

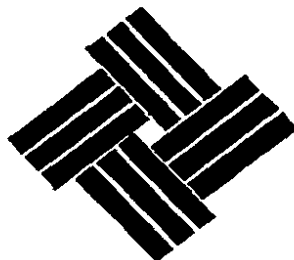
881201

7
23

UNIVERSIDAD ANAHUAC

ESCUELA DE ACTUARIA

Con estudios incorporados a la Universidad Nacional Autónoma de México



EL EMPLEO DE TRANSFORMACIONES POTENCIA PARA ANALIZAR SERIES DE TIEMPO

TESIS CON
FALLA DE ORIGEN

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
MARIA DEL CARMEN MINOR MOLINA

MEXICO, D. F.

1989.



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

INDICE

	Pag.
Introducción.....	1
Capítulo 1.- Conceptos básicos e introducción al problema.....	6
1.1.- Introducción al estudio de las series de tiempo.....	7
1.2.- Modelos para representar series de tiempo como procesos estocásticos.....	8
1.3.- Importancia de la validez de los supuestos.....	11
1.4.- Transformaciones y problema inicial.....	17
1.5.- Revisión de la literatura existente.....	23
Capítulo 2.- Diferentes usos de las transformaciones en el análisis de series de tiempo.....	29
2.1.- Transformaciones para estabilizar la varianza.....	31
2.2.- Transformaciones para normalizar la distribución de los errores.....	33
2.3.- Transformaciones para seleccionar modelos en desestacionalización de series.....	35
Capítulo 3.- Presentación y comparación de los diferentes métodos para elegir una transformación potencia.....	40
3.1.- Método de Box y Cox.....	42
3.2.- Método de la gráfica recorrido-media.....	47
3.3.- Método de Guerrero.....	50
3.4.- Comparaciones Teóricas.....	54
Capítulo 4.- Presentación y comparación de los diferentes métodos para corregir pronósticos por sesgo.....	60
4.1.- Método de Neyman y Scott.....	62
4.2.- Método de Granger y Newbold.....	74
4.3.- Método de Miller.....	84
4.4.- Método de Taylor.....	91
4.5.- Método de Guerrero.....	95
4.6.- Comparaciones teóricas.....	99
Capítulo 5.- Algunas aplicaciones empíricas.....	103
5.1.- Caso Chatfield y Prothero.....	105
5.2.- Caso "Índice de Precios al Consumidor".....	109
Conclusiones.....	117
Apéndice.....	122
Bibliografía.....	135

INTRODUCCION

Afortunadamente en estos últimos años, en México se ha manifestado un desarrollo sin precedente en el área de Series de Tiempo. Estadísticos, Ingenieros, Economistas e Investigadores de Operaciones con una gran variedad de necesidades y objetivos en mente, han desarrollado nuevas metodologías para el análisis de estas series. El resultado de estos trabajos, no es más que conocimientos teóricos con considerable valor práctico y que pueden ser usados en diferentes campos.

No está por demás decir que estas metodologías han tenido su cuna en países altamente desarrollados, sin embargo en México nos hemos dado cuenta de la vital importancia de los resultados que este tipo de análisis ofrece, ya que con ellos es posible tener un buen control y una mejor planeación en todas las áreas económicas. Un ejemplo de esto puede observarse en algunas compañías Mexicanas que se dedican a la venta de un cierto producto y cuya preocupación está en pronosticar correctamente sus ventas, para así poder disminuir su inventario e incrementar sus ganancias.

El análisis de series de tiempo es muy amplio y existen muchas y variadas metodologías para llevarse a cabo. Específicamente, este trabajo estará basado en la metodología desarrollada por George E.P. Box y Gwilym M. Jenkins publicado en su libro "Time Series Analysis: Forecasting and Control", en 1970.

Como en toda nueva área por desarrollar, hay siempre

discordancia de opiniones entre los estudiosos del tema, sobre lo que es importante, útil y de fácil aplicación. En específico no hay una opinión generalizada sobre qué aproximación es la mejor, o que metodología es la más precisa, con menor error o costo y que además sea la menos compleja para su implementación.

Como ejemplo de dichas discordancias está el artículo escrito por C. Chatfield y D.L. Prothero (1973) titulado "Box-Jenkins Seasonal Forecasting: Problems in a Case-Study", en donde dichos autores aplican la metodología propuesta por Box y Jenkins a una serie de Ventas de la compañía X. Dentro de sus conclusiones, las más importantes para efectos de esta tesis, fueron: a) que dicha metodología no es eficiente para analizar este tipo de series, b) que la transformación utilizada para volver estacionaria la serie no fue la adecuada c) y que dicha transformación introduce un cierto sesgo por lo cual se obtienen pronósticos muy pobres.

Este artículo despertó gran interés entre la comunidad estadística, generando discusiones sobre la metodología seguida y planteando la interrogante de ¿cómo seleccionar la transformación adecuada para la serie en estudio? y poder obtener pronósticos óptimos.

Cuando esta discusión se suscitó, pocos eran los trabajos escritos para resolver esta interrogante. De hecho no existían metodologías formales y las disponibles, no eran fáciles de implementar, más aún, actualmente no son muchos los métodos existentes para resolver dicho problema.

El objetivo de esta tesis, es el de mostrar un método

para seleccionar el índice λ (lambda) de las transformaciones potencia aplicadas a las series de tiempo. También se presenta una técnica para la corrección de sesgo de los pronósticos que introduce la aplicación de la transformación inversa. Los métodos que serán presentados fueron desarrollados por Víctor Guerrero del Banco de México.

El interés de este trabajo surge de la necesidad de tener métodos prácticos y de relativamente fácil aplicación y cálculo, pretendiendo mostrar que estos métodos son de gran utilidad y que permiten realizar un buen análisis de las series de tiempo.

Para poder llevar a cabo esta investigación se presentarán los métodos más conocidos, como son el método de Box y Cox y el de la Gráfica Recorrido-Media para seleccionar el parámetro λ y los de Granger y Newbold, Neyman y Scott, Taylor y Miller para corrección de sesgo, haciendo una comparación práctica con los métodos propuestos por Guerrero que serán introducidos en este trabajo.

Uno de los problemas que se presentó al desarrollar esta tesis, fué el de la revisión y condensación de la literatura existente. Desgraciadamente en México son pocas las Instituciones o Universidades que cuentan con revistas científicas de esta rama, esto es quizá, debido a que el desarrollo de estos métodos se ha hecho principalmente en los países desarrollados haciendo su acceso un poco difícil, sin embargo, la mayoría de la literatura manejada en esta tesis se obtuvo de la biblioteca del Banco de México. Otros artículos interesantes se encontraron en la Universidad Anáhuac y los más recientes en algunas

Universidades de Estados Unidos.

Para llevar a cabo las comparaciones prácticas de los métodos, primeramente se recrearon los resultados obtenidos por Chatfield y Prothero con su serie de ventas, así como los obtenidos por Wilson y Box y Jenkins. Posteriormente se comparan estos resultados con los que resultan al aplicar el método de Guerrero (los cuales serán presentados a lo largo de esta tesis). Es conveniente indicar que esta investigación se llevó a cabo en la computadora HP 3000 de la Universidad Anáhuac, con el paquete estadístico SPSS, que tiene implementada la metodología de Box y Jenkins.

El capítulo 1 de esta tesis proporcionará al lector una breve introducción a la metodología de Box y Jenkins, así como el planteamiento del problema en estudio. El capítulo 2 tratará el uso de las transformaciones en este tipo de análisis. En los capítulos 3 y 4 se presentarán los métodos más conocidos para seleccionar una transformación potencia y para corregir pronósticos por sesgo. Finalmente, en el último capítulo de esta tesis se presentará el caso de Chatfield y Prothero, comparando los resultados obtenidos por otros autores con los resultados obtenidos al aplicar el método de Guerrero. En este mismo capítulo se aplicarán los métodos introducidos en este trabajo a la Serie "Índice de Precios al Consumidor", utilizando para su cálculo una computadora personal y el paquete estadístico STATISTIX.

Agradezco la ayuda del Dr. Miguel Angel Villalobos para

el uso e interpretación de los resultados del paquete SPSS y especialmente agradezco al Dr. Víctor Manuel Guerrero por permitirme presentar los métodos que él desarrolló, así como su valioso tiempo.

CAPITULO 1

Conceptos básicos e introducción al problema

1.1.- Introducción al estudio de las Series de Tiempo.

Actualmente se desarrollan una gran variedad de modelos matemáticos para describir o explicar el comportamiento de fenómenos físicos, químicos, económicos, etcétera; pero desgraciadamente los fenómenos que se presentan en la vida diaria no son completamente deterministas, ya que existen factores aleatorios que no pueden ser medidos o conocidos y que afectan de sobremanera al fenómeno en estudio.

Para estos fenómenos que se ven afectados por factores desconocidos, no es posible elaborar modelos deterministas que proporcionen pronósticos óptimos, por lo cual, en estos casos se recurre al uso de modelos estocásticos.

Una serie de tiempo, la cual se define como una sucesión de valores $\{Z_t\}$ tomados u observados a intervalos con longitud de tiempo fija, puede ser considerada como una realización de un proceso estocástico.

Los modelos estocásticos que actualmente ya son de uso común, suponen que una serie de tiempo, cuyos valores sucesivos pueden ser altamente dependientes, puede considerarse generada a partir de una serie de choques aleatorios independientes $\{a_t\}$. A esta serie de variables aleatorias ordenadas cronológicamente, independientes e idénticamente distribuidas, se le llama "Ruido Blanco". La distribución asociada con cada una de las variables generalmente es una distribución Normal con media cero y varianza

σ^2 .

1.2.- Modelos para representar Series de Tiempo como procesos estocásticos.

Una importante clase de modelos estocásticos que sirven para representar a las Series de Tiempo, son los procesos estacionarios. En estos modelos se supone que el proceso se encuentra en un estado particular de equilibrio estadístico en relación a un nivel constante de la media. Es decir, no existen cambios sistemáticos en la media, o cambios en la varianza y no existen variaciones estrictamente periódicas.

En la práctica generalmente nos topamos con series de tiempo que son no-estacionarias, por lo que su análisis se hace más complicado que si fueran estacionarias.

Existen varios enfoques para el análisis de series de tiempo, como son los expuestos por Winters (1960), Brown (1963), Harrison (1965), etc., pero el que actualmente se ha utilizado con mayor éxito en la práctica, es el de Box y Jenkins (1970). Su análisis consiste en la construcción de modelos del tipo ARIMA (autorregresivos, integrados y de promedios móviles) para representar tanto procesos estacionarios como no-estacionarios, y que por ello representan de manera adecuada a las diferentes series que se presentan en la práctica.

De una manera muy superficial, haré mención a los que forman la familia de los modelos ARIMA.

1) Modelo Autorregresivo: AR(p)

Se expresa como una combinación lineal de los valores pasados del proceso y de un término de error o "choque" aleatorio etc.

Son de la forma:

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

donde $\{Z_t\}$ es la serie observada, μ es el nivel de la serie y $\phi_1, \phi_2, \dots, \phi_p$ son los parámetros autorregresivos.

2) Modelos de Promedios Móviles: MA(q)

Son una combinación lineal del Ruido Blanco $\{a_t\}$, q periodos atrás. Son de la forma:

$$Z_t - \mu = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t$$

con $\theta_1, \theta_2, \dots, \theta_q$ los parámetros de Promedios Móviles.

3) Modelos Autorregresivos de Promedios Móviles. ARMA(p,q)

Son la combinación de los modelos anteriores y toman la forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(Z_t - \mu) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t$$

4) Modelos Autorregresivos Integrados y de Promedios Móviles: ARIMA(p,d,q)

El modelo ARIMA, supone una aplicación del operador $(1-B)Z_t = Z_t - Z_{t-1}$ d veces a la serie original y adquiere la forma:

$$W_t - \phi_1 W_{t-1} - \dots - \phi_p W_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

donde:

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

o de manera más compacta:

$$\Phi(B)W_t = \Theta(B)a_t \quad (1.2.1)$$

con $\Phi(B)$ y $\Theta(B)$ que representan a los polinomios autorregresivos y de promedios móviles, respectivamente.

