938	3876	3876	3876

EXTENSION= 19

14467	1632	15504	25296	-153
9248	1193	14144	36856	5372
5168	832	12601	43696	9162
2108	542	10924	48439	11427
-51	316	9162	45708	12377
-1428	147	7364	42126	12222
-2142	28	5579	38316	11172
-2312	-48	3856	28901	9437
-2057	-88	2244	20504	7227
-1496	-99	792	11748	4752
-748	-88	-451	3256	2222
68	-62	-1436	-4349	-153
833	-28	-2114	-10444	-2169
1428	7	-2436	-14406	-3598
1734	36	-2353	-15612	-4248
1632	52	-1816	-13439	-3903
1003	48	-776	-7264	-2353
-272	17	816	3536	612
-2312	-48	3009	19584	5202
24871	4389	74613	298452	74613

-1224	-918	-27744	-561
714	68	-3264	-136
2104	797	15424	204
3009	1297	28901	466
3492	1596	37748	657
3616	1722	42546	784
3444	1703	43876	854
3039	1567	42319	874
2464	1342	38456	851
1782	1056	32868	792
1056	737	26136	704
349	413	18841	594
-276	112	11564	469
-756	-138	4886	336
-1028	-309	-612	202
-1029	-373	-4349	74
-696	-302	-5744	-41
34	-68	-4216	-136
1224	357	816	-204
21318	10659	298452	6783

EXTENSION= 21

5781	1938	22287	32946	15447
3876	1437	19380	38760	46740
2346	1020	16593	41820	68610
1156	680	13940	42469	82100
271	410	11435	41050	88253
-344	203	9092	37906	88112
-724	52	6925	33380	82720
-904	-50	4948	27815	73120
-919	-110	3175	21554	60355
-804	-135	1620	14940	45468
-594	-132	297	8316	29502
-324	-108	-780	2025	13500
-29	-70	-1597	-3598	-1495
256	-25	-2140	-8186	-14440
496	20	-2395	-11420	-24292
656	58	-2348	-12949	-30008
701	82	-1985	-12430	-30545
596	85	-1292	-9520	-24860
306	60	-255	-3876	-11910
-204	0	1140	4845	9348
-969	-102	2907	16986	39957
10626	5313	100947	302841	605682

-9804	-20634	-25764	-52383	-2546
11571	2964	-2850	-12540	-855
27276	20775	14844	19050	540
37906	33380	27815	43108	1660
44056	41360	36560	60355	2526
46321	45296	41576	71512	3159
45296	45769	43360	77300	3580
41576	43360	42409	78440	3810
35756	38650	39220	75653	3870
28431	32220	34290	69660	3781
20196	24651	28116	61182	3564
11646	16524	21195	50940	3240
3376	8420	14024	39655	2830
-4019	920	7100	28048	2355
-9944	-5395	920	16840	1836
-13804	-9944	-4019	6752	1294
-15004	-12146	-7220	-1495	750
-12949	-11420	-8186	-7180	225
-7044	-7185	-6420	-9582	-260
3306	1140	-1425	-7980	-684
18696	14136	7296	-1653	-1026
302841	302841	302841	605682	33649

```

10 REM "PROGRAMA PARA PROMEDIO MOVIL"
20 DISP "EXTENSION DEL PROMEDIO"
30 WAIT 1000
40 INPUT E
50 PRINT "EXTENSION="E
60 FOR I=2 TO (E-1)/2
70 IF (E=2*I+1) THEN 120
80 NEXT I
90 DISP "EXTENSION IMPROPIA"
100 WAIT 1000
110 GOTO 20
120 K=(E-1)/2
130 DIM W(21),A(21),B(21),C(21),I(21),F(10,21),U(120),P(120),T(120)
140 LOAD DATA 2,U
150 T2=T4=T6=0
160 FOR I=1 TO K
170 T2=T2+2*I+2
180 T4=T4+2*I+4
190 T6=T6+2*I+6
200 NEXT I
210 T0=T4/T2
220 IF (T0=INT(T0)) THEN 280
230 R=5*(E*T0-T2)
240 FOR I=1 TO K+1
250 W(I)=5*(T0-(K+1-I)+2)
260 NEXT I
270 GOTO 320
280 R=E*T0-T2
290 FOR I=1 TO K+1
300 W(I)=T0-(K+1-I)+2
310 NEXT I
320 FOR I=1 TO K
330 W(E+1-I)=W(I)
340 NEXT I
350 R1=R3=T2+T6-T4+2
360 R2=E*T4-T2+2
370 E1=E
380 T7=T2
390 T8=T4
400 FOR I=21 TO 2 STEP -1
410 IF (R1/I=INT(R1/I)) AND (T4/I=INT(T4/I)) AND (T6/I=INT(T6/I)) THEN 2160
420 NEXT I
430 FOR I=21 TO 2 STEP -1
440 IF (E1/I=INT(E1/I)) AND (R2/I=INT(R2/I)) AND (T2/I=INT(T2/I)) THEN 2200
450 NEXT I
460 FOR I=21 TO 2 STEP -1
470 IF (R3/I=INT(R3/I)) AND (T7/I=INT(T7/I)) AND (T8/I=INT(T8/I)) THEN 2240
480 NEXT I
490 FOR I=1 TO E
500 A(I)=-T6*(K+1-I)+T4*(K+1-I)+2
510 B(I)=E1*(K+1-I)+2-T2
520 C(I)=-T7*(K+1-I)+T8*(K+1-I)
530 NEXT I
540 FOR L=21 TO 2 STEP -1
550 IF (R1/L=INT(R1/L)) THEN 570
560 GOTO 640

```

```
570 FOR J=1 TO E
580 IF (A[J]/L#INT(A[J]/L)) THEN 640
590 NEXT J
600 FOR I=1 TO E
610 A[I]=A[I]/L
620 NEXT I
630 R1=R1/L
640 NEXT L
650 FOR L=21 TO 2 STEP -1
660 IF (R2/L=INT(R2/L)) THEN 680
670 GOTO 750
680 FOR J=1 TO E
690 IF (B[J]/L#INT(B[J]/L)) THEN 750
700 NEXT J
710 FOR I=1 TO E
720 B[I]=B[I]/L
730 NEXT I
740 R2=R2/L
750 NEXT L
760 FOR L=21 TO 2 STEP -1
770 IF (R3/L=INT(R3/L)) THEN 790
780 GOTO 860
790 FOR J=1 TO E
800 IF (C[J]/L#INT(C[J]/L)) THEN 860
810 NEXT J
820 FOR I=1 TO E
830 C[I]=C[I]/L
840 NEXT I
850 R3=R3/L
860 NEXT L
870 Q=1
880 R0=R
890 FOR I=25 TO 2 STEP -1
900 J=I
910 S=R0/J
920 T=R1/J
930 U=R2/J
940 V=R3/J
950 IF (S=INT(S)) AND (T=INT(T)) AND (U=INT(U)) AND (V=INT(V)) THEN 2280
960 NEXT I
970 H0=R0
980 H1=R1
990 H2=R2
1000 H3=R3
1010 Q1=1
1020 FOR I=25 TO 2 STEP -1
```

```

1030 J=I
1040 IF (H0/J=INT(H0/J)) AND (H1/J=INT(H1/J)) AND (H2/J=INT(H2/J)) THEN 2340
1050 IF (H0/J=INT(H0/J)) AND (H2/J=INT(H2/J)) AND (H3/J=INT(H3/J)) THEN 2390
1060 IF (H0/J=INT(H0/J)) AND (H1/J=INT(H1/J)) AND (H3/J=INT(H3/J)) THEN 2440
1070 IF (H1/J=INT(H1/J)) AND (H2/J=INT(H2/J)) AND (H3/J=INT(H3/J)) THEN 2490
1080 NEXT I
1090 Q2=1
1100 FOR I=25 TO 2 STEP -1
1110 J=I
1120 IF (H0/J=INT(H0/J)) AND (H1/J=INT(H1/J)) THEN 2540
1130 IF (H0/J=INT(H0/J)) AND (H2/J=INT(H2/J)) THEN 2580
1140 IF (H0/J=INT(H0/J)) AND (H3/J=INT(H3/J)) THEN 2620
1150 IF (H1/J=INT(H1/J)) AND (H2/J=INT(H2/J)) THEN 2660
1160 IF (H1/J=INT(H1/J)) AND (H3/J=INT(H3/J)) THEN 2700
1170 IF (H2/J=INT(H2/J)) AND (H3/J=INT(H3/J)) THEN 2740
1180 NEXT I
1190 R4=H0+H1+H2+H3+Q1*Q2
1200 R0=R4/R0
1210 R1=R4/R1
1220 R2=R4/R2
1230 R3=R4/R3
1240 FOR I=1 TO E
1250 DCI)=DCI)+R0
1260 ACI)=ACI)+R1
1270 BCI)=BCI)+R2
1280 CCI)=CCI)+R3
1290 NEXT I
1300 FOR T=1 TO K
1310 FOR I=1 TO E
1320 FCT, I)=DCI)+ACI)+T+BCI)+T+2+CCI)+T+3
1330 NEXT I
1340 NEXT T
1350 FOR T=1 TO K
1360 RCT)=R4
1370 FOR L=38 TO 2 STEP -1
1380 IF (RCT/L=INT(RCT/L)) THEN 1400
1390 GOTO 1470
1400 FOR J=1 TO E
1410 IF (FCT, J)/L=INT(FCT, J)/L) THEN 1470
1420 NEXT J
1430 FOR I=1 TO E
1440 FCT, I)=FCT, I)/L
1450 NEXT I
1460 RCT)=RCT)/L
1470 NEXT L
1480 NEXT T