Para tener en cuenta ambos tipos de efectos, estacionales y no-estacionales, Box y Jenkins (1970) han propuesto un modelo

general del tipo:

$$\Theta(BE) \nabla^d E(Z_t - \mu) = \Theta(BE) \alpha_t \quad (1.2.2)$$

donde las variables $\{\alpha_t\}$ no se suponen Ruido Blanco, sino generadas por un proceso ARIMA (p,d,q) o sea

$$\Phi(B) \nabla^d \alpha_t = \Theta(B) a_t \quad (1.2.3)$$

con $\{a_t\}$ un proceso de Ruido Blanco. De estas dos últimas expresiones se obtiene el Modelo Multiplicativo Estacional

$$\Phi(B) \Theta(BE) \nabla^d E(Z_t - \mu) = \Theta(B) \Theta(BE) a_t$$

denotado por ARIMA (p,d,q) X (P,D,Q)E. La importancia de este modelo radica en que, para series con observaciones mensuales, se tiene en cuenta posibles relaciones: 1) entre meses contiguos dentro de los años, debido a (1.2.3) y 2) entre años para los mismos meses, a causa de (1.2.2).

El método para la construcción de modelos para series de tiempo desarrollado por Box y Jenkins, consta de cuatro etapas:

1) IDENTIFICACION: Aquí se determinan los órdenes de los polinomios autorregresivo y de promedios móviles, así como el número de veces que se debe aplicar el operador diferencia.

2) ESTIMACION: de los parámetros involucrados en el modelo tentativo previamente identificado.

3) VERIFICACION: de los supuestos implícitos en el modelo y de que el modelo proporciona un ajuste adecuado.

4) USO DEL MODELO: para los fines por el cual se haya realizado el estudio. Esta etapa está íntimamente ligada con la realización de pronósticos.

La técnica de Box y Jenkins es un proceso iterativo, es decir, pueden cometerse errores en la identificación de patrones

característicos de cada modelo, lo cual nos llevaría a la obtención de modelos sobre o sub-parametrizados. En ambos casos es necesario regresar a la etapa de identificación y el proceso se repite hasta encontrar un modelo que satisfaga la verificación.

1.3.- Importancia de la validez de los supuestos del modelo.

Supóngase que se ha identificado un modelo para una serie de tiempo en particular y que además se han estimado sus parámetros. La pregunta que ahora el investigador debe responder es si el modelo elegido es adecuado o no. Si el modelo no es adecuado, necesitamos saber en qué y cómo debe ser modificado el modelo tentativo, según el proceso iterativo de Box y Jenkins.

La etapa de verificación que sugieren Box y Jenkins, tiene su origen en la idea de que TODO MODELO ES ERRONEO. Si tenemos dos o más prospectos de modelos, se elegirá aquel que presente menos fallas o el que tenga fallas sin trascendencia. Dichas fallas se miden como violaciones a los supuestos que fundamentan al modelo.

Esta verificación de la "no-violación" a los supuestos de los modelos, se hace a través del análisis de RESIDUALES. Se entiende por RESIDUAL a la parte de las observaciones que no es explicada por el modelo. Para el modelo ARIMA $\hat{\phi}(B)w_t = \theta(B)a_t$, los residuales se definen como:

$$\begin{aligned}\hat{a}_t &= [\hat{\theta}(B)]^{-1} \hat{\phi}(B)w_t \\ &= \hat{\pi}(B)w_t\end{aligned}$$

donde $\hat{\pi}(B)$ es un polinomio de retraso con coeficientes estimados

a partir de las estimaciones de los β 's y θ 's. De esta relación se sigue que:

$$W_t = \hat{a}_t + \hat{\pi}_1 W_{t-1} + \hat{\pi}_2 W_{t-2} + \dots$$

ahora, la estimación de W_t , a partir de información hasta el tiempo $t-1$, viene a ser:

$$\hat{W}_t = \hat{\pi}_1 W_{t-1} + \hat{\pi}_2 W_{t-2} + \dots = \hat{W}_{t-1}(1)$$

por lo que:

$$\hat{a}_t = W_t - \hat{W}_{t-1}(1)$$

Es decir, el residual \hat{a}_t mide la discrepancia entre el valor observado y el valor estimado por el modelo a partir de información hasta el momento $t-1$. Ya que los residuales representan entonces una "estimación" del error, al analizar los residuales (\hat{a}_t) se analiza lo que debería ser una realización del proceso de Ruido Blanco (a_t).

Los supuestos que se deben checar para determinar si un modelo puede o no ser utilizado, son los siguientes: (véase Guerrero, 1983 para mayores detalles).

SUPUESTO 1.- $\{a_t\}$ TIENE MEDIA CERO.

Se verifica usando el cociente:

$$C = \frac{\sqrt{N-d-p} \quad m(\hat{a})}{\hat{\sigma}_a}$$

donde

$$m(\hat{a}) = \frac{\sum_{t=1}^N \hat{a}_t}{N-d-p} \quad \text{y} \quad \hat{\sigma}_a = \sqrt{\frac{\sum_{t=1}^N (\hat{a}_t - m(\hat{a}))^2}{N-d-p-q}}$$

donde $t' = d+p+1$

Si $|C| < 2$ No rechazamos el supuesto.

Si $|C| > 2$ Rechazamos la hipótesis de que $E(a_t) = 0$, debido a que $m(\hat{a})$ es significativamente distinta de cero.

SUPUESTO 2.- $\{a_t\}$ TIENE VARIANZA CONSTANTE.

La verificación se hace con una gráfica de los residuales contra el tiempo, observando si la varianza parece ser o no constante a simple vista.

Si la varianza presenta algún patrón de crecimiento o de decrecimiento, debemos rechazar el supuesto.

SUPUESTO 3.- $\{a_t\}$ SON MUTUAMENTE INDEPENDIENTES.

Calculamos la función de autocorrelación de $\{\hat{a}_t\}$ y esperamos que:

$$r_k(\hat{a}) = \text{Corr}(\hat{a}_t, \hat{a}_{t-k}) = 0 \quad \text{para } k \neq 0$$

Se estima la desviación estándar de $r_k(a)$ como:

$$\sqrt{\text{var}[r_k(\hat{a})]} = 1/\sqrt{N-d-p}$$

donde $r_k(\hat{a})$ está dada por:

$$r_k(\hat{a}) = \frac{\sum_{t=1}^{N-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^N \hat{a}_t^2} \quad \begin{matrix} k=1, 2, \dots \\ t'=d+p+1 \end{matrix}$$

$$\text{Si } |r_k(a)| > \frac{2}{\sqrt{N-d-p}}$$

se dice que la k -ésima autocorrelación es significativamente distinta de cero.

Para retrasos pequeños ($k \leq 3$) esta prueba no es válida y se requiere utilizar el estadístico Q de Box y Pierce (1970) definido por:

$$Q = (N-d-p) \sum_{k=1}^K r_k^2(\hat{a})$$

el cual, para K grande, sigue aproximadamente una distribución χ^2 con $K-p-q$ grados de libertad. La Q debe ser comparada con valores de tablas. Si $Q > \chi_{K-p-q}^2(\alpha)$ rechazamos el que las K primeras autocorrelaciones sean cero al nivel de significación α .

Una mejor aproximación a la distribución χ^2 se obtiene si se recurre al estadístico Q' de Ljung y Box (1978):

$$Q' = (N-d-p)(N-d-p+2) \sum_{k=1}^K r_k^2(\hat{a}) / (N-d-p-k)$$

SUPUESTO 4.- a_t TIENE UNA DISTRIBUCION NORMAL χ_t .

Para verificar esto, se sugiere utilizar la misma gráfica de los residuales contra el tiempo que se utilizó para verificar el supuesto de varianza constante.

También es útil hacer un histograma de los residuales para visualizar la forma de la distribución.

El supuesto de normalidad no es estricto para los residuales $\{\hat{a}_t\}$ sino para los errores $\{a_t\}$, pero si la violación es muy notoria en los residuales debemos rechazar el supuesto.

SUPUESTO 5.- IMPLICITAMENTE SE SUPONE QUE NO EXISTEN OBSERVACIONES ABERRANTES.

La gráfica de los residuales contra el tiempo nos permite observar si algún residual esta fuera del intervalo $(-3\hat{\sigma}_a, 3\hat{\sigma}_a)$. Si esto sucede, se pensará que el residual corresponde a una observación que no fué generada por el proceso y podría considerarse como observación "sospechosa".

Para estas observaciones, debe investigarse la causa, ya que podría deberse a un cambio estructural en el comportamiento del fenómeno, causado quizá, por una "intervención" exógena a la serie en estudio o por un simple error de copiado de los datos.

Los supuestos que a continuación se verán, no se verifican con el análisis de residuales, sin embargo son

supuestos relacionados con el modelo y por ello deben verificarse.

SUPUESTO 6.- EL MODELO CONSIDERADO ES PARSIMONIOSO.

El principio de parsimonia implica que no se puede reducir el número de parámetros involucrados, puesto que todos son necesarios para explicar el comportamiento del fenómeno.

Para la verificación se construyen intervalos de confianza ($\hat{\theta} \pm 2\sqrt{\text{var}(\hat{\theta})}$) para cada parámetro y se observa si el valor cero se encuentra dentro del intervalo, si esto es así se concluye que la estimación no es significativa y el parámetro puede cancelarse del modelo.

SUPUESTO 7.- EL MODELO ES ADMISIBLE.

El supuesto puede verificarse de acuerdo con las condiciones de estacionariedad y/o invertibilidad correspondientes a los polinomios de retraso que intervengan.

Un proceso AR(P) podrá ser estacionario o no-estacionario, dependiendo de los valores que tomen las raíces del polinomio asociado al proceso.

$\phi(B)$ puede escribirse como:

$$\phi(B) = (1 - g_1 B)(1 - g_2 B) \dots (1 - g_p B) \quad (1.3.1)$$

de tal manera que el proceso AR definido por $\phi(B)$ será ESTACIONARIO siempre y cuando :

$$|g_i| < 1 \quad \text{para } i=1,2,\dots,p$$

es decir, si y sólo si, las raíces de $\phi(B)$ se encuentran fuera del círculo unitario (en el plano complejo).

En la práctica, el teorema de Schur (vease Chipman, 1951) puede ser utilizado para encontrar las condiciones de

estacionariedad en términos de los parámetros ϕ_1, \dots, ϕ_p .

Cuando un proceso puede ser expresado mediante un modelo AR, se dirá que dicho proceso es INVERTIBLE lo cual significa que tiene como posible representación a:

$$\pi(B) Z_t = a_t \quad \text{donde } Z_t = z_t - \mu$$

en donde:

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$$

es un polinomio de retraso que cumple con que la suma:

$$\pi(x) = 1 - \sum_{i=1}^{\infty} \pi_i x^i$$

converge DENTRO O SOBRE EL CIRCULO UNITARIO $|x| \leq 1$.

Las condiciones para invertibilidad de un proceso MA pueden obtenerse de manera similar a las condiciones de estacionariedad para procesos AR.

SUPUESTO B.- EL MODELO ES ESTABLE EN LOS PARAMETROS.

Para esto, se calculan las correlaciones entre parejas de parámetros estimados, teniendo cuidado de correlaciones altas (positivas o negativas) entre los parámetros estimados, ya que son indicadores de posibles causas de inestabilidad en la estimación.

Además de verificar los supuestos anteriores, Box y Jenkins sugieren verificar la bondad del modelo por medio de la "sobreparametrización". Esto último consiste en introducir parámetros extra en el modelo, previendo la posible necesidad de incluirlos, pero que, en caso de no ser necesarios se verían rechazados por la verificación del supuesto de parsimonia.

1.4.- Transformaciones y Problema inicial.

En vista de que cualquier modelo utilizado es únicamente una representación simplificada de la realidad y por lo tanto es posible que presente fallas, es de vital importancia que los supuestos anteriormente expuestos sean verificados. Si se cumplen, el modelo puede ser utilizado, si no, se tratará de diagnosticar la causa, repitiendo el ciclo de Box y Jenkins y corregir el supuesto que ha sido violado.

Por ejemplo, si hemos detectado violación al supuesto de que $\{a_t\}$ tiene media cero, podemos corregir el modelo incluyendo un término de tendencia determinista (B_0) o bien chequeando si no es necesario aplicar una diferencia más o un término autorregresivo adicional. (Si se desea profundizar en este punto, remítase al texto de Box y Jenkins, 1970 o al de Guerrero, 1983).

Es necesario mencionar entonces, que en cualquier caso de violación a los supuestos debe hacerse una "DEBIDA CORRECCION". Supóngase que el supuesto 2 o el 4 ha sido violado. En el primer caso, la aplicación de una transformación potencia podría ser útil para estabilizar la varianza. Dicha transformación es del tipo:

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

En el segundo caso, el supuesto de distribución normal para el Ruido Blanco podría validarse por medio de una

transformación normalizante, como podría ser la transformación potencia antes mencionada, pero modificada por Box y Cox (1964) a

$$T(Z_t) = \begin{cases} \frac{(Z_t^\lambda - 1)}{\lambda} & \text{si } \lambda \neq 0 \\ \log(Z_t) & \text{si } \lambda = 0 \end{cases} \quad Z_t > 0$$

En ambos casos parece ser que la corrección a realizar no es difícil, sin embargo, el índice de estas transformaciones, puede seleccionarse de manera "errónea" y conducirnos a un fracaso. En el capítulo siguiente se presentarán algunos métodos para poder elegir λ .

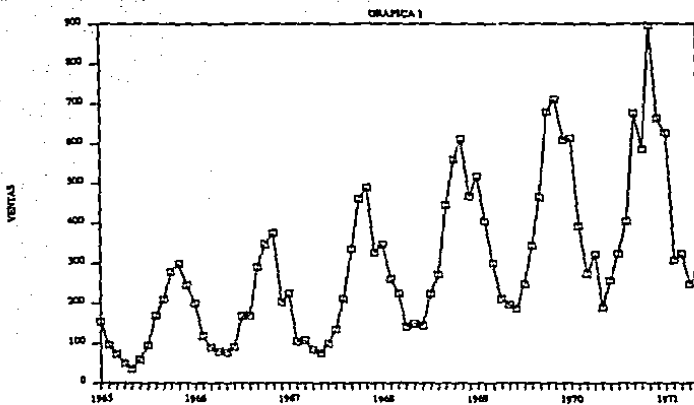
Como ejemplo introductorio de estos "errores", haré mención al artículo de Chatfield y Prothero (1973), en el cual, con el propósito de pronosticar 12 meses adelante, estudian una serie de las ventas de un producto de ingeniería. Los datos proporcionados comprenden 6 años con 5 meses, hasta mayo de 1971; es decir, un total de 77 observaciones.

En el artículo se describe paso a paso el análisis y la discusión de varios problemas prácticos que se encontraron en "el camino", los cuales me voy a permitir reproducir brevemente.

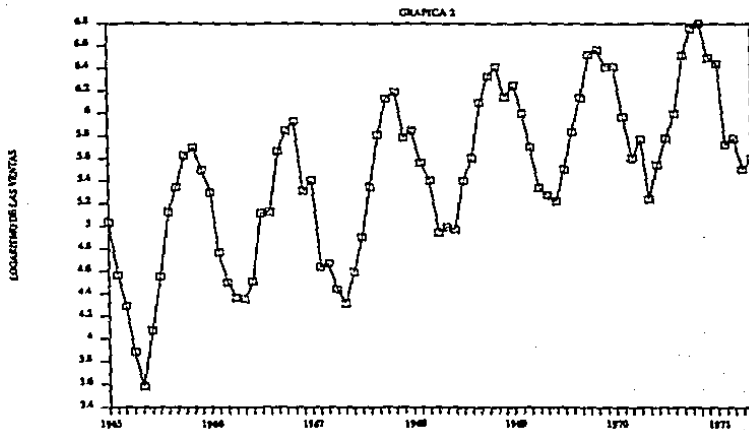
El método que se utilizó para el estudio, es el de Box y Jenkins, para lo cual se utilizó el paquete de computadora ICL disponible en la Universidad de Bath y al cual tuvieron acceso Chatfield y Prothero.

Los datos $\{Z_t\}$ se muestran en la tabla 1, así como su gráfica en la figura 1. Como se puede ver, la serie muestra una tendencia muy marcada y también un marcado patrón estacional. El

VENTAS DE LA COMPA-IA X



LOGARITMOS DE LAS VENTAS



tamaño del efecto estacional se incrementa con el nivel de las ventas, "indicando que el efecto estacional es multiplicativo". Por lo que Chatfield y Prothero aplicaron una transformación logarítmica para obtener un efecto estacional aditivo. (Ver gráfica 2).

TABLA 1
VENTAS DE LA COMPANIA X.

	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
1965	154	96	73	49	36	59	95	169	210	278	298	245
1966	200	118	90	79	78	91	167	169	289	347	375	203
1967	223	104	107	85	75	99	135	211	335	460	488	326
1968	346	261	224	141	148	145	223	272	445	560	612	467
1969	518	404	300	210	196	186	247	343	464	680	711	610
1970	613	392	273	322	189	257	324	404	677	858	895	664
1971	628	308	324	248	272							

Como la serie presenta una tendencia y un patrón estacional de un ciclo de 12 observaciones, examinaron las funciones de autocorrelación de las series $\nabla^d \nabla_{12}^p Z_t$ donde, $Z_t = \log_{10} x_t$ para $d=0.1$ y $D=0.1$ alternativamente y concluyeron (siguiendo el ejemplo de Box y Jenkins de la serie del número de pasajeros total mensual de una línea aérea) que el operador $\nabla \nabla_{12}$ estabilizaría la tendencia y el patrón estacional. Prosiguieron con la identificación del modelo haciendo uso de la FAC y encontraron que un modelo tentativo era:

$$(1 - \phi B) W_t = (1 - \theta B^{12}) a_t$$

donde

$$W_t = \nabla \nabla_{12} Z_t$$

Se estimaron los parámetros del modelo por mínimos cuadrados, con valores iniciales en cero. Alternativamente se calcularon por procedimientos de estimación no-lineal resultando

$\hat{\theta} = -0.47$ y $\hat{\theta} = 0.81$ por lo tanto el modelo resulta ser:

A) $W_t = -0.47 W_{t-1} + a_t - 0.81 a_{t-12}$

Chatfield y Prothero procedieron enonces a verificar los "supuestos del modelo" y realizaron un análisis de residuales. Se muestra la FAC de los residuales en la tabla 2, donde se observa que los retrasos 7 y 11 tienen un valor absoluto mayor que dos veces su desviación estándar ($=1/\sqrt{64} = .125$). Chatfield y Prothero no se sorprenden de encontrar dos valores significativos en 24 coeficientes y proceden a calcular el

estadístico de Box y Pierce $Q = 64 \sum_{k=1}^{36} r_k^2 = 29.07$ el cual comparan con χ_{34}^2 de tablas y encuentran que no hay indicios de que el modelo sea inadecuado al nivel de significación del 5%.

TABLA 2
FUNCION DE AUTOCORRELACION DE
LOS RESIDUALES DEL MODELO A

RETRASOS	AUTOCORRELACIONES					
1-12	0.01	0.10	-0.10	-0.10	-0.16	-0.08
	-0.29	-0.06	-0.04	-0.06	0.34	0.08
13-24	0.02	-0.08	-0.02	0.07	-0.04	-0.10
	0.04	0.05	0.02	0.04	-0.30	-0.20

Posteriormente pronosticaron 12 meses adelante (tabla 3) y notaron que hay un cierto sesgo introducido por la transformación, ellos hacen notar que las propiedades de esta transformación no son discutidas por Box y Jenkins, sin embargo Chatfield y Prothero hacen uso de ella. El porcentaje de sesgo es aproximadamente dado por $\sigma^2_{\epsilon} (\log_{10})^2 / 2$ donde σ^2_{ϵ} es la varianza de $\{Z_{t+1} - Z_t(1)\}$. En este caso, el porcentaje de sesgo varía de 1.4% cuando el horizonte de pronóstico es uno, a 8.6% cuando el

horizonte es 12.

TABLA 3
PRONOSTICOS PARA Z_{t+h} DEL MODELO "A"
(Y LIMITES DE TOLERANCIA)

ADELANTO	1	2	3	4	5	6
PRONOSTICO	305	482	673	990	1297	1387
	±103	±186	±315	±521	±759	±899

En realidad Chatfield y Prothero encuentran el pronóstico muy "pobre", debido a que un pico (nov.1971) es mucho más alto de lo razonablemente esperado. Debido a esto, vuelven a pronosticar usando la técnica de pronóstico hacia atrás para calcular los residuales, con lo cual mejoraron sustancialmente los pronósticos, pero el pico de noviembre de 1971, así como sus límites de tolerancia, seguían siendo considerablemente altos. Por lo que se concluyó que el modelo A no era satisfactorio para propósitos de pronóstico.

Por lo anterior, consideran 3 modelos más:

$$B) (1-\theta B)(1-\theta B^{12})W_t = a_t$$

$$C) (1-\theta B^{12})W_t = (1-\theta B) a_t$$

$$D) W_t = (1-\theta B)(1-\theta B^{12}) a_t$$

Analizaron la FAC de los residuales de cada modelo y encontraron que los 3 modelos presentan altas autocorrelaciones en los retrasos 7, 11 y 24. Las Q calculadas se muestran en la tabla 4. La Q del modelo A es la mínima, por lo que se supone que es el modelo más adecuado. Sin embargo ninguno de los valores de la Q son significativos al nivel del 5%, comparados con la χ^2_{24} , lo cual indica que todos los modelos pueden considerarse adecuados. Se calculó $\hat{\sigma}_a^2$ para cada modelo y se vió nuevamente

que el mínimo se lograba con el modelo A.

TABLA 4

MODELO	VALORES DE Q	VALORES DE $\hat{\sigma}_a^2$
A	29.07	0.00523
B	38.60	0.00629
C	35.87	0.00694
D	33.94	0.00539

Se obtuvieron los pronósticos con cada uno de los modelos y se encontró que los modelos A y D daban pronósticos similares, así como B y C, pero en conjunto los pronósticos son sustancialmente diferentes; sin embargo Chatfield y Prothero seleccionaron el modelo B por dar un mejor ajuste a las observaciones de Z_t .

En la discusión que se presenta al final del artículo, Chatfield y Prothero hacen mención a todas sus "inconveniencias", como son que la diferenciación introduce "autocorrelaciones espurias"; que el tener tantos modelos opcionales no es bueno; sugieren que los pronósticos cumplan con un conjunto de condiciones dadas y no sólo con la optimalidad, etc.

Ellos concluyen (por el resultado de su análisis) que el procedimiento de Box y Jenkins puede ser inadecuado para el pronóstico de ventas cuando hay una componente estacional multiplicativa y sugieren una investigación más profunda, así como la obtención de experiencia práctica con el método. Sin embargo, Chatfield y Prothero aceptan que el estudio no fué satisfactorio por una causa que para ellos no está muy clara, pero al mismo tiempo sospechan que el uso de la transformación logarítmica pudo ser la raíz del problema.

El trabajo de esta tesis esta enfocado a los distintos usos de las transformaciones y a los métodos a los que se recurre para seleccionar λ , de tal forma que se tenga una base más sólida para usar una cierta transformación, como puede ser la logarítmica.

Después de la publicación del artículo de Chatfield y Prothero, Box y Jenkins (1973) estudiaron la serie de ventas aplicando su método y publicaron sus resultados en la misma revista. Estos resultados se mostrarán en los capítulos subsecuentes.