```

```

1490 FOR I=1 TO K
1500 R[I]=R[I]*Q
1510 NEXT I
1520 FOR J=1 TO 120-2*K
1530 S=0
1540 FOR I=J TO E+J-1
1550 S=S+W[I+1-J]*U[I]
1560 NEXT I
1570 P[J+K]=S/R
1580 NEXT J
1590 FOR T=K TO 1 STEP -1
1600 S1=0
1610 FOR J=E TO 1 STEP -1
1620 S1=S1+F[T,J]*U[E-J+1]
1630 NEXT J
1640 P[K-T+1]=S1/R[T]
1650 NEXT T
1660 FOR T=1 TO K
1670 S2=0
1680 FOR J=1 TO E
1690 S2=S2+F[T,J]*U[120-E+J]
1700 NEXT J
1710 P[120-K+T]=S2/R[T]
1720 NEXT T
1730 REM EN "T" QUEDA LA TENDENCIA
1740 LOAD DATA 4,T
1750 E1=0
1760 FOR I=1 TO 12
1770 E1=E1+(T[I]-P[I])^2
1780 NEXT I
1790 PRINT "E.C.M. 1-12 "E1
1800 E2=0
1810 FOR I=13 TO 100
1820 E2=E2+(T[I]-P[I])^2
1830 NEXT I
1840 PRINT "E.C.M. 13-100 "E2
1850 E3=0
1860 FOR I=109 TO 120
1870 E3=E3+(T[I]-P[I])^2
1880 NEXT I
1890 PRINT "E.C.M 109-120 "E3
1900 A=E1+E2+E3
1920 PRINT "E.C.M. TOTAL "A
1930 SCALE 0,120,0,10000
1940 XAXIS 120,1
1950 YAXIS 0,1000

```

```
1960 X=120
1970 FOR I=0 TO 10000 STEP 10000
1980 PLOT X,I
1990 NEXT I
2000 WAIT 3000
2010 PEN
2020 Y=10000
2030 FOR I=0 TO 120 STEP 120
2040 PLOT I,Y
2050 NEXT I
2060 WAIT 3000
2070 PEN
2080 FOR I=1 TO 120
2090 PLOT I,UC(I)
2100 NEXT I
2110 WAIT 3000
2120 FOR I=1 TO 120
2130 PLOT I,PC(I)
2140 NEXT I
2150 END
2160 T4=T4/I
2170 T6=T6/I
2180 R1=R1/I
2190 GOTO 420
2200 E1=E1/I
2210 R2=R2/I
2220 T2=T2/I
2230 GOTO 450
2240 R3=R3/I
2250 T7=T7/I
2260 T8=T8/I
2270 GOTO 480
2280 R0=R0/J
2290 R1=R1/J
2300 R2=R2/J
2310 R3=R3/J
2320 Q=Q*J
2330 GOTO 910
2340 H0=H0/J
2350 H1=H1/J
2360 H2=H2/J
2370 Q1=Q1*J
2380 GOTO 1040
2390 H0=H0/J
2400 H2=H2/J
2410 H3=H3/J
2420 Q1=Q1*J
```

2430 GOTO 1050
2440 H0=H0/J
2450 H1=H1/J
2460 H3=H3/J
2470 Q1=Q1*J
2480 GOTO 1060
2490 H1=H1/J
2500 H2=H2/J
2510 H3=H3/J
2520 Q1=Q1*J
2530 GOTO 1070
2540 H0=H0/J
2550 H1=H1/J
2560 Q2=Q2*J
2570 GOTO 1120
2580 H0=H0/J
2590 H2=H2/J
2600 Q2=Q2*J
2610 GOTO 1130
2620 H0=H0/J
2630 H3=H3/J
2640 Q2=Q2*J
2650 GOTO 1140
2660 H1=H1/J
2670 H2=H2/J
2680 Q2=Q2*J
2690 GOTO 1150
2700 H1=H1/J
2710 H3=H3/J
2720 Q2=Q2*J
2730 GOTO 1160
2740 H2=H2/J
2750 H3=H3/J
2760 Q2=Q2*J
2770 GOTO 1170

APENDICE B

En este apéndice, primeramente se muestra el teorema de Holladay mencionado en el capítulo 2. Además se incluyen los programas correspondientes a splines, que son:

- a) Spline de aproximación *
- b) Spline promedio *
- c) Spline por mínimos cuadrados

Y una corrida del programa de splines por mínimos cuadrados.

* Estos splines se corrieron en una microcomputadora HP-9830.

Teorema de Holladay.

Dados $\Delta: a = x_1 < x_2 < \dots < x_n = b$ y $Y = \{y_k \mid y_k \in \mathbb{R}, k = \overline{1, n}\}$, entonces entre todas las funciones $g(x) \in C^2 [a, b]$ que cumplen con $g(x_k) = y_k$ con $k = \overline{1, n}$, la función spline $f(x)$ que se une en cada nodo x_k con $k = \overline{2, n-1}$ y donde $f''(a) = f''(b) = 0$ minimiza la integral $\int_a^b (f''(x))^2 dx$.

Demostración:

$$\int_a^b (g'' - f'')^2 dx = \int_a^b (g'')^2 dx - \int_a^b (f'')^2 dx - 2 \sum_{k=2}^n \int_{x_{k-1}}^{x_k} (g'' - f'') f'' dx$$

$$\begin{aligned} \text{como: } \int_{x_{j-1}}^{x_j} (g'' - f'') f'' dx &= (g' - f') f'' \Big|_{x_{j-1}}^{x_j} - (g - f) f''' \Big|_{x_{j-1}}^{x_j} \\ &= (g' - f') f'' \Big|_{x_{j-1}}^{x_j} \end{aligned}$$

pues $g(x_k) = f(x_k)$ con $k = \overline{1, n}$. Y debido a la continuidad de la expresión anterior en $[a, b]$ se tiene que:

$$\int_a^b (g'' - f'') f'' dx = (g' - f') f'' \Big|_a^b = 0$$

pues $f''(a) = f''(b) = 0$. De aquí que:

$$\int_a^b (g'' - f'')^2 dx = \int_a^b (g'')^2 dx - \int_a^b (f'')^2 dx \geq 0$$

por lo tanto:

$$\int_a^b (g'')^2 dx \geq \int_a^b (f'')^2 dx$$

PLINE DE APROXIMACION

```

10 DIM X(122),Y(122),Z(122),A(122),B(122),C(122),D(122),E(122)
20 DIM S(122),Q(122),U(122),T(122),O(122),V(122),F(122),G(122),W(122),G(122)
40 M=9=0
90 FOR I=1 TO 120
100 Z(I+1)=1
110 NEXT I
160 Z(I)=0
170 FOR I=2 TO 121
180 X(I)=1
190 NEXT I
200 LOAD DATA 2,M
210 FOR I=1 TO 120
220 Y(I+1)=X(I)
230 NEXT I
240 W(I)=X(I)=0
250 INPUT S
260 PRINT "SUAVIDAD"S
270 N1=2
280 N2=121
290 M1=N1-1
300 M2=N2+1
310 R(M1)=R(N1)=S(N2)=Q(N2)=O(N2)=U(M1)=U(N1)=U(N2)=U(M2)=P=0
320 M1=M1+1
330 M2=M2-1
340 H=X(M1)-X(N1)
350 F=(Y(M1)-Y(N1))/H
360 FOR I=M1 TO M2
370 G=H
380 H=X(I+1)-X(I)
390 E=F
400 F=(Y(I+1)-Y(I))/H
410 A(I)=F-E
420 T(I)=2*(G+H)/3
430 O(I)=H/3
440 Q(I)=Z(I-1)/G
450 R(I)=Z(I+1)/H
460 S(I)=(Z(I)-Z(I+1))/G-Z(I)/H
470 NEXT I
480 FOR I=M1 TO M2
490 B(I)=R(I)*Q(I)+S(I)*S(I)+U(I)*O(I)
500 C(I)=R(I)*S(I+1)+S(I)*O(I+1)
510 D(I)=R(I)*O(I+2)
520 NEXT I
530 FOR I=