1.5 Revisión de la literatura existente.

Debido a que el objetivo de esta tesis es el de presentar los métodos propuestos por Guerrero (1983) para la selección del parámetro λ en el uso de las transformaciones potencia así como el método para corrección del sesgo introducido por la misma, es necesario revisar los métodos anteriores y de ser posible, los posteriores a los de Guerrero.

Por razones obvias, la mayoría de las metodologías existentes para analizar series de tiempo, se han originado en los países desarrollados, haciendo un tanto difícil su acceso, sin embargo en México existen algunas Instituciones y Universidades que cuentan con revistas científicas en donde se pueden encontrar algunas publicaciones interesantes sobre el tema.

Por lo que respecta al uso de transformaciones, varios son los trabajos escritos sobre el tema, pero quizá uno de los trabajos más importantes que dieron lugar a la consolidación de

algunos métodos fue el de Bartlett (1947), en el cual estudia las transformaciones que se aplicaban a datos estadísticos pero haciendo especial énfasis en análisis de varianza, tratando de determinar empírica o teóricamente la relación entre la varianza y la media de la variable transformada.

En 1964, Box y Cox publican su artículo "An Analysis of Transformations" en el cual, la idea general es la de restringir la atención a las transformaciones indexadas por un parámetro desconocido λ y luego estimar λ y los otros parámetros del modelo simultáneamente por métodos estándares de inferencia. El método desarrollado por Box y Cox será presentado en capítulos subsecuentes.

Posteriormente Jenkins (1979) publica el método de la "Gráfica Media-Recorrido", que se basa en la inspección visual de la gráfica entre la Media y el Recorrido de los datos observados. Este método también será discutido más ampliamente en el capítulo 3 de esta tesis.

Varios son los artículos escritos sobre transformaciones, aparte de Bartlett (1947) y Box y Cox (1964), Tukey (1957) y Box y Tidwell (1962) discuten las transformaciones de la variable independiente en modelos de regresión. Draper y Hunter (1969), Nelson y Granger (1979) estudian las consecuencias de la aplicación de varias transformaciones en un conjunto particular de datos: sin embargo son pocos los nuevos métodos propuestos.

Otro artículo que es interesante en este tema, es el algoritmo numérico desarrollado por Ansley, Spivey y Wroblewski (1977) para estimar el parámetro λ de la transformación de Box y

Cox en un modelo ARIMA, conjuntamente con los otros parámetros del modelo.

Hopwood, McKeown y Newbold (1981) aplicaron la metodología de Box y Jenkins a series trimestrales de las ganancias obtenidas en cierto tipo de inversiones. Estos autores consideraron cuatro estrategias para seleccionar el parámetro λ e hicieron comparaciones de acuerdo a los resultados obtenidos.

Consideraron : a) No usar transformaciones, b) Usar la transformación logarítmica, c) usar $\hat{\lambda}$ estimada por Máxima Verosimilitud y d) Una "Regla de Decisión" basada en los intervalos de confianza al 95% para λ . Si el intervalo contiene $\lambda=1$ pero no $\lambda=0$, no se usará transformación alguna. Si el intervalo contiene $\lambda=0$ pero no $\lambda=1$, se usará la transformación Logarítmica. Si el intervalo no contiene ni $\lambda=0$ ni $\lambda=1$ se usará $\hat{\lambda}$. Si contiene ambos, se usará cualquiera de las dos que esté más cercana a $\hat{\lambda}$. Esta estrategia fué motivada por el trabajo de Box y Cox (1964), pero no se considera un método formal.

Estos autores, de acuerdo a los resultados que obtuvieron, indican que el uso indiscriminado de la transformación logarítmica produce pronósticos que, en general, son marcadamente inferiores a aquellos resultados obtenidos cuando no se usa ninguna transformación, pero el pronóstico mejora cuando se utiliza cualquiera de las λ s obtenidas por Máxima Verosimilitud o por la "Regla de Decisión".

Posteriormente, estos mismos autores (1984) proponen un método heurístico que encontraron de gran utilidad en la selección de modelos ARIMA cuando se considera aplicar una

transformación potencia a la serie en estudio.

Esta aproximación heurística considera una estimación preliminar del parámetro. La estimación preliminar, $\hat{\lambda}^m$, se obtiene estimando autorregresiones por mínimos cuadrados, con varios valores para λ , la k-ésima autoregresión es:

$$x(\lambda)_t = \sum_{j=1}^k B_j x(\lambda)_{t-j} + e_t$$

en donde e_t es el término de error.

El valor inicial de λ es entonces el valor $\hat{\lambda}^m$ que, sobre los valores escogidos, maximiza

$$g(\lambda) = -\frac{n}{2} \text{Log } \hat{\sigma}^2_{e,\lambda} + (\lambda-1) \sum_{i=1}^n \text{Log } x_t$$

en donde n es el tamaño de la serie en estudio y el segundo término es el logaritmo del Jacobiano de la Transformación. (Este método no será discutido en este trabajo por lo que si se desea profundizar, remítase a Hopwood, McKeown, Newbold (1982)).

El artículo de Chatfield y Prothero (1973) despertó interés sobre la correcta selección del parámetro en el uso de las transformaciones, así como la corrección del sesgo introducido por esta. Por lo que respecta a este último, Neyman y Scott publican su artículo "Correction for bias introduced by a transformation of variables" (1960), en donde el propósito principal es el de obtener un estimador insesgado con varianza mínima de los resultados del estudio, expresado en las unidades originales de la serie.

Posteriormente, Granger y Newbold (1976) analizan la estructura de las autocovarianzas de la serie y sugieren métodos

alternativos para pronosticar las series transformadas y una expresión para el sesgo, cuando no se utiliza el pronóstico óptimo. Ellos presentan resultados teóricos, aplicables al pronóstico de la serie transformada, cuando esta es Gaussiana, para una amplia clase de transformaciones. Estos autores atacan el problema usando polinomios hermitianos.

Miller (1984) propone soluciones al sesgo introducido por un conjunto de transformaciones de las más comunmente usadas, considerando la necesidad de obtener un estimador de la "Respuesta Media" en modelos de regresión, en términos de la escala original de la variable, explicando la razón del sesgo y dando remedios apropiados al problema.

Taylor (1986) da un método aproximado para estimar la media de la variable dependiente después de que un modelo lineal ha sido adecuado para la variable transformada, estudiando sus propiedades y mostrando que hay un cierto "costo", debido a la estimación de λ , comparado con la situación que presenta una λ conocida a priori, aunque en general este costo es muy pequeño.

Quizá, el más reciente método es el de Pankratz y Dudley (1987) quienes proponen un algoritmo para encontrar el sesgo introducido por la transformación de Box y Cox, calculando la media con el mínimo error cuadrático para el pronóstico en la variable original. Este método no será discutido en este trabajo.

Nótese que el método de Box y Cox (1964) y el de la gráfica Recorrido-Media (1973) son métodos previos al de Guerrero así como el de Neyman y Scott (1960), Granger y Newbold (1976); pero el de Hopwood, Mckeen y Newbold (1981), Miller (1984),

Taylor (1986) y Pankratz (1987) son posteriores.

En los siguientes capítulos se presentarán los métodos más conocidos y algunos resultados obtenidos al aplicarse a la serie de Chatfield y Prothero.

CAPITULO 2

**Diferentes usos de las transformaciones
en el análisis de Series de Tiempo.**

Una razón básica para transformar los datos de la serie en estudio, es simplemente hacer el análisis lo más simple que sea posible para llegar a obtener un modelo poco complejo; es decir, la serie original puede requerir el uso de un modelo ARMA(1,1), mientras que después de transformar la serie, un modelo AR(1) quizá pueda ser un modelo adecuado.

Otra razón lógica para el uso de transformaciones en el estudio de series de tiempo, es la de encontrar la "escala de medición ideal" que satisfaga los supuestos teóricos requeridos para el análisis.

Como se mencionó anteriormente, en la práctica lo más común es que las series de tiempo que se analizan sean no-estacionarias, ya sea porque exhiben algún tipo de tendencia, porque su varianza no sea constante o porque estén influenciadas por algún factor de tipo semideterminista. Se recurre al uso de transformaciones básicamente :

a) Para estabilizar la varianza.- Si hay una tendencia en la serie y la varianza parece incrementarse con la media, entonces puede ser aconsejable usar una transformación.

b) Para normalizar la distribución de los errores.- Si existe algún patrón que indique la presencia de no-normalidad en la distribución empírica, una transformación normalizante puede ser el remedio.

c) Para hacer que el efecto estacional sea aditivo.- Si hay una tendencia en la serie y el tamaño del efecto estacional parece incrementarse con la media, entonces puede ser provechoso el uso de transformaciones para hacer que el efecto estacional

sea constante.

De acuerdo con Draper y Hunter (1969), Box y Cox (1964), Taylor (1986), Makridakis (1976) se ha visto que la transformación Potencia da mejores resultados que otras transformaciones por lo que actualmente se ha hecho de uso común.

2.1.- Transformaciones para estabilizar la varianza.

Si la no-estacionariedad se debe a que tanto la varianza como el nivel de la serie no son constantes, quizá la causa sea que en cada punto de observación t , la variable Z_t tiene varianza σ^2_t la cual es función de su media μ_t ; de ocurrir esto, un argumento derivado del trabajo de Bartlett(1947), conduce a determinar una transformación potencia que puede ser útil para estabilizar la varianza de la serie.(como lo indica Guerrero 1983, Cap. 3).

El propósito de Bartlett al escribir su artículo, fué el hacer un resumen de las transformaciones que se aplicaban en el análisis de datos estadísticos con particular referencia al análisis de varianza, afirmando que " Si la variación tiende a cambiar con el nivel medio de las mediciones, la varianza solamente será estabilizada con un cambio apropiado de escala."

Supongamos que la varianza σ^2_t de la variable aleatoria Z_t puede expresarse como función de su media μ_t , o sea:

$$\sigma^2_t = f(\mu_t)$$

entonces, si $T(x)$ es una función cuya primera derivada existe, puede obtenerse la siguiente aproximación lineal a $T(Z_t)$:

$$T(Z_t) \approx T(\mu_t) + \left(\frac{dT}{dZ_t} \Big|_{Z_t=\mu_t} \right) (Z_t - \mu_t)$$

de donde se obtiene que una aproximación lineal a la varianza de $T(Z)$ es:

$$\text{var}(T) \doteq \left(\frac{dT}{dZ_t} \Big|_{Z_t=\mu_t} \right)^2 f(\mu_t)$$

así, ya que se desea que $T(Z_t)$ tenga varianza constante C^2 , se sigue que:

$$\frac{dT}{dZ_t} \Big|_{Z_t=\mu_t} \doteq \frac{C}{\sqrt{f(\mu_t)}}$$

y por lo tanto:

$$T(\mu_t) \doteq \int \frac{C}{\sqrt{f(\mu_t)}} d\mu_t \quad (2.1.1)$$

Como se observa, es necesario conocer f para poder estar en condiciones de utilizar (2.1.1) y así poder determinar la transformación T que estabilice la varianza. En esta situación, Guerrero (1983) restringe el tipo de transformación a la FAMILIA DE TRANSFORMACIONES POTENCIA, para que sea factible emplear (2.1.1): es decir, si la variable Z_t es positiva y si es razonable suponer que $\sigma^2_{Z_t}$ es proporcional a $\mu^{2(1-\lambda)}$ para algún valor de λ ; es decir, si:

$$f(\mu_t) \propto \mu_t^{2(1-\lambda)}$$

entonces se sigue que:

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

Este último resultado sugiere que la función T , que vuelve aproximadamente constante a la varianza de Z_t , debe ser la

transformación potencia:

$$\pi(Z_t) = \begin{cases} Z_t^\lambda & \text{si } \lambda \neq 0 \\ \log(Z_t) & \text{si } \lambda = 0 \end{cases} \quad (2.1.2)$$

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

2.2.- Transformaciones para normalizar la distribución de los errores.

Como se mencionó en el capítulo anterior, el supuesto de normalidad se debe cumplir para los errores aleatorios $\{a_t\}$, pero no tiene por qué ser satisfecho exactamente por los residuales $\{\hat{a}_t\}$.

Dicho supuesto podría validarse en cierta medida mediante la aplicación de una transformación "Normalizante", como podría ser la transformación potencia (2.1.2), que fué modificada por Box y Cox (1964), según se indicó ya en el capítulo 1.

El procedimiento que ellos sugieren, es el resultado de una síntesis de los procedimientos propuestos principalmente por Bartlett (1947), por Tukey (1949, 1950), Tukey y Moore (1954) y por Kleezkonski (1949); Todos ellos con la idea general de restringir la atención a transformaciones indexadas por un parámetro λ desconocido, y entonces estimar dicho parámetro junto con los otros parámetros del modelo, por medio de métodos estándares de inferencia. Esta idea permite de hecho, validar el modelo completo a través de una selección apropiada de λ .

Box y Cox distinguen dos tipos de análisis: a) en donde la transformación particular sea el punto de interés y b) en

donde el interés principal sean los efectos que esta produce. La finalidad es escoger λ , en parte a través de la información que dá la serie y y en parte, escoger λ para que el modelo resulte simple y de fácil entendimiento. Por ejemplo, dentro del contexto de Modelos de Análisis de Varianza, ellos consideran que quizá para un análisis formal, el uso de $\sqrt{Z_t}$, podría considerarse como la mejor "escala" para normalizar, pero al mismo tiempo, se dan cuenta que para los "casos de interpretación" es mejor trabajar con $\log(Z_t)$.

Ellos trabajaron con una familia de transformaciones que convierte a Z_t en $Z_t(\lambda)$, en donde el parámetro λ (o el vector λ_1 y λ_2) define una transformación en particular, entonces se considera a:

$$Z_t^{(\lambda)} = \begin{cases} \frac{(Z_t^\lambda - 1)}{\lambda} & \text{si } \lambda \neq 0, Z_t > 0 \\ \log(Z_t) & \text{si } \lambda = 0, Z_t > 0 \end{cases} \quad (2.2.1)$$

o bien a:

$$Z_t^{(\lambda)} = \begin{cases} \frac{(Z_t + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \text{si } \lambda_1 \neq 0 \\ \log(Z_t + \lambda_2) & \text{si } \lambda_1 = 0 \end{cases} \quad (2.2.2)$$

La transformación (2.2.1) es una modificación de (2.1.2) que sirve para evitar la discontinuidad en $\lambda=0$, mientras que (2.2.2) es sólo una generalización de (2.2.1) que se utiliza en caso de que $Z_t < 0$, pero $Z_t + \lambda_2 > 0$.

A pesar del frecuente uso de (2.2.1), no debe esperarse obtener estricta normalidad en los residuales, pues el que Z_t sea

positiva, implica que $Z_t > -1/\lambda$ según $\lambda > 0$, es decir, Z_t está acotada y por lo tanto no puede tener una distribución Normal de manera estricta.

Por otro lado cabe subrayar que el fin de aplicar la transformación de (2.2.1), es el de validar todo el modelo, es decir el parámetro λ se elige con el criterio de que el modelo ARIMA expresado en (1.2.1) y el supuesto de normalidad de los errores (con varianza constante) sean válidos.

2.3.- Transformaciones para seleccionar modelos en desestacionalización de series.

El análisis de la desestacionalización o el ajuste estacional de series de tiempo, se basa en la idea de que la serie observada está constituida por varias componentes que pueden ser separadas o bien, que las series de tiempo pueden visualizarse como series constituidas por varias componentes no observables (Nerlove, Grether y Carvalho, 1979) como son: la tendencia, el ciclo, la estacionalidad y el movimiento irregular.

Posiblemente la razón más importante para llevar a cabo el ajuste estacional de una serie, sea la que propuso Pearson en 1919 de que cada una de las componentes de una serie es causada por fenómenos distintos. Granger (1978) en lo que toca a la componente estacional, encuentra que existen al menos cuatro posibles causas de las fluctuaciones estacionales, las cuales no tienen por qué ser completamente ajenas o distintas entre sí y son: 1) el calendario mismo (festividades y/o número de días en cada mes) 2) fechas fijadas por instituciones para realizar

ciertas actividades (vacaciones escolares) 3) el clima o dicho de otra manera las estaciones del año y 4) las expectativas de fluctuaciones estacionales (como la e evada producción de juguetes en los meses previos a la Navidad, que es causada por la expectativa de ventas elevadas en diciembre).

Como Kallek (1978) dice, con un ajuste estacional "uno pretende eliminar lo más de la fluctuación que oscurece a la componente de tendencia-ciclo de la serie".

Los modelos básicos que muestran explícitamente la relación que guardan las componentes de una serie observada $\{O_1, O_2, \dots, O_N\}$ son: el multiplicativo

$$O_t = T_t * E_t * I_t \quad (t=1,2,\dots,N)$$

o el aditivo

$$O_t = T_t + E_t + I_t \quad (t=1,2,\dots,N)$$

en donde T_t , E_t e I_t son las componentes de tendencia-ciclo, estacionalidad e irregularidad respectivamente.

A partir de estos modelos básicos, el problema de la desestacionalización se convierte en el de estimación de las componentes T_t , E_t e I_t para cada uno de los periodos de observación $t=1,2,\dots,N$.

Existen varios métodos desarrollados para llevar acabo la desestacionalización de series y que se han agrupado como : 1) Métodos de Regresión , 2) Metodos de Promedios Móviles y 3) Métodos Basados en Modelos; dichos métodos podrían aplicarse a una serie de tiempo dada y el problema ahora sería el decidir cual método es el más apropiado para la serie en estudio.

Una vez que se haya elegido el método de ajuste que habrá

de utilizarse, el siguiente paso deberá consistir en discriminar entre el modelo aditivo y el multiplicativo. Para eso, Durbin y Murphy (1978) y posteriormente Durbin y Kenny (1978), sugieren un procedimiento con base en métodos de regresión para tener en cuenta una componente estacional cuyo comportamiento se refleje en que el modelo sea mixto (en el sentido de que admita la representación aditiva y multiplicativa, simultáneamente). De acuerdo con estos autores, conviene calcular la desviación media absoluta respecto a la tendencia para cada año y graficarla contra la tendencia media anual, de esta manera, la inspección visual de estas gráficas proporciona la indicación de que modelo deberá emplearse.

Para evitar la subjetividad en la apreciación de gráficas, sería preferible utilizar un procedimiento numérico que discrimine entre los modelos multiplicativos y aditivos. El procedimiento numérico que sugiere Guerrero (DI 1983) surge del siguiente argumento: "ya que el modelo multiplicativo $O_t = T_t \times E_t \times I_t$ con $t=1,2,\dots,N$, puede expresarse de manera equivalente como el modelo aditivo $O'_t = T'_t + E'_t + I'_t$ con $t=1,2,\dots,N$, donde (') denota al logaritmo de la componente en el modelo anterior, entonces el modelo a utilizar sería el aditivo y el problema se convierte en realidad en discriminar entre utilizar el logaritmo de las observaciones o emplear las observaciones originales."

De acuerdo con lo expuesto por Durbin y Kenny, si la amplitud de la componente estacional es proporcional (constante respecto) a la tendencia media anual, el modelo será

multiplicativo (aditivo). Así pues, si se supone que la variación que se observe en grupos anuales de datos es causada esencialmente por las fluctuaciones estacionales, el que se obtenga una relación aproximada del tipo: 1) $S_i = KM_i$ o 2) $S_i = K$ con $i=1,2,\dots,n$ donde S_i y M_i representan respectivamente la desviación estándar y la media para las observaciones del año i y $K \neq 0$ es una constante de proporcionalidad, implicaría entonces emplear un modelo multiplicativo en el caso 1) y aditivo en el caso 2).

Guerrero hace notar "esto, desde luego, es una aproximación, al igual que lo es la sugerencia de Durbin y Kenny; no obstante, debe notarse que ahora el supuesto en que se basa la aproximación, entre otras cosas, es de que la tendencia crece de año a año, pero es esencialmente constante dentro del año", por otro lado; el procedimiento propuesto por Guerrero puede justificarse también con el siguiente argumento: Una relación más general que contiene a las expresiones (1) y (2) anteriores viene a ser 3) $S_i/M_i^\delta = K$ donde, si $\delta=1$ se obtiene (1) y si $\delta=0$ se obtiene (2).

Puede demostrarse que esta expresión surge de una adaptación de los resultados de Bartlett (1947), para ello puede consultarse a Guerrero (1983 cap. 3 y 4) en donde se detalla la operacionalización de la expresión (3). Ahora bien, la relación (3) surge en la búsqueda de una transformación estabilizadora de varianza para los datos, ya que si dicha relación se satisface razonablemente para alguna δ , será adecuado aplicar la transformación potencia

$$O'_t = \begin{cases} O^{1-\delta} & \text{si } \delta \neq 1 \\ \log(O_t) & \text{si } \delta = 1 \end{cases}$$

En este caso, el interés radica en utilizar el valor $\delta=0$ o $\delta=1$, dependiendo de si la δ estimada está más cerca de 0 o de 1.

Nótese que los modelos por sí mismos denotan ciertas restricciones para su uso, en especial un modelo multiplicativo no puede emplearse con una serie que contenga ceros o datos negativos, de igual manera como no puede aplicarse el logaritmo a cifras negativas. Así pues, con este tipo de series que contiene datos negativos o ceros, el modelo aplicable es el aditivo. Nótese también que se requiere que los datos originales sean todos positivos para aplicar la transformación potencia.

Otro procedimiento que utiliza transformaciones y que es usado actualmente, es el desarrollado en los laboratorios Bell de los EUA el cual hace uso de gráficas (Cleveland y Terpenning, 1982) y genera modelos para la serie mediante transformaciones, incluyendo como casos particulares a los modelos aditivos y multiplicativos, o sea, el modelo supuesto por SABL es

$$O^{(p)}_t = T_t + E_t + I_t \quad t=1,2,\dots,N$$

con

$$O^{(p)}_t = \begin{cases} O^p_t & \text{si } p > 0 \\ \log(O_t) & \text{si } p = 0 \text{ para } O_t > 0 \\ -O^p_t & \text{si } p < 0 \end{cases} \quad t=1,2,\dots,N$$

en donde la potencia p se elige de tal manera que la descomposición sea precisamente aditiva.

CAPITULO 3

**Presentación y comparación de los diferentes métodos
para elegir una transformación potencia.**

Resumiendo un poco lo expuesto en los capítulos anteriores, podemos decir que en la práctica en ocasiones se presentan series con una variabilidad no constante y no-normalidad en los datos recopilados, violando los supuestos de varianza constante y normalidad de $\{a_t\}$, los cuales deben ser verificados para seguir adelante con el uso de un modelo ARIMA en lo que toca a la inferencia estadística .

Estudiando los cambios de varianza en las series de tiempo, Box y Jenkins (1970) sugieren el uso de la transformación potencia (2.2.1). El uso de esta transformación en estos últimos años, ha sido incorporada en la metodología de construcción de modelos de series de tiempo y el parámetro λ es considerado como otro parámetro a estimar, junto con los demás parámetros del modelo seleccionado.

Aunque la posibilidad de utilizar la transformación potencia fue brevemente expuesta por Box y Jenkins (1970), su uso ha sido de gran interés en la construcción de modelos ARIMA. Este interés fue estimulado por los resultados de Chatfield y Prothero (1973) quienes obtuvieron pronósticos no satisfactorios de la serie de ventas, de lo cual surgió una discusión de si el uso de la transformación logarítmica era apropiado o no.

Podría decirse que la forma apropiada del modelo ARIMA depende en gran parte, de la correcta selección del parámetro λ . Sin embargo, no son muchos los métodos desarrollados para llevar a cabo una correcta selección de λ , en realidad, podemos hablar del método de Box y Cox (Máxima Verosimilitud), el cual es catalogado como "muy sofisticado" y el de la gráfica

recorrido-media, el cual se clasifica como "muy empírico" (Chatfield y Prothero, 1973).