```

```

540 FOR I=M1 TO M2
550 S[I-1]=F*RC[I-1]
560 Q[I-2]=G*RC[I-2]
570 RC[I]=1/(P+BC[I]+TC[I]-F*SC[I-1]-L*Q[I-2])
580 UC[I]=RC[I]-S[I-1]*UC[I-1]-Q[I-2]*UC[I-2]
590 F=P*CC[I]+Q[I]-H*SC[I-1]
600 G=H
610 H=DC[I]*P
620 NEXT I
630 FOR I=M2 TO M1 STEP -1
640 UC[I]=RC[I]*UC[I]-S[I]*UC[I+1]-Q[I]*UC[I+2]
650 NEXT I
660 E=H=0
670 FOR I=M1 TO M2
680 G=H
690 H=(UC[I+1]-UC[I])/(XC[I+1]-XC[I])
700 VC[I]=(H-G)*Z[I]*Z[I]
710 E=E+VC[I]*(H-G)
720 NEXT I
730 G=VC[N2]=-H*Z[N2]*Z[N2]
740 E=E-G*H
750 G=F2
760 F2=E*P*P
770 IF F2 >= S OR F2 <= G THEN 910
780 F=0
790 H=(VC[M1]-VC[N1])/(XC[M1]-XC[N1])
800 FOR I=M1 TO M2
810 G=H
820 H=(VC[I+1]-VC[I])/(XC[I+1]-XC[I])
830 G=H-G-SC[I-1]+RC[I-1]-Q[I-2]*RC[I-2]
840 F=F+G*RC[I]*G
850 RC[I]=G
860 NEXT I
870 H=E-P*F
880 IF H <= 0 THEN 910
890 P=P+(S-F2)/(SQR(S/E)+P)*H)
900 GOTO 540
910 FOR I=N1 TO N2
920 AC[I]=VC[I]-P*VC[I]
930 CC[I]=UC[I]
940 NEXT I
950 FOR I=N1 TO N2
960 H=(XC[I+1]-XC[I])
970 DC[I]=(CC[I+1]-CC[I])/(3*H)
980 RC[I]=(AC[I+1]-AC[I])/H-(H*DC[I]+CC[I])/H

```

```

990 NEXT I
1000 C[N2]=0
1040 X[121]=0
1050 FOR I=1 TO 120
1060 X[I]=1
1070 NEXT I
1080 FOR I=2 TO 121
1090 Y[I-1]=Y[I]
1100 AC[I-1]=AC[I]
1110 CC[I-1]=CC[I]
1120 NEXT I
1130 FOR I=2 TO 120
1140 DC[I-1]=DC[I]
1150 BC[I-1]=BC[I]
1160 NEXT I
1165 AC[120]=BC[120]=CC[120]=DC[120]=0
1264 LOAD DATA 4,E
1265 A=B=C=D=0
1270 FOR I=1 TO 120
1280 D=D+(E[I]-AC[I])2
1281 NEXT I
1282 D=SQR(D/120)
1283 FOR I=1 TO 12
1284 A=A+(E[I]-AC[I])2
1285 NEXT I
1286 A=SQR(A/12)
1287 FOR I=13 TO 108
1288 B=B+(E[I]-BC[I])2
1289 NEXT I
1290 B=SQR(B/96)
1291 FOR I=109 TO 120
1292 C=C+(E[I]-AC[I])2
1293 NEXT I
1294 C=SQR(C/12)
1295 PRINT "TOTAL=",D,"INICIAL=",A,"CENTRAL=",B,"FINAL=",C
1296 STOP
1300 PRINT TAB5"A"TAB21"B"TAB36"C"TAB51"D"
1310 PRINT
1320 FOR I=1 TO 120
1330 PRINT AC[I];BC[I];CC[I];DC[I]
1340 NEXT I
1350 END

```

```

10 DIM X(145),Y(145),Z(145),A(145),B(145),C(145),D(145),E(145)
20 DIM S(145),Q(145),U(145),T(145),O(145),V(145),F(145),R(145)
30 DIM H(120)
40 LOAD DATA 2,H
50 DISP "VALOR DE I2"
60 WAIT 3000
70 INPUT I2
80 FOR I=1 TO 10
90 Z[I2*I+12]=1
100 NEXT I
110 FOR I=1 TO 10
120 X[I2*I+12]=12*I+I2
130 NEXT I
140 FOR I=1 TO 10
150 Y[I2*I+12]=H[I2*I-6]
160 NEXT I
170 DISP "SUAVIDAD"
180 WAIT 3000
190 INPUT S
200 PRINT "SUAVIDAD" S
210 DISP "VALORES DE P1 P2"
220 WAIT 3000
230 INPUT P1,P2
240 N1=P1
250 N2=P2
260 M1=N1-12
270 M2=N2+12
280 R[M1]=R[N1]=S[M2]=Q[M2]=O[M2]=U[M1]=U[N1]=U[M2]=U[M2]=F=0
290 M1=N1+12
300 M2=N2-12
310 H=X[M1]-X[N1]
320 F=(Y[M1]-Y[N1])/H
330 FOR I=M1 TO M2 STEP 12
340 G=H
350 H=X[I+12]-X[I]
360 E=F
370 F=(Y[I+12]-Y[I])/H
380 A[I]=F-E
390 T[I]=2*(G+H)/3
400 O[I]=H/3
410 Q[I]=Z[I-12]/G
420 R[I]=Z[I+12]/H
430 S[I]=-Z[I]/G-Z[I]/H
440 NEXT I
450 FOR I=M1 TO M2 STEP 12
460 B[I]=R[I]*R[I]+S[I]*S[I]+Q[I]*Q[I]

```

```

470 C[I]=R[I]*S[I+12]+S[I]*Q[I+12]
480 D[I]=R[I]*Q[I+24]
490 NEXT I
500 F2=-S
510 FOR I=M1 TO M2 STEP 12
520 S[I-12]=F*R[I-12]
530 Q[I-24]=G*R[I-24]
540 R[I]=1/(P*B[I]+T[I]-F*S[I-12]-G*Q[I-24])
550 U[I]=R[I]-S[I-12]*U[I-12]-Q[I-24]*U[I-24]
560 F=P*C[I]+Q[I]-H*S[I-12]
570 G=H
580 H=D[I]*P
590 NEXT I
600 FOR I=M2 TO M1 STEP -12
610 U[I]=R[I]*U[I]-S[I]*U[I+12]-Q[I]*U[I+24]
620 NEXT I
630 E=H*Q
640 FOR I=M1 TO M2 STEP 12
650 G=H
660 H=(U[I+12]-U[I])/((X[I+12]-X[I]))
670 V[I]=(H-G)*Z[I]*Z[I]
680 E=E+V[I]*(H-G)
690 NEXT I
700 G=V[N2]*(-H*Z[N2])/Z[N2]
710 E=E-G*H
720 G=F2
730 F2=E*P*P
740 IF F2 >= S OR F2 <= G THEN 880
750 F=Q
760 H=(V[M1]-V[N1])/((X[M1]-X[N1]))
770 FOR I=M1 TO M2 STEP 12
780 G=H
790 H=(V[I+12]-V[I])/((X[I+12]-X[I]))
800 G=H-G-S[I-12]*R[I-12]-Q[I-24]*R[I-24]
810 F=F+G*R[I]*G
820 R[I]=G
830 NEXT I
840 H=E-P*F
850 IF H <= 0 THEN 880
860 P=P+((S-F2)/((SQR(S/E)+P))*H)
870 GOTO 510
880 FOR I=M1 TO M2 STEP 12
890 A[I]=Y[I]-P*V[I]
900 C[I]=U[I]
910 NEXT I
920 FOR I=M1 TO M2 STEP 12
930 H=X[I+12]-X[I]

```

```
940 DC[I]=(CC[I+12]-CC[I])/(3*H)
950 BC[I]=(AC[I+12]-AC[I])/H-(H*DC[I]-CC[I])*H
960 NEXT I
970 C[N2]=0
980 SCALE 0,120,0,10000
990 XAXIS 0,1
1000 YAXIS 0,1000
1010 FOR I=1 TO 120
1020 X[I]=I
1030 NEXT I
1040 FOR I=1 TO 120
1050 PLOT X[I],H[I]
1060 NEXT I
1070 I1=I2-1
1080 FOR I=0 TO 9
1090 AC[12*I+I1]=AC[12*(I+1)+I2]
1100 CC[12*I+I1]=CC[12*(I+1)+I2]
1110 NEXT I
1120 FOR I=0 TO 8
1130 DC[12*I+I1]=DC[12*(I+1)+I2]
1140 BC[12*I+I1]=BC[12*(I+1)+I2]
1150 NEXT I
1160 DISP "VALORES V1 V2"
1170 WAIT 3000
1180 INPUT V1,V2
1190 FOR I=V1 TO V2 STEP 12
1200 FOR X=I TO I+11.99 STEP 0.1
1210 H=X-X[I]
1220 F=((DC[I]*H+CC[I])*H+BC[I])*H+AC[I]
1230 PLOT X,F
1240 NEXT X
1250 NEXT I
1260 FOR J=V1 TO V2+11 STEP 12
1270 FOR I=J TO J+11
1280 H=I-X[J]
1290 FC[I]=((DC[J]*H+CC[J])*H+BC[J])*H+AC[J]
1300 NEXT I
1310 NEXT J
1320 END
```

SPLINE POR MINIMOS CUADRADOS

250

ESTE PROGRAMA SE CORRIÓ EN PDP 11/34 DEL CENTRO DE INFORMATICA DE LA ENEP-ACATLAN

LOS COMENTARIOS DEL PROGRAMA PRINCIPAL Y DE LAS SUBROUTINAS SE PUEDEN CONSULTAR EN EL LIBRO 'A PRACTICAL GUIDE TO SPLINES' POR CARL DE BOOR

```

LPKMAX=200
NTMAX=200
LTKMAX=2000
INTEGER I, ICOUNT, II, J, K, L, LBEGIN, LNEW, LL, N, NT, NTIMES, NTAU
1 / ON, PRBCO, PRFUN, PRPCO
REAL ADDRK, BCOEF(100), BREAK, COEF, QTAU, Q(2000), SCRCH(200)
1 / T(200), TAU, TOTALW, WEIGHT
COMMON /DATA/ NTAU, TAU(200), QTAU(200), WEIGHT(200), TOTALW
COMMON /APPROX/BREAK(100), COEF(2000), L, K
DATA ON /ZMON /
ICOUNT=0
CALL ASSIGN(3, 'RESULT. DAT')
1 CALL SETDAT(ICOUNT)
CALL LZKNTS(BREAK, L, K, T, N)
TYPE 600
600 FORMAT('NTIMES - I3')
ACCEPT 300, NTIMES
300 FORMAT(I3)
ADDRK=0.0
PRBCO='ON'
PRPCO='ON'
PRFUN='ON'
TYPE 3000, NTIMES
3000 FORMAT(3X, I3)
LBEGIN=L
NT=1
10 CALL LZAPPR(T, N, K, Q, SCRCH, BCOEF)
IF(PRBCO .EQ. ON) WRITE(3, 609) (BCOEF(I), I=1, N)
609 FORMAT(//, 10X, 'COEFICIENTES B-SPLINES'/(5E20.10))
CALL BSPLPP(T, BCOEF, N, K, Q, BREAK, COEF, L)
WRITE(3, 610) K, L, (BREAK(LL), LL=2, L)
610 FORMAT(//, 10X, 'APROXIMACION POR SPLINES DE ORDEN ', I3, 'ON',
1 I3, 'INTERVALOS. PUNTOS DE QUIEBRE -',/(5E16.10))
IF(PRPCO .NE. ON) GOTO 15
WRITE(3, 611)
611 FORMAT(//, 10X, 'REPRESENTACION-PP PARA LA APROXIMACION ')
DO 12 I=1, L
II=(I-1)*K
12 WRITE(3, 613) BREAK(I), (COEF(II+J), J=1, K)
613 FORMAT(F9.3, 5E20.10/(11X, 5E20.10))
15 CALL LZERR(PRFUN, SCRCH, Q)
IF(NT .GE. NTIMES) GOTO 1
LNEW=LBEGIN+FLOAT(NT)*ADDRK
CALL NZWNT(BREAK, COEF, L, K, SCRCH, LNEW, T)
CALL LZKNTS(SCRCH, LNEW, K, T, N)
NT=NT + 1
GOTO 10
END

```

```

SUBROUTINE BCHFAC(W,NBANDS,NROW,DIAG)
INTEGER NBANDS,NROW,I,IMAX,J,JMAX,N
REAL W(NBANDS,NROW),DIAG(NROW),RATIO
IF(NROW.GT.1)
  IF(W(1,1).GT.0.) W(1,1)=1./W(1,1)
  RETURN
9   DO 10 N=1,NROW
10  DIAC(N)=W(1,N)
   DO 20 N=1,NROW
   IF(W(1,N)+DIAC(N).GT.DIAC(N))
14  DO 14 J=1,NBANDS
     W(J,N)=0.
   GOTO 20
15  W(1,N)=1./W(1,N)
   IMAX=MIN0(NBANDS-1,NROW-N)
   IF(IMAX.LT.1)
17  GOTO 20
   JMAX=IMAX
   DO 18 I=1,IMAX
     RATIO=W(I+1,N)*W(1,N)
17  DO 17 J=1,JMAX
     W(J,N+I)=W(J,N+I)-W(J+I,N)*RATIO
   JMAX=JMAX-1
18  W(I+1,N)=RATIO
20  CONTINUE
   RETURN
END

```

```

SUBROUTINE BSPLPP(T, BCOEF, N, K, SCRTCH, BREAK, COEF, L)
KMAX=20
INTEGER K, L, N, I, J, JP1, KMJ, LEFT, LSOFAR
REAL BCOEF(N), BREAK(1), COEF(K, 1), T(1), SCRTCH(K, K),
1 BIATX(20), DIFF, FKMJ, SUM
LSOFAR=0
BREAK(1)=T(K)
DO 50 LEFT=K, N
    IF(T(LEFT+1) .EQ. T(LEFT)) GOTO 50
    LSOFAR=LSOFAR + 1
    BREAK(LSOFAR+1)=T(LEFT+1)
    IF(K .GT. 1) GOTO 9
    COEF(1, LSOFAR)=BCOEF(LEFT) GOTO 50
9 DO 10 I=1, K
10 SCRTCH(I, 1)=BCOEF(LEFT-K+I)
DO 20 JP1=2, K
    J=JP1 - 1
    KMJ=K-J
    FKMJ=FLOAT(KMJ)
    DO 20 I=1, KMJ
        DIFF=T(LEFT+I) - T(LEFT+I -KMJ)
        IF(DIFF .GT. 0.) SCRTCH(I, JP1)=
            ((SCRTCH(I+1, J)-SCRTCH(I, J))/DIFF)*FKMJ
20 I CONTINUE
CALL BSPLVB(T, 1, 1, T(LEFT), LEFT, BIATX)
COEF(K, LSOFAR)=SCRTCH(1, K)
DO 30 JP1=2, K
    CALL BSPLVB(T, JP1, 2, T(LEFT), LEFT, BIATX)
    KMJ=K+1-JP1
    SUM=0.
    DO 28 I=1, JP1
28 SUM=BIATX(I)*SCRTCH(I, KMJ)+SUM
30 COEF(KMJ, LSOFAR)=SUM
90 CONTINUE
L=LSOFAR
RETURN
END