En este capítulo, expondremos estos dos métodos y se introducirá también el método desarrollado por Guerrero. Dichos métodos serán aplicados al problema previamente expuesto en la sección 1.4 mostrando los resultados obtenidos.

3.1.- Método de Box y Cox.

Box y Cox (1964) al exponer (2.2.1) indican que el índice de esta transformación λ , puede estimarse de manera simultánea con los demás parámetros: las ϕ 's, θ_0 y θ 's, mediante Máxima Verosimilitud.

Suponemos que para algún valor del parámetro λ , los valores transformados Z_t dados por (2.2.1), siguen un proceso ARIMA (p,d,q). Los errores aleatorios a_t se suponen con una distribución Normal con media cero y varianza σ^2 .

Debe hacerse notar que en (2.2.1) es necesario excluir valores negativos por lo que, estrictamente hablando, $\{a_t\}$ puede sólo tener una distribución Normal aproximada.

De aquí se obtiene la función de densidad conjunta de los errores aleatorios:

$$(3.1.1)$$

$$f(a_{d+p+1}, a_{d+p+2}, \dots, a_N) = (2\pi)^{-N} \sigma^{-N+d+p} \exp\left\{-\sum_{t=d+p+1}^N \frac{a_t^2}{2\sigma^2}\right\}$$

Ya que $\{Z_t^{(\lambda)}\}$ sigue un proceso ARIMA (p,d,q), el operador diferencia aplicado a esta serie:

$$u_t = \nabla^d Z_t^{(\lambda)}$$

conduce a un proceso ARMA (p,q) para $\{w_t\}$, la expresión (1.2.1) implica que:

$$a_t = w_t - \phi_1 w_{t-1} - \dots - \phi_p w_{t-p} - \theta_0 + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

esto nos permite obtener la función de densidad conjunta de $\{w_t\}$ a partir de (3.1.1) como:

$$f(\{w_t\} | \underline{\phi}, \theta_0, \underline{\theta}, \lambda, \sigma_a^2) = (2\pi)^{-N} \sigma_a^{-N+d+p} \exp(-S/2\sigma_a^2) \quad (3.1.2)$$

donde:

$$S = S(\hat{\underline{\phi}}, \hat{\theta}_0, \hat{\underline{\theta}}, \hat{\lambda}) = \sum_{t=d+p+1}^N (w_t - \phi_1 w_{t-1} - \dots - \phi_p w_{t-p} - \theta_0 + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q})^2$$

Las transformaciones deben ser consideradas en la derivación: la transformación del operador diferencia, cuyo Jacobiano es uno, y la transformación de Box y Cox, cuyo Jacobiano es:

$$J = J_{N-d-p}(\lambda, Z_t) = \prod_{t=1}^{N-d-p} \left| \frac{\partial Z_t^{(\lambda)}}{\partial Z_t} \right| = \prod_{t=1}^{N-d-p} \{Z_t\}^{\lambda-1}$$

La función de densidad (3.1.2) permite calcular probabilidades de la distribución normal multivariada, una vez que se conoce a los parámetros $\underline{\phi} = (\phi_1, \dots, \phi_p)$, $\theta_0, \underline{\theta} = (\theta_1, \dots, \theta_q)$, λ y σ_a^2 . Ahora, lo que se conoce es $\{w_t\}$ y se desconoce $\underline{\phi}, \theta_0, \underline{\theta}, \lambda$ y σ_a^2 por lo que (3.1.2) se considera una FUNCIÓN DE VEROSIMILITUD de los parámetros, la cual depende del vector de observaciones $\{w_t\}$, esto es:

(3.1.3)

$$L(\phi, \theta_0, \theta, \lambda, \sigma_a^2 | \{u_t\}) = \frac{1}{(2\pi)^{-(N-d+p)}} \frac{1}{\sigma_a^{N-d+p}} \exp(-S/2\sigma_a^2) \quad (3.1.3)$$

Dicha función de verosimilitud (3.1.3) debe maximizarse con respecto a los parámetros, para obtener el modelo (1.2.1) más apropiado a la serie en estudio.

Para maximizar (3.1.3), primero se deben elegir los valores ϕ, θ_0, θ y λ que minimicen S , y posteriormente se determina el estimador de σ_a^2 .

Supóngase que se ha encontrado el mínimo de S , entonces se procede a maximizar la función de verosimilitud con respecto a σ_a^2 o equivalentemente a maximizar al logaritmo de la función de verosimilitud, que viene dado por:

$$L^* = -(N-d-p)/2 [\log(2\pi) - \log(\sigma_a^2)] - S/2\sigma_a^2 + \log J$$

Para su maximización considérese la ecuación:

$$\frac{\partial L^*}{\partial \sigma_a^2} \Big|_{\sigma_a^2 = \hat{\sigma}_a^2} = \frac{-N+d+p}{2\sigma_a^2} + \frac{S(\hat{\phi}, \hat{\theta}_0, \hat{\theta}, \hat{\lambda})}{2\sigma_a^4} = 0$$

que produce:

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}, \hat{\theta}_0, \hat{\theta}, \hat{\lambda})}{N-d-p}$$

El problema de maximizar (3.1.3) se reduce a minimizar S . Al tratar de minimizar S , se presentan serios problemas de cálculo, por lo que en la práctica se utiliza algún procedimiento numérico para encontrar el mínimo.

Dado un valor de λ , Box y Jenkins (1970) sugieren un método de estimación no lineal basado en el algoritmo de Marquardt (1963), el cual tiene como fundamento un desarrollo en serie de Taylor que linealiza a a_t condicionada en que se conocen los valores muestrales $\{W_t\}$ y unos valores iniciales de los parámetros del modelo ARIMA a calcular. Los valores iniciales serán corregidos iterativamente con el objetivo final de minimizar S y el proceso iterativo termina cuando se logra "convergencia".

Una manera de estimar el parámetro λ , es por medio de un algoritmo numérico desarrollado por Ansley, Spivey y Wroblewski (1977), cuya aproximación surge al reescribir (3.1.3) como una función monótona de la suma de cuadrados, lo cual permite el uso de un algoritmo de mínimos cuadrados no-lineales para dar una solución numérica.

En el caso de la serie de ventas, se puede utilizar la transformación de Box y Cox (1964). De acuerdo con el método de Máxima Verosimilitud, Wilson (1973) encontró, como se muestra en la gráfica 3, que el parámetro λ para el modelo A, es $\lambda=0.34$. Los intervalos de confianza de λ al 95% de confianza son (0.1,0.6), por lo que Wilson rechaza rotundamente la transformación logarítmica para este caso. Así pues, Wilson fue uno de los estudiosos que analizó la serie de ventas de Chatfield y Prothero, concluyendo que la serie había sido erróneamente transformada.

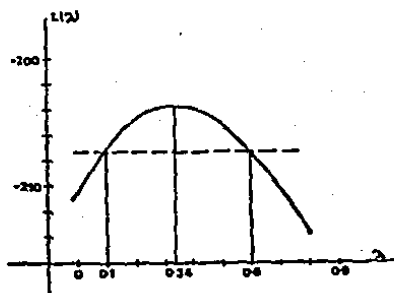
El modelo obtenido por Wilson es:

$$(1+0.378) \nabla \nabla_{12} Z_t^{.34} = (1-0.798^{12}) a_t$$

Los pronósticos del modelo A, se muestran en la gráfica 4.

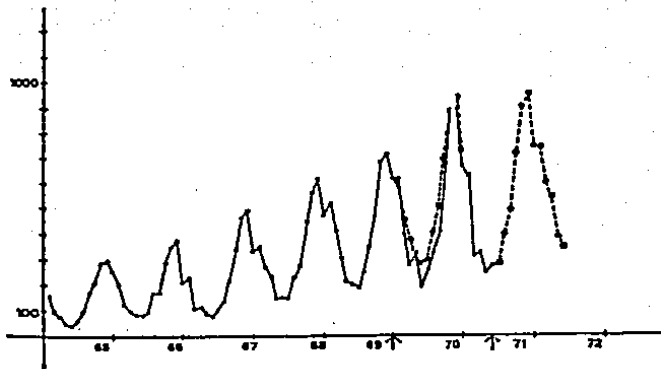
Los pronósticos 12 meses adelante hechos desde noviembre de 1964, siguen bastante bien el patrón de la serie. Los pronósticos hechos desde mayo 1971 hacia adelante, no muestran el comportamiento indeseado que obtuvieron Chatfield y Prothero del modelo A usando la transformación logarítmica.

GRAFICA 3



Función de Máxima Verosimilitud para el parámetro λ .

GRAFICA 4



Pronósticos usando el modelo A con $\lambda = 0.34$

3.2.- Método de la gráfica recorrido-media.

El método de la gráfica recorrido-media, se considera como un método "empírico" para obtener una estimación preliminar de un parámetro λ aproximado para la transformación que se le aplicará a la serie en estudio (Jenkins 1979).

Primero, se divide a la serie en subconjuntos (es útil pensar en subconjuntos de 4 a 12 observaciones y es ideal que estén relacionados con el tamaño del período estacional que presenta la serie en estudio).

Después se calcula el "recorrido" (es decir, la diferencia entre los valores extremos), y la media de cada subconjunto, de tal manera que se obtienen parejas ordenadas para ser graficadas.

A la gráfica del recorrido contra la media, de cada uno de los subconjuntos, se le llama: "La Gráfica Recorrido-Media".

El "recorrido" de cada subconjunto se usa preferentemente en lugar de la desviación estándar, como un estimador de la variabilidad local, debido a su simplicidad; la media se usa como una estimación del nivel local de la serie.

Si el recorrido es independiente de la media, no es necesario hacer uso de una transformación ($\lambda=1$) y la relación entre el recorrido y la media se mostrará con puntos aleatoriamente dispersos alrededor de una línea horizontal. En la gráfica 5 se muestra con líneas la dirección que pueden tomar las parejas ordenadas al ser graficadas.

La transformación logarítmica puede ser útil cuando es sugerida por la gráfica recorrido-media al mostrar una dispersión aleatoria alrededor de una línea inclinada ($\lambda=0$ en la gráfica 5).

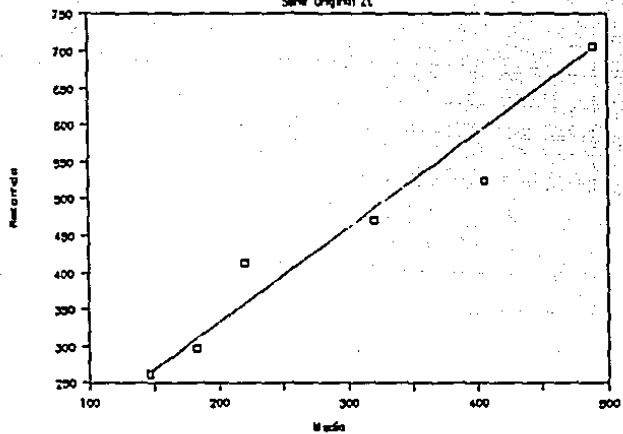
El objetivo del método, no es fijar el valor exacto para λ , sino distinguir entre valores específicos que representen algún interés, por ejemplo $\lambda=1, 0.5$ o 0 , como se muestra en la gráfica 5.

La gráfica 6 muestra (siguiendo la aplicación de este método a la serie de ventas de Chatfield y Prothero) las gráficas del recorrido contra la media cuando los datos han sido divididos en grupos de 12 observaciones de acuerdo con el tamaño del período estacional.