```

```

SUBROUTINE BSPLVB  JHIGH, DEX, LEFT, BIATX)
INTEGER INDEX, JHIGH, LEFT, I, JP
JMAX=20
REAL BIATX(JHIGH), X, DELTAL, DELTAR(20), SAVED, TERM
DATA J/1/

```

253

10

```

J=1
BIATX(1)=1.
IF(J .GE. JHIGH)
GOTO 99
JP1=J+1
DELTAR(J)=T(LEFT+J
DELTAL(J)=X-T(LEFT J)
SAVED=0.
DO 26 I=1,J
TERM=BIATX I/(DELT I)- DELTAL(JP1-I)
BIATX(I)=S I+DELT I+ERM
26 SAVED=DELTAL(JP1-I) TERM
BIATX(JP1)=SAVED
J=JP1
IF(J .LT. JHIGH)
GOTO 20
99 RETURN
END

```

20

26

99

END

```

SUBROUTINE BCHSLV(W, NBANDS, NROW, B)
  INTEGER NBANDS, NROW, J, JMAX, N, NBNDM1
  DOUBLE PRECISION B(NROW), W(NBANDS, NROW)
  IF(NROW .GT. 1)
    B(1)=B(1)*W(1,1)
    GOTO 21
  RETURN
21  NBNDM1=NBANDS - 1
    DO 30  N=1, NROW
      JMAX=MINO(NBNDM1, NROW-N)
      IF(JMAX .LT. 1)
        GOTO 30
      DO 25  J=1, JMAX
        B(J+N)=B(J+N) - W(J+1, N)*B(N)
25  CONTINUE
30  DO 40  N=NROW, 1, -1
      B(N)=B(N)*W(1, N)
      JMAX=MINO(NBNDM1, NROW-N)
      IF(JMAX .LT. 1)
        GOTO 40
      DO 35  J=1, JMAX
        B(N)=B(N) - W(J+1, N)*B(J+N)
35  CONTINUE
40  CONTINUE
    RETURN
  END

```

```

SUBROUTINE INTERV(XT,LXT,X,LEFT,MFLAG)
INTEGER LEFT,LXT,MFLAG,IHI,ILO,ISTEP,MIDDLE
REAL X,XT(LXT)
DATA ILO /1/
IHI=ILO + 1
IF(IHI .LT. LXT)
                                GOTO 20
    IF(X .GE. XT(LXT))
                                GOTO 110
    IF(LXT .LE. 1)
                                GOTO 90
    ILO=LXT - 1
    IHI=LXT
20  IF(X .GE. XT(IHI))
                                GOTO 40
    IF(X .GE. XT(ILO))
                                GOTO 100
    ISTEP=1
31  IHI=ILO
    ILO=IHI - ISTEP
    IF(ILO .LE. 1)
                                GOTO 35
    IF(X .GE. XT(ILO))
                                GOTO 50
    ISTEP=ISTEP*2
                                GOTO 31
35  ILO=1
    IF(X .LT. XT(1))
                                GOTO 90
                                GOTO 50
40  ISTEP=1
41  ILO=IHI
    IHI=ILO + ISTEP
    IF(IHI .GE. LXT)
                                GOTO 45
    IF(X .LT. XT(IHI))
                                GOTO 50
    ISTEP=ISTEP*2
                                GOTO 41
45  IF(X .GE. XT(LXT))
                                GOTO 110
    IHI=LXT
50  MIDDLE=(ILO + IHI)/2
    IF(MIDDLE .EQ. ILO)
                                GOTO 100
    IF(X .LT. XT(MIDDLE))
                                GOTO 53
    ILO=MIDDLE
                                GOTO 50
53  IHI=MIDDLE
                                GOTO 50
                                GOTO 50
90  MFLAG=-1
    LEFT=1
                                RETURN
100 MFLAG=0
    LEFT=ILO
                                RETURN
110 MFLAG=1
    LEFT=LXT
                                RETURN
END

```

```

SUBROUTINE LZAPPR(T, N, K, Q, DIAQ, BCOEF)
INTEGER K, N, I, J, JJ, LEFT, LEFTMK, LL, MM, NTAU
REAL BCOEF(N), DIAQ(N), Q(K, N), T(1), DW, QTAU, TAU, WEIGHT
KMAX=20
NTMAX=200
REAL BIATX(20)
COMMON /DATA/ NTAU, TAU(200), QTAU(200), WEIGHT(200)
DO 7 J=1, N
    BCOEF(J)=0.
    DO 7 I=1, K
        Q(I, J)=0.
7
    LEFT=K
    LEFTMK=0
    DO 20 LL=1, NTAU
        IF(LEFT .EQ. N) GOTO 15
        IF(TAU(LL) .LT. T(LEFT+1)) GOTO 15
        LEFT=LEFT + 1
        LEFTMK=LEFTMK + 1
        GOTO 10
15
    CALL BSPLVB(T, K, 1, TAU(LL), LEFT, BIATX)
    DO 20 MM=1, K
        DW=BIATX(MM)*WEIGHT(LL)
        J=LEFTMK + MM
        BCOEF(J)=DW+QTAU(LL) + BCOEF(J)
        I=1
        DO 20 JJ=MM, K
            Q(I, J)=BIATX(JJ)*DW+Q(I, J)
20
        I=I+1
    CALL BCFAC(Q, K, N, DIAQ)
    CALL BCBVLV(Q, K, N, BCOEF)
RETURN
END

```

```

SUBROUTINE LZERR (PRFUN, FTAU, ERROR)
  INTEGER PRFUN, IE, K, L, LL, NTAU, ON
  REAL FTAU(1), ERROR(1), BREAK, COEF, ERR, ERRMAX, ERRL1, ERRL2
  1      , GTAU, SCALE, TAU, TOTALW, WEIGHT, TENDEN(120)
  LPKMAX=200
  NTKMAX=200
  LTKMAX=2000
  COMMON /DATA/NTAU, TAU(200), GTAU(200), WEIGHT(200), TOTALW
  COMMON /APPROX/BREAK(100), COEF(2000), L, K
  CALL ASSIGN(2, 'TENDENCIA. DAT')
  DATA ON /2NDON /
  ERRL1=0.
  ERRL2=0.
  ERRMAX=0.
  ECM=0.
  DO 10 LL=1, NTAU
    FTAU(LL)=PVALU (BREAK, COEF, L, K, TAU(LL), 0)
    READ(2, *) TENDEN(LL)
    ERROR(LL)=TENDEN(LL) - FTAU(LL)
    ECM=ECM + (TENDEN(LL) - FTAU(LL))**2
    ERR=ABS(ERROR(LL))
    IF(ERRMAX .LT. ERR) ERRMAX=ERR
    ERRL1=ERRL1+ERR*WEIGHT(LL)
    ERRL2=ERRL2+ERR**2*WEIGHT(LL)
  10 ERRL1=ERRL1/TOTALW
    ERRL2=SQRT(ERRL2/TOTALW)
    TYPE 615, ERRL2, ERRL1, ERRMAX
    FORMAT(///, 'ERROR MINIMO CUADRADO=', E20.6/
    1 , 'ERROR PROMEDIO=', E20.6/
    2 , 'ERROR MAXIMO =', E20.6//)
    IF(PRFUN .NE. ON) RETURN
    IE=0
    SCALE=1.
    IF(ERRMAX .GE. 10.) GOTO 18
    DO 17 IE=1, 9
      SCALE=SCALE*10.
      IF(ERRMAX*SCALE .GE. 10.) GOTO 18
  17 CONTINUE
  18 DO 19 LL=1, NTAU
    ERROR(LL)=ERROR(LL)*SCALE
    ECM=SQRT(ECM/120)
    TYPE 630, ECM
  630 FORMAT(/, 9X, 'ERROR CUADRATICO MEDIO TOTAL=', F19.2)
    ECM=0.
    DO 20 I=1, 12
      ECM=ECM+(TENDEN(I)-FTAU(I))**2
  20 ECM=SQRT(ECM/12)
    TYPE 700, ECM
  700 FORMAT(9X, 'E. C. M. INICIAL=', F19.2)

```

ECM=0.

DO 21 I=13, 108

ECM=ECM+(TENDEN(I)-FTAU(I))*2

ECM=SQRT(ECM/96)

TYPE 710, ECM

258

FORMAT(5X, 'E. C. M. CENTRAL=', F15.2)

ECM=0.