Las gráficas se hicieron para $Z_t, Z_t^{1/2}, Z_t^{1/4}$ y $\log Z_t$. En ambas gráficas el método de la gráfica recorrido-media, sugiere

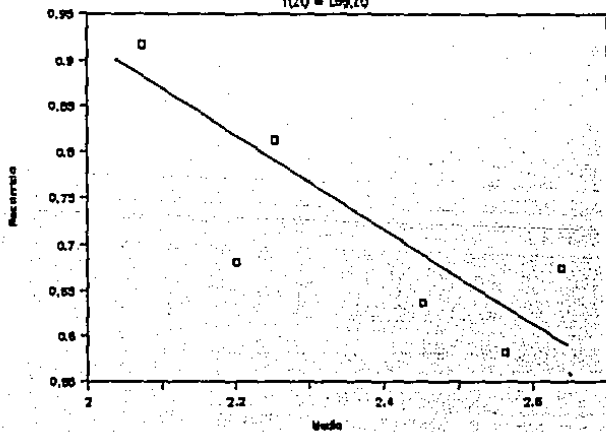
GRAFICA 6 a
Recorrido-Medio

Serie Original II

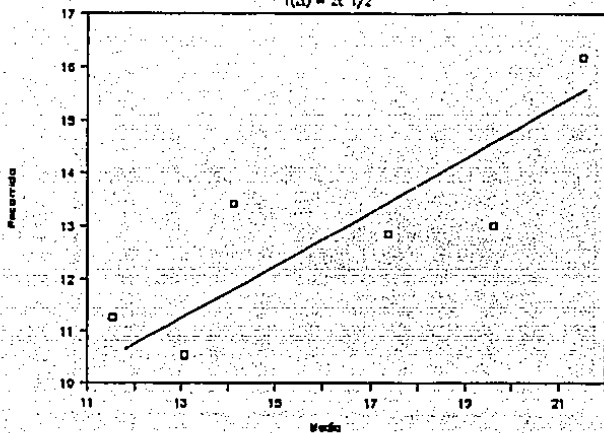


GRAFICA 6 b
Recorrido-Medio

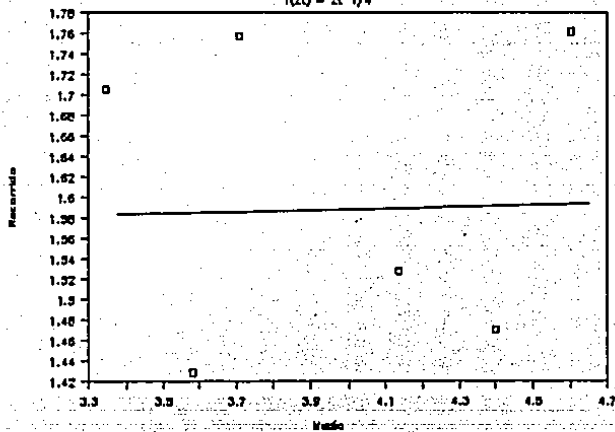
$T(Z) = \text{Log}(Z)$



GRAFICA 6 c
 Recorrido—Media
 $f(x) = 2x^{-1/2}$

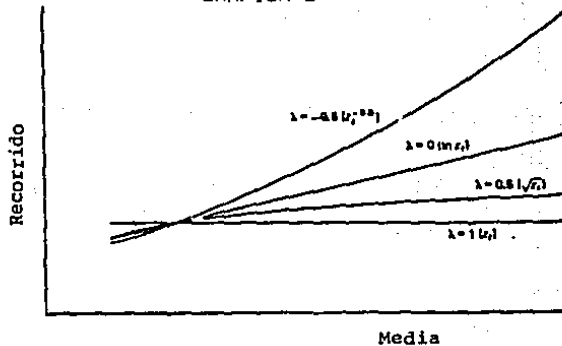


GRAFICA 6 d
 Recorrido—Media
 $f(x) = 2x^{-1/4}$



la transformación $T(Z_t) = Z_t^{0.25}$

GRAFICA 5



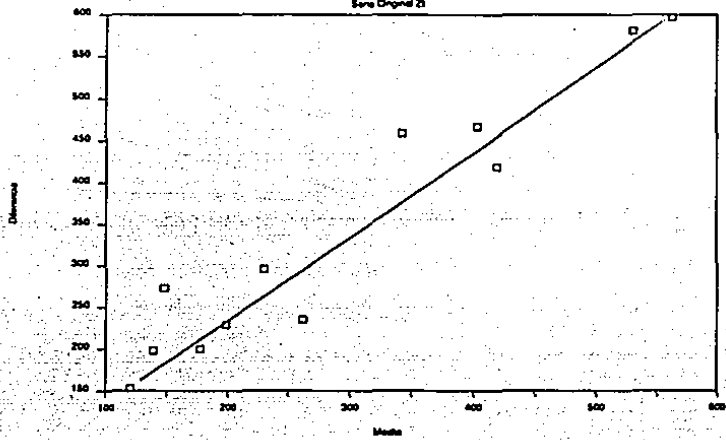
Las curvas muestran la relación teórica entre el recorrido y la media para diferentes tipos de transformaciones. (En realidad, la relación, que se basa en los datos, se mostrará como puntos dispersos aleatoriamente siguiendo el patrón de la relación teórica).

Curiosamente Box y Jenkins llegaron a este mismo resultado en el artículo que escribieron sobre la serie de Chatfield y Prothero (1973), pero usando un método más sencillo. Ellos muestran la tendencia de cada mes, que ha sido estimada como la diferencia entre los valores de 1970 (que es el último año que contienen datos completos) con los valores de 1965.

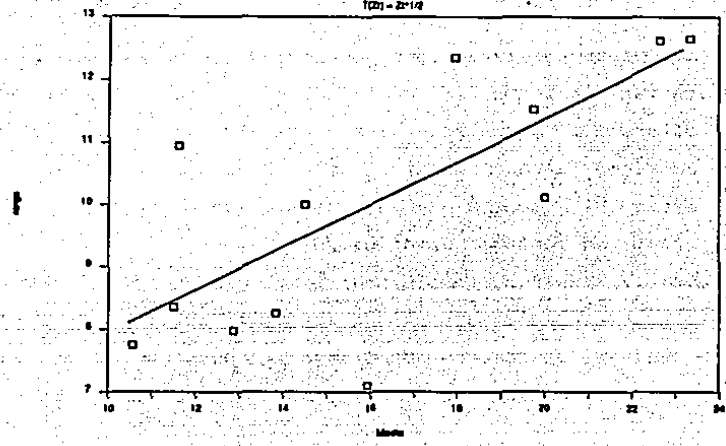
Estas diferencias se han graficado contra la media mensual del mismo periodo. (gráfica 7).

Box y Jenkins concluyeron que era evidente que el logaritmo sobre-transformaba y la raíz cuadrada sub-transformaba, mientras que la raíz cuarta de Z_t era la mejor transformación. Ahora, la transformación $T(Z) = Z^{0.25}$, sugerida por las figuras 6

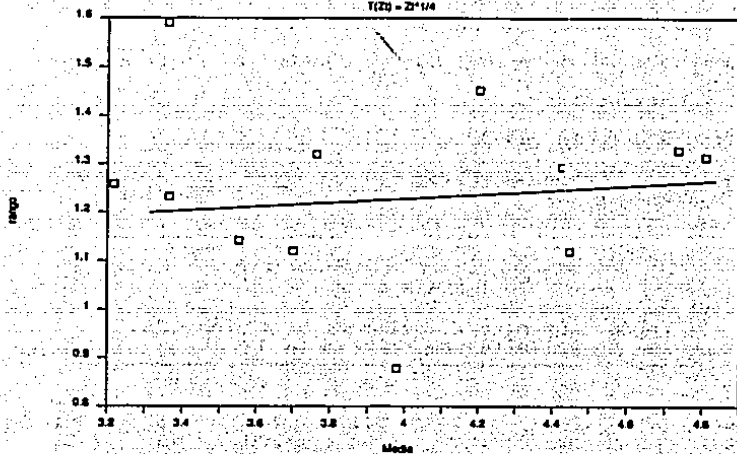
GRAFICA 7 a
Sens Original 21



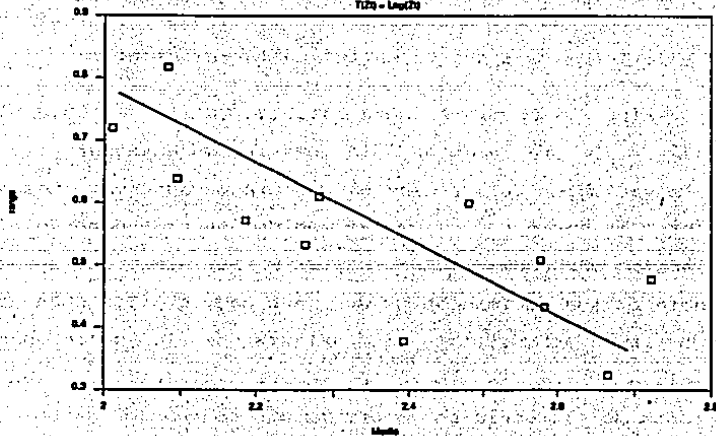
GRAFICA 7 b
 $r(2) = 21/2$



GRAFICA 7 c
 $T(Z) = Z^{1/4}$



GRAFICA 7 d
 $T(Z) = \text{Log}(Z)$



y 7 dan esencialmente resultados similares a $Z^{0.34}t$, que es la transformación obtenida por máxima verosimilitud.

Debido a lo anterior, Box y Jenkins se inclinan por el modelo A, y obtienen:

$$(1 + .58)W_t = (1 - .88I^2) a_t$$

$$\text{con: } W_t = \nabla \nabla_{12} T(Z_t) = \nabla \nabla_{12} Z_t^{0.25} \quad \text{y } \sigma_a^2 = .02336$$

La tabla 5, muestra los resultados obtenidos con el modelo A y $\lambda = 0.25$

TABLA 5
Pronósticos para Z_t del modelo "A" ($\lambda = 0.25$)
(y límites de tolerancia)

Adelanto	1	2	3	4	5	6
Pronóstico	286	409	511	761	966	1091
95%	±83	±128	±184	±280	±373	±420
50%	±29	±44	±63	±96	±128	±145

3.3.- Método de Guerrero.

El método que propone Guerrero (1983), sugiere elegir la potencia λ de tal manera que se satisfaga la relación:

$$\sigma_t / \mu_t^{1-\lambda} = \text{constante para } t=1, 2, \dots, N \quad (3.3.1)$$

en donde σ_t y μ_t representan a la desviación estándar y a la media de la variable Z_t , y N es el número de observaciones que se tiene para la serie $\{Z_t\}$. Como es sabido, en cada momento t se hace solamente una observación de Z_t y por lo tanto no es posible obtener una estimación de σ_t para cada t. Sin embargo, el método para estabilizar la varianza de la serie puede operacionalizarse de la siguiente manera: divídanse a las N observaciones de la serie en H grupos que contengan $R = (N-n)/H$ OBSERVACIONES CONTIGUAS cada uno, dejando fuera de los cálculos a

un total de n observaciones ($O \ll n \ll R$) ya sean del principio o del final de la serie: lo importante es cuidar que exista HOMOGENEIDAD ENTRE LOS GRUPOS y que todos ellos contengan el MISMO NUMERO DE OBSERVACIONES, para calcular estimaciones de la media y la desviación estándar dentro de cada grupo, de esta forma se tendrán H parejas de valores comparables $\{S_h, Z_h\}$, $h=1, \dots, H$ con los cuales podrá construirse la tabla 6, en donde, si $Z_{h,r}$ es la r -ésima observación del grupo h ,

$$Z_h = \sum_{r=1}^R Z_{h,r} / R, \quad S_h = \sqrt{\sum_{r=1}^R (Z_{h,r} - \bar{Z}_h)^2 / (R-1)}$$

además, el coeficiente de variación que aparece en dicha tabla, se calcula como la desviación estándar, de (λ) , entre la media $M(\lambda)$, es decir:

TABLA 6
Cálculos para Seleccionar la Potencia λ de la
Transformación Estabilizadora de Varianza.