DO 22 I=109, 120

ECM=ECM+(TENDEN(I)-FTAU(I))*2

ECM=SQRT(ECM/12)

TYPE 720, ECM

FORMAT(5X, 'E. C. M. FINAL=', F15.2)

TYPE 620, IE, (LL, TENDEN(LL), FTAU(LL), ERROR(LL), LI=1, NTAU)

FORMAT(//14X, 'APROXIMACION Y ERROR DE CURVATURA EN ESCALA', /7X

1, 'DATO', /7X, 'APROXIMACION', /3X, 'DESVIACION X 10**', I1/

2 (I4, F16.8, F16.8, F17.6))

CALL CLOSE(2)

RETURN

END

SUBROUTINE L2KNTS(BREAK, L, K, T, N)

INTEGER K, L, N, I, KM1

REAL BREAK(1), T(1)

259

KM1=K-1

DO 5 I=1, KM1

5 T(I)=BREAK(I)

DO 6 I=1, L

6 T(KM1+I)=BREAK(I)

N=KM1 + L

DO 7 I=1, K

7 T(N+I)=BREAK(L+I)

RETURN

END

```

SUBROUTINE NEWNOT(BREAK, COEF, L, K, BRKNEW, LNEW, COEFG)
  INTEGER K, L, LNEW, I, IPRINT, J
  REAL BREAK(1), BRKNEW(1), COEF(K, L), COEFG(2, L), DIF, DIFPRV, 260
  1 ONEOVK, STEP, STEPI
  DATA IPRINT /0/
  BRKNEW(1)=BREAK(1)
  BRKNEW(LNEW + 1)=BREAK(L+1)
  IF(L .LE. 1)                                GOTO 90
  ONEOVK=1./FLOAT(K)
  COEFG(1, 1)=0.
  DIFPRV=ABS(COEF(K, 2)-COEF(K, 1))/(BREAK(3)-BREAK(1))
  DO 10 I=2, L
    DIF=ABS(COEF(K, I) - COEF(K, I-1))/(BREAK(I+1)-BREAK(I-1))
    COEFG(2, I-1)=(DIF+DIFPRV)*ONEOVK
    COEFG(1, I)=COEFG(1, I-1)+COEFG(2, I-1)*(BREAK(I)-BREAK(I-1))
    DIFPRV=DIF
  10 COEFG(2, L)=(2.*DIFPRV)+ONEOVK
  STEP=(COEFG(1, L)+COEFG(2, L)*(BREAK(L+1)-BREAK(L)))/FLOAT(LNEW+1)
  IF(IPRINT .GT. 0) TYPE 600, STEP, (I, COEFG(1, I), COEFG(2, I), I=1, L)
  600 FORMAT(5X, 'STEP=', E16.7/(15, 2E16.5))
  IF(STEP .LE. 0.)                                GOTO 90
  J=1
  DO 30 I=2, LNEW
    STEPI=FLOAT(I-1)*STEP
    21 IF(J .EQ. L)                                GOTO 27
    IF(STEPI .LE. COEFG(1, J+1)) GOTO 27
    J=J+1
    27 IF(COEFG(2, J) .EQ. 0.)                      GOTO 21
    BRKNEW(I)=BREAK(J)+(STEPI-COEFG(1, J))/COEFG(2, J)
    29 GOTO 30
  30 BRKNEW(I)=(BREAK(J)+BREAK(J+1))/2.
  CONTINUE
  RETURN
  90 STEP=(BREAK(L+1) - BREAK(1))/FLOAT(LNEW)
  DO 93 I=2, LNEW
    93 BRKNEW(I)=BREAK(1)+FLOAT(I-1)*STEP
  RETURN
  END

```

REAL FUNCTION PPVALU(BREAK, COEF, L, K, X, JDERIV)

INTEGER JDERIV, K, L, I, M, NDUMMY

261

REAL BREAK(L), COEF(K, L), X, FMMJDR, H

PPVALU=0

FMMJDR=K - JDERIV

IF(FMMJDR .LE. 0.)

GO TO 99

CALL INTERV (BREAK, L, X, I, NDUMMY)

H=X - BREAK(I)

DO 10 M=K, JDERIV+1, -1

PPVALU=(PPVALU/FMMJDR)*H + COEF(M, I)

10 FMMJDR=FMMJDR - 1.

99

RETURN

END

APROXIMACION POR SPLINES MINIMOS CUADRADOS DE ORDEN CUATRO
EN TRES INTERVALOS

262

COEFICIENTES B-SPLINES :

0.6929155884E+03 0.1574598340E+05 0.3464114666E+00
-0.64193004013E+00 -1.3866278231E+03 0.0000000000E+00

PARAMETROS DE LOS POLINOMIOS CUBICOS :

1 - 0.6929155884E+03 0.9031840625E+05 -0.4872387187E+06
 0.1058451000E+07
2 - 0.6996341309E+04 -0.2099457812E+05 0.4198678516E+05
 -0.419852140E+05
3 - -0.3804363906E+00 -0.4015281796E+00 0.1569964124E+01
 -0.1510452987E-01

ERROR CUADRATICO MEDIO :

E. C. M. TOTAL = 844.51
E. C. M. INICIAL = 844.51
E. C. M. CENTRAL = 395.64
E. C. M. FINAL = 103.71

DATO	APROXIMACION	SERIE DE TIEMPO	DESVIACION
1	908.5	692.9	295.6
2	978.0	874.7	103.3
3	968.5	-0.4	968.9
4	960.0	0.8	959.2
5	952.5	3.5	949.0
6	946.0	7.7	938.3
7	940.5	13.5	927.0
8	936.0	20.7	915.3
9	932.5	29.5	903.0
10	930.0	39.7	890.3
11	928.5	51.3	877.2
12	928.0	64.4	863.6
13	928.5	79.0	849.5
14	930.0	94.9	835.1
15	932.5	112.3	820.2
16	936.0	131.0	805.0
17	940.5	151.1	789.4
18	946.0	172.5	773.5
19	952.5	195.3	757.2
20	960.0	219.3	740.7
21	968.5	244.7	723.0

21	700.0	270.7	723.8
22	978.0	271.4	706.6
23	988.5	299.4	689.1
24	1000.0	328.6	671.4
25	1012.5	359.1	653.4
26	1026.0	390.7	635.3
27	1040.5	423.6	616.9
28	1056.0	457.7	598.3
29	1072.5	493.0	579.5
30	1090.0	529.5	560.5
31	1108.5	567.1	541.4
32	1128.0	605.8	522.2
33	1148.5	645.7	502.8
34	1170.0	686.7	483.3
35	1192.5	728.8	463.7
36	1216.0	771.9	444.1
37	1240.5	816.1	424.4
38	1266.0	861.4	404.6
39	1292.5	907.7	384.8
40	1320.0	955.0	365.0
41	1348.5	1003.3	345.2
42	1378.0	1052.6	325.4
43	1408.5	1102.9	305.6
44	1440.0	1154.1	285.9
45	1472.5	1206.3	266.2
46	1506.0	1259.4	246.6
47	1540.5	1313.4	227.1
48	1576.0	1368.3	207.7
49	1612.5	1424.1	188.4
50	1650.0	1480.7	169.3
51	1688.5	1538.2	150.3
52	1728.0	1596.6	131.4
53	1768.5	1655.7	112.8
54	1810.0	1715.7	94.3
55	1852.5	1776.5	76.0
56	1896.0	1838.0	58.0
57	1940.5	1900.3	40.2
58	1986.0	1963.4	22.6
59	2032.5	2027.1	5.4
60	2080.0	2091.6	-11.6
61	2128.5	2156.8	-28.3
62	2178.0	2222.7	-44.7
63	2228.5	2289.2	-60.7
64	2280.0	2356.4	-76.4
65	2332.5	2424.3	-91.8
66	2386.0	2492.8	-106.8
67	2440.5	2561.8	-121.3
68	2496.0	2631.5	-135.5
69	2552.5	2701.8	-149.3
70	2610.0	2772.6	-162.6
71	2668.5	2843.9	-175.4
72	2728.0	2915.8	-187.3

72	2720.0	2710.0	107.0
73	2788.5	2903.3	-199.8
74	2850.0	3061.2	-211.2
75	2912.5	3134.6	-222.1
76	2976.0	3208.5	-232.5
77	3040.5	3282.8	-242.3
78	3106.0	3357.6	-251.6
79	3172.5	3432.8	-260.3
80	3240.0	3508.4	-268.4
81	3308.5	3584.5	-276.0
82	3378.0	3660.9	-282.9
83	3448.5	3737.7	-289.2
84	3520.0	3814.8	-294.8
85	3592.5	3892.3	-299.8
86	3666.0	3970.1	-304.1
87	3740.5	4048.2	-307.7
88	3816.0	4126.6	-310.6
89	3892.5	4205.2	-312.7
90	3970.0	4284.2	-314.2
91	4048.5	4363.3	-314.8
92	4128.0	4442.8	-314.8
93	4208.5	4522.4	-313.9
94	4290.0	4602.2	-312.2
95	4372.5	4682.2	-309.7
96	4456.0	4762.4	-306.4
97	4540.5	4842.8	-302.3
98	4626.0	4923.3	-297.3
99	4712.5	5003.9	-291.4
100	4800.0	5084.6	-284.6
101	4888.5	5165.4	-276.9
102	4978.0	5246.3	-268.3
103	5068.5	5327.3	-258.8
104	5160.0	5408.3	-248.3
105	5252.5	5489.4	-236.9
106	5346.0	5570.5	-224.5
107	5440.5	5651.5	-211.0
108	5536.0	5732.6	-196.6
109	5632.5	5813.7	-181.2
110	5730.0	5894.7	-164.7
111	5828.5	5975.6	-147.1
112	5928.0	6056.5	-128.5
113	6028.5	6137.3	-108.8
114	6130.0	6218.0	-88.0
115	6232.5	6298.6	-66.1
116	6336.0	6379.0	-43.0
117	6440.5	6459.3	-18.8
118	6546.0	6539.5	6.5
119	6652.5	6619.5	33.0
120	6760.0	6699.3	60.7

APENDICE C

En este apéndice se incluye material correspondiente al capítulo tres:

- a) La tabulación de la serie de tiempo simulada, para cada punto y para cada componente.
- b) El programa que hace posible la generación de tal serie.

* Este programa se corrió en la microcomputadora HP-9830.

VALORES DE LAS COMPONENTES DE LA SERIE DE TIEMPO

N	U	T	E	I
1	692.9155965	988.5	0.6999	1.001538549
2	874.7084269	978	0.8	1.117981118
3	845.8674060	968.5	0.9	0.970420933
4	1025.406253	960	1.1	0.971028649
5	1130.725043	952.5	1.2	0.989260755
6	1068.995716	946	1.12	1.008943405
7	994.9056424	940.5	1.1	0.961679612
8	862.7556130	936	0.896	1.028735999
9	891.7294510	932.5	0.97	0.985853847
10	826.4908577	930	0.9	0.987444274
11	1109.288805	928.5	1.2	0.995592178
12	1193.971726	928	1.3	0.989698048
13	654.217477	926.5	0.6999	1.00670967
14	685.8966110	930	0.8	0.895822327
15	799.9436751	932.5	0.9	0.953164939
16	1043.569242	936	1.1	1.01356764
17	1066.352542	940.5	1.2	0.944845421
18	1081.18045	946	1.12	1.020443644
19	1099.717254	952.5	1.1	1.049598906
20	904.823959	960	0.896	1.051925176
21	955.9737356	968.5	0.97	1.017594149
22	857.8340326	978	0.9	0.974589903
23	1163.773565	988.5	1.2	0.981093883
24	1270.887438	1000	1.3	0.977605722
25	731.5727557	1012.5	0.6999	1.032348897
26	864.463274	1026	0.8	1.053195997
27	886.1363862	1040.5	0.9	0.946271970
28	1173.209480	1056	1.1	1.009994387
29	1324.153310	1072.5	1.2	1.02886815
30	1298.66274	1090	1.12	1.063780096
31	1213.150146	1108.5	1.1	0.994915443
32	986.1413505	1128	0.896	0.975712931
33	1170.196242	1148.5	0.97	1.050403029
34	1043.070423	1170	0.9	0.9905778
35	1380.89034	1192.5	1.2	0.964988358
36	1560.872588	1216	1.3	0.987394097
37	853.9452492	1240.5	0.6999	0.983551861
38	1132.549901	1266	0.8	1.118236474
39	1058.454668	1292.5	0.9	0.909911500
40	1450.048262	1320	1.1	0.998655828
41	1642.921505	1348.5	1.2	1.015277163
42	1536.299159	1378	1.12	0.995425020
43	1534.101271	1406.5	1.1	0.990157983
44	1343.599519	1440	0.896	1.041356274
45	1360.320323	1472.5	0.97	0.952388513

46	1420.983166	1506	0.9	1.048386577
47	1957.554153	1540.5	1.2	1.058938739
48	2064.379012	1576	1.3	1.007603969
49	1239.575284	1612.5	0.6999	1.098340989
50	1293.068038	1650	0.8	0.979590938
51	1661.725672	1688.5	0.9	1.093492364
52	1858.190721	1728	1.1	0.977583502
53	2020.912916	1768.5	1.2	0.952272602
54	1954.798968	1810	1.12	0.964285205
55	1919.090969	1852.5	1.1	0.941769584
56	1671.685662	1896	0.896	0.984029855
57	1894.527112	1940.5	0.97	1.006503657
58	1674.412029	1986	0.9	0.93678641
59	2337.956852	2032.5	1.2	0.958571895
60	2732.988879	2080	1.3	1.010720739
61	1515.516290	2128.5	0.6999	1.017304489
62	1785.766515	2178	0.8	1.024888955
63	2063.309679	2228.5	0.9	1.026748625
64	2569.065924	2280	1.1	1.024348455
65	2719.994051	2332.5	1.2	0.971773509
66	2623.209202	2386	1.12	0.981622411
67	2750.107047	2440.5	1.1	1.024420125
68	2384.177850	2496	0.896	1.066070825
69	2409.056286	2552.5	0.97	0.972992431
70	2297.798752	2610	0.9	0.978202960
71	3188.93045	2668.5	1.2	0.995856115
72	3666.963953	2728	1.3	1.033996152
73	2010.571350	2788.5	0.6999	1.030179367
74	2249.877438	2850	0.8	0.986788350
75	2681.892503	2912.5	0.9	1.023134956
76	2982.971862	2976	1.1	0.911220632
77	3532.752623	3040.5	1.2	0.968248814
78	3448.576711	3106	1.12	0.991334948
79	3214.879849	3172.5	1.1	0.921235002
80	3167.213979	3240	0.896	1.090999083
81	3210.488018	3308.5	0.97	1.000387324
82	3070.812278	3378	0.9	1.010069166
83	4410.461454	3448.5	1.2	1.065792242
84	4389.407224	3520	1.3	0.959223607
85	2490.997727	3592.5	0.6999	0.990696346
86	3028.689253	3666	0.8	1.032695463
87	3559.201917	3740.5	0.9	1.057256729
88	4081.041052	3816	1.1	0.972232002
89	4755.11496	3892.5	1.2	1.018007913
90	4601.337229	3970	1.12	1.034845545
91	4355.515949	4048.5	1.1	0.978031358
92	3554.185493	4128	0.896	0.960931415