Grupo	Potencia (λ)				
	-1	-0.5	0	0.5	1
1	S_1/Z_1^2	$S_1/Z_1^{1.5}$	S_1/Z_1	$S_1/Z_1^{0.5}$	S_1
2	S_2/Z_2^2	$S_2/Z_2^{1.5}$	S_2/Z_2	$S_2/Z_2^{0.5}$	S_2
...
h	S_h/Z_h^2	$S_h/Z_h^{1.5}$	S_h/Z_h	$S_h/Z_h^{0.5}$	S_h
...
H	S_H/Z_H^2	$S_H/Z_H^{1.5}$	S_H/Z_H	$S_H/Z_H^{0.5}$	S_H
Coficiente de Variación	$CV(-1)$	$CV(-0.5)$	$CV(0)$	$CV(0.5)$	$CV(1)$

$$cv(\lambda) = de(\lambda)/M(\lambda) \quad \text{con} \quad M(\lambda) = \sum_{h=1}^H (S_h/Z_h)^{1-\lambda} / H$$

y

$$de(\lambda) = \sqrt{\sum_{h=1}^H [(S_h/Z_h)^{1-\lambda} - M(\lambda)]^2 / (H-1)}$$

entonces, ya que se desea satisfacer la relación (3.3.1) con los valores muestrales, deberá elegirse la potencia λ que proporcione el MINIMO COEFICIENTE DE VARIACION, debido a que ese valor satisfará de manera aproximada la relación

$$\frac{S_h}{\bar{Z}_h^{1-\lambda}} = \text{constante, para } h=1, \dots, H$$

ya así, la transformación resultante será (2.1.2).

Nótese el espaciamento entre valores consecutivos de λ en la tabla 6, está es una mera sugerencia y podría ensayarse un espaciamento diferente al mostrado para encontrar, con la exactitud deseada, aquel valor de λ que vuelva mínimo el coeficiente de variación.

El método de Guerrero fué aplicado a la serie de ventas de Chatfield y Prothero que contiene $N=77$ observaciones. Se formaron $H=6$ grupos de $R=(77-5)/6=12$ observaciones cada uno. Inicialmente, el espaciamento se hizo de acuerdo a la sugerencia de Guerrero (tabla 6.1), despues fue de 0.1, en un intervalo de -1 a 1. Posteriormente, se hizo un refinamiento quedando el espaciamento en 0.001.

TABLA 6.1
DETERMINACION DE λ PARA LA SERIE DE VENTAS

Grupo	Potencia				
	-1	-0.5	0	0.5	1
1	0.004299	0.052098	0.63130	7.64976	92.6958
2	1.003082	0.041783	0.56520	7.68114	104.1448
3	0.003042	0.045184	0.67120	9.97065	148.1125
4	0.001602	0.028681	0.51333	9.18749	164.4364
5	0.001154	0.023245	0.46823	9.43161	189.9833
6	0.001012	0.022386	0.49503	9.89323	242.0687
$M(\lambda)$	0.002365	0.035562	0.55760	9.14456	156.9069
$DE(\lambda)$	0.001310	0.012469	0.08035	1.29541	55.4752
$CV(\lambda)$	0.554197	0.350623	0.14410	0.14165	0.3535

En la tabla 7 se muestran los valores para $\lambda = 0, 0.25, 0.253, 0.254, 0.255, 0.34$. En la tabla se observa que el mínimo coeficiente de variación (dentro del grupo de λ 's elegido) se logra con $\lambda = 0.254$.

La transformación $T(Z) = Z_t^{0.254}$, sugerida por la tabla 7, es prácticamente igual a la transformación sugerida por Box y Jenkins $T(Z) = Z^{0.25t}$, las cuales dan resultados muy similares a $Z^{0.34t}$.

Por lo anterior, para el modelo A se tiene:

$$(1 + 0.498)W_t = (1 - 0.798^{12}) a_t$$

$$\text{con } W_t = \nabla \nabla_{12} Z_t = \nabla \nabla_{12} Z_t^{0.254} \quad \text{y} \quad \sigma_a^2 = 0.021385$$

La tabla 8 muestra los pronósticos obtenidos con el modelo A y con $\lambda = 0.254$.

TABLA 7
DETERMINACION DE λ PARA LA SERIE DE VENTAS

Grupo	Potencia						
	0	0.25	0.253	0.254	0.255	0.34	1
1	.63130	2.19756	2.23070	2.24186	2.25308	3.44315	92.69582
2	.56520	2.08603	2.11891	2.12999	2.14112	3.33517	104.14486
3	.67120	2.58696	2.62918	2.64341	2.65771	4.20461	148.11257
4	.51333	2.17168	2.20960	2.22238	2.23524	3.65006	164.43641
5	.46823	2.10146	2.13967	2.15256	2.16552	3.60799	189.98331
6	.49503	2.32786	2.37151	2.38624	2.40106	4.06439	242.06874
$M(\lambda)$.55760	2.24526	2.28326	2.29607	2.30896	3.71756	156.90695
$DE(\lambda)$.08035	0.18831	0.19145	0.19252	0.19360	0.34514	55.47525
$CV(\lambda)$.14410	.08387	.0838497	.0838481	.0838490	.09284	0.35356

TABLA 8
Pronósticos para Z_t del modelo "A" ($\lambda=0.254$)
(y límites de tolerancia)

Adelanto	1	2	3	4	5	6
Pronóstico	284	407	523	735	924	976
95%	± 82	± 120	± 171	± 243	± 316	± 352

3.4.- Comparaciones teóricas.

Como se ha visto, es importante que antes de empezar un análisis de series de tiempo, el estadístico debe de considerar que "escala" o transformación debe ser usada, si es que el caso lo requiere.

También se debe tener en mente el fin para el cual se hace el análisis, es decir, si sólo se quiere explicar el fenómeno en estudio o se desea obtener un pronóstico, etc. y de

esa manera poder seleccionar el método más adecuado para obtener el parámetro λ .

De los tres métodos presentados en este trabajo, podemos hacer los siguientes comentarios: El Método de Box y Cox es un método que tiene una base teórica sólida, pero es un método muy elaborado (Box y Jenkins, 1973), que da precisión en el cálculo de la λ , pero su desarrollo se torna muy complejo (Chatfield y Prothero, 1973).

Este método fué aplicado a la serie de Chatfield y Prothero por Wilson (1973) usando 60 observaciones para el modelo A (previamente discutido) obteniendo $\lambda=0.34$.

Repetiendo este método pero utilizando todas las observaciones (77 casos) también para el modelo A, se obtiene $\hat{\lambda}=0.24$ (Chatfield y Prothero, 1973).

Chatfield y Prothero hacen referencia al tiempo que se tomó la computadora de la Universidad de Bath utilizando el paquete ICL, en calcular por Máxima Verosimilitud el parámetro λ , este fue más de 40 minutos, lo cual no fué de su agrado.

Definitivamente, este método necesita el auxilio de la rapidez de cálculo que tienen las computadoras, ya que si se desarrolla a "mano" su cálculo tomaría muchas horas-hombre.

Afortunadamente el desarrollo que ha tenido la tecnología en lo que a computadoras se refiere desde 1973, ha sido substancial, por lo que actualmente se pueden conseguir paquetes estadísticos para computadoras personales con la metodología de Box y Jenkins, en donde el tiempo de cálculo de estos parámetros se ha visto reducido considerablemente.

Esto nos indica que si el estadístico desea tener "precisión" en la λ , necesita tener acceso a una computadora que tenga implementada la metodología de Box y Jenkins, así como el método de Box y Cox para el cálculo de la λ o tener paciencia y tener disponibles "unas cuantas horas de trabajo". Desafortunadamente en el mercado de software, la mayoría de los paquetes estadísticos no contienen el método de Box y Cox para el cálculo de la λ .

El método de la gráfica Recorrido-Media es considerado por su propio autor, como un método totalmente empírico y muy poco "refinado"; esto es debido a que el objetivo del método es el de tratar de encontrar la "tendencia" de cada año y obtener una idea de que tipo de transformación se requiere inicialmente. Es definitivo que este método no daría una λ exacta pero sí algo aproximado, sin embargo existen series de tiempo que son poco sensibles a la selección del parámetro λ y esta situación podría ser reconocida ya que las gráficas de Z_t y $T(Z_t)$ son iguales o muy similares. En este caso no se aconsejaría utilizar este método y se podría recurrir a cualquiera de los otros dos aquí presentados.

Chatfield y Prothero opinan que este método es sumamente inapropiado para obtener aunque sea una idea de la "tendencia" de la serie, ya que si se decide utilizar sólo una porción de los datos el resultado será diferente, lo cual también sucede con el método de Máxima Verosimilitud.

El valor obtenido para el parámetro λ^* utilizando este método, fue de 0.25, el cual es muy cercano al $\hat{\lambda}=0.24$.

Como se puede ver, el método de Guerrero tiene una base teórica más sólida que el método anterior. En este método se pueden ensayar espaciamentos diferentes entre los valores de λ , haciendo refinamientos hasta lograr la exactitud deseada para encontrar el valor de λ que minimice el coeficiente de variación. El mismo autor señala que debe tenerse en cuenta que "el método está basado en aproximaciones y que, por consiguiente, conviene utilizarlo esencialmente para discriminar entre transformaciones que a priori se consideran apropiadas, más no para obtener estimaciones del parámetro λ con mucha exactitud". De acuerdo con esto, el método podría considerarse como un método complementario o preliminar.

Sin embargo la λ^o obtenida por este método no difiere mucho de la $\hat{\lambda}$ obtenida por Máxima Verosimilitud. Si el propósito del análisis es el de pronóstico, la transformación encontrada por el método propuesto por Guerrero podría resultar benéfica, ya que el supuesto de estacionariedad se vería satisfecho.

Este método es de fácil implementación, no necesita de una programación laboriosa y puede ser calculado en una hoja electrónica de cálculo (i.e. LOTUS 123) o en el peor de los casos, se puede desarrollar a mano con una calculadora que tenga la función de desviación estándar implementada. La $\hat{\lambda}$ obtenida con el método de Guerrero para la serie de Chatfield y Prothero es 0.254.

Nótese que el método de la Gráfica Recorrido-Media también es muy fácil de desarrollar y su cálculo no tomaría mucho tiempo.

Resumiendo los resultados obtenidos al aplicar los métodos aquí expuestos a la serie de Chatfield y Prothero, se tiene la siguiente tabla. (Recuérdese que Wilson utilizó solo 60 observaciones por lo tanto no está en igualdad de circunstancias que los demás métodos)

TABLA 9
Pronósticos con las diferentes transformaciones

1	Chatfield &Prothero	M.M.V. Wilson	Máxima Verosim.	Recorrido - Media	Método Guerrero	Obs. Reales
	$\log Z_t$	$Z^{0.34}_t$	$Z^{0.24}_t$	$Z^{0.25}_t$	$Z^{0.254}_t$	
78	305	304	284	286	284	260
79	482	423	410	409	407	304
80	673	574	529	511	523	390
81	990	706	743	761	735	614
82	1297	858	935	966	924	783
83	1387	902	988	1091	976	872
M_1	-318.5	-90.6	-111	-133.5	-104.3	
M'_1	-55.72%	-21.87%	-22.24%	-24.66%	-21.14	

Como se puede observar en la Tabla 10 con los métodos de Box y Cox y el de Guerrero se obtienen parámetros iguales (a dos decimales) y sus pronósticos son parecidos y muy cercanos a las observaciones reales (Jun-Nov 1971).

TABLA 10
PARAMETROS DEL MODELO A

	$\log Z_t$	$Z^{0.34}_t$	$Z^{0.24}_t$	$Z^{0.25}_t$	$Z^{0.254}_t$
AR	-0.47	-0.37	-0.49	-0.5	-0.49
SMA	+0.81	0.79	0.79	0.8	-0.79

Los resultados que presenta el de la Gráfica Recorrido-Media también son muy cercanos a los de los otros métodos, pero los obtenidos con el logaritmo definitivamente no son apropiados.

Nótese que se presenta la media de los errores de pronóstico (M_1) siendo todas negativas por lo que concluimos que los pronósticos tienden a sobreestimar los valores de la serie, sin embargo el mínimo error se obtiene para $Z_0.25^4$ (M'_1).

Se define a M_1 como:

$$M_1 = \sum_{h=1}^H e_t(h)/H$$

donde :

$$e_t(h) = T(Z_{t+h}) - \hat{T}(Z_t)(h)$$

Los pronósticos obtenidos se comparan sólo con seis observaciones debido a que son las únicas observaciones reales