93	4087.236684	4208.5	0.97	1.001222779
94	3823.761124	4290	0.9	0.990355121
95	5428.801095	4372.5	1.2	1.034648579
96	6232.648407	4456	1.3	1.07592019
97	3079.431089	4540.5	0.6999	0.969015707
98	3494.534640	4626	0.8	0.944264656
99	4231.07265	4712.5	0.9	0.997600389
100	5297.057484	4800	1.1	1.003230584
101	5914.949718	4888.5	1.2	1.008316272
102	5862.937468	4978	1.12	1.051580072
103	5760.699203	5068.5	1.1	1.033244407
104	4857.144664	5160	0.896	1.050565966
105	4759.908443	5252.5	0.97	0.934245046
106	4729.720676	5346	0.9	0.983023793
107	6689.200660	5440.5	1.2	1.024599556
108	7382.970179	5536	1.3	1.025862466
109	3747.688599	5632.5	0.6999	0.950662370
110	4355.167096	5730	0.8	0.950020082
111	5283.511437	5828.5	0.9	1.007217683
112	6433.206616	5920	1.1	0.98656708
113	7699.539161	6028.5	1.2	1.064324896
114	6983.369112	6130	1.12	1.017153506
115	6573.55814	6232.5	1.1	0.958838660
116	5571.198669	6336	0.896	0.981353481
117	6537.681906	6440.5	0.97	1.046483697
118	6121.081160	6546	0.9	1.038985837
119	8148.921196	6652.5	1.2	1.020764316
120	9328.808778	6760	1.3	1.06153946

```

10 REM SIMULACION DE LA SERIE DE TIEMPO HIPOTETICA
20 DIM U(120),T(120),E(120),I(256),X(120)
30 FOR I=1 TO 120
40 X(I)=I
50 NEXT I
60 FOR X=1 TO 120
70 T(X)=0.5*X^2-12*X+1000
80 NEXT X
90 GOSUB 570
100 E[1]=E[13]=E[25]=E[37]=E[49]=E[61]=E[73]=E[85]=E[97]=E[109]=0.6999
110 E[2]=E[14]=E[26]=E[38]=E[50]=E[62]=E[74]=E[86]=E[98]=E[110]=0.8
120 E[3]=E[15]=E[27]=E[39]=E[51]=E[63]=E[75]=E[87]=E[99]=E[111]=0.9
130 E[4]=E[16]=E[28]=E[40]=E[52]=E[64]=E[76]=E[88]=E[100]=E[112]=1.1
140 E[5]=E[17]=E[29]=E[41]=E[53]=E[65]=E[77]=E[89]=E[101]=E[113]=1.2
150 E[6]=E[18]=E[30]=E[42]=E[54]=E[66]=E[78]=E[90]=E[102]=E[114]=1.12
160 E[7]=E[19]=E[31]=E[43]=E[55]=E[67]=E[79]=E[91]=E[103]=E[115]=1.1
170 E[8]=E[20]=E[32]=E[44]=E[56]=E[68]=E[80]=E[92]=E[104]=E[116]=0.896
180 E[9]=E[21]=E[33]=E[45]=E[57]=E[69]=E[81]=E[93]=E[105]=E[117]=0.97
190 E[10]=E[22]=E[34]=E[46]=E[58]=E[70]=E[82]=E[94]=E[106]=E[118]=0.9
200 E[11]=E[23]=E[35]=E[47]=E[59]=E[71]=E[83]=E[95]=E[107]=E[119]=1.2
210 E[12]=E[24]=E[36]=E[48]=E[60]=E[72]=E[84]=E[96]=E[108]=E[120]=1.3
220 GOSUB 690
230 FOR J=1 TO 256
240 U=RNDO.999999999999
250 V=RNDO.12541254129
260 I[J]=(SQR(-2*LOG(U))*SIN(2*PI*V))
270 NEXT J
280 GOSUB 810
285 GOTO 920
290 FOR X=1 TO 120
295 R(X)=I(X+136)/25
300 UC(X)=T(X)*E(X)*EXP(I(X+136)/25)
310 NEXT X
312 STORE DATA 2,U
315 STORE DATA 5,R
320 GOSUB 450
330 PRINT TAB1"n"TAB22"U"TAB33"T"TAB49"E"TAB66"J"
340 PRINT
350 FOR I=1 TO 120
360 PRINT I,TAB3,UC(I),TAB16,T(I),TAB22,E(I),TAB31,EXP(I(I+136)/25)
370 NEXT I
380 SCALE 0,120,0,10000
390 XAXIS 0,1
400 YAXIS 0,1000
410 FOR I=1 TO 120
420 PLOT X(I),UC(I)

```

```
430 NEXT I
440 END
450 M=0
460 FOR I=1 TO 120
470 M=M+U(I)
480 NEXT I
490 M=M/120
500 S=0
510 FOR I=1 TO 120
520 S=S+(U(I)-M)^2
530 NEXT I
540 S=S/119
550 PRINT "MEDIA DE U";M;"VARIANZA DE U";S
560 RETURN
570 M=0
580 FOR I=1 TO 120
590 M=M+T(I)
600 NEXT I
610 M=M/120
620 S=0
630 FOR I=1 TO 120
640 S=S+(T(I)-M)^2
650 NEXT I
660 S=S/119
670 PRINT "MEDIA DE T";M;"VARIANZA DE T";S
680 RETURN
690 M=0
700 FOR I=1 TO 120
710 M=M+E(I)
720 NEXT I
730 M=M/120
740 S=0
750 FOR I=1 TO 120
760 S=S+(E(I)-M)^2
770 NEXT I
780 S=S/119
790 PRINT "MEDIA DE E";M;"VARIANZA DE E";S
800 RETURN
810 M=0
820 FOR I=1 TO 120
830 M=M+I[(I+136)]
840 NEXT I
845 M=M/25
850 S=0
860 FOR I=1 TO 120
870 S=S+(I[(I+136)]-M)^2
```

```
880 NEXT I
890 S=S/119
900 PRINT "MEDIA DE I";M;"VARIANZA DE I";S
910 RETURN
920 M=0
930 FOR I=1 TO 120
940 M=M+EXP((I+136)/25)
950 NEXT I
960 M=M/120
970 FOR I=1 TO 120
980 S=S+(EXP((I+136)/25)-M)^2
990 NEXT I
1000 PRINT "MEDIA DEL ERROR DE LA SERIE "M
1010 PRINT "VARIANZA DEL ERROR DE LA SERIE "S
1020 RETURN
```

BIBLIOGRAFIA

AHLBER , J. H.,; NILSON, E. N. & WALSH, J. L.
The Theory of Splines and Their Applications
Academic Press, New York
1977,

APOSTOL, TOM M.
Mathematical Analysis
1974, Second edition, pag. 322.

BOOR, CARL DE
A Practical Guide to Splines
Applied Mathematics Science 27
1978.

BOX - JENKINS

Time Series Analysis Forecasting and Control
1969.

COURANT

Introducción al Cálculo y Análisis Matemático
1971.

COWDEN, D. J.

Weights for Fitting Polynomial Secular Trends
Technical Paper No. 4
School of Business Administration
University of North Carolina
1962.

DRAPPER , NORMAN & SMITH HARRY

Applied Regression Analysis
1966.

DUFFIN, R. T. & KARLOVITZ, L. A.

Formulation of Linear Programs in Analysis Approximation
Theory
University of Maryland,
Technical Note BN - 486
1967

DURBIN, J. & MURPHY, M. J.

Seasonal adjustment based on mixed additive multiplicative model

1965.

FORSYTHE, G. E. Y MOLER, L. B.

Computer Solutions of Linear Algebraic Systems

Prentice - Hall, Englewood Cliffs, New Jersey

1967

FUNK, P.

Variationsrechnung und ihre Anwendung in Physik und --
Technik

Berlin-Go Hingen Heildelberg: Springer

1962

HAYES, J. G.

Numerical Approximation to Functions and Data

The Athlone Press of University of London

1970

HILDEBRAND, F. B.

Introduction to Numerical Analysis

Mac Graw-Hill Inc., New York

1974

HUOT GUY

Stochastic Simulation of Time Series Components

1978.

JOHNSTON

Econometrics Methods

Second Edition

1972.

JONES, R. C. & KARLOVITZ, L. A.

Iterative Construction of Constrained Tschebycheff Approximation of Continuous Functions

University of Maryland

Technical Note BN - 491

1967.

KENDALL, MAURICE

Time Series

Second Edition

1976.

ROW-MILLER

Applied Econometrics

1971.

RUDIN, B.D.

Convex Polynomial Approximation

Stanford University

Technical Report CS23

1967.

SPATH, H.

Spline Algorithms for Curves and Surfaces

Utilitas Mathematica

Publ. Inc. Winnipeg, Manitoba

1974.

SARD, A. & WEINTRAUB, S.

A Book of Splines

Wiley, New York

1971.

SCHUMAKER LARRY, L.

Toward a Constructive Theory of Generalized Spline
Functions

Vol. 501

Springer-Verlog

1915.

STEPHEN, P. SHAO

Estadística para Economistas Y Administradores de
Empresas

Herrero Hermanos

México

1960.

STEPHENSON, JAMES A, & FARR, HELEN T,

Seasonal adjustment of Economic Data by Application of
the General Linear Statistical Model.

Journal of America Statistical Association

Vol. 67 No. 337

March 1972.

THOMOPOULOS, NICK T.

Applied Forecasting Methods

1980.

VARGA, P. S.

Functional Analysis and Approximation Theory Numerical
Analysis

SIAM, Philadelphia CBMS vol. 3

1971.

REINSCH, C.H.

Smoothing by Spline Functions

Num. Math. 10

177-183 (1967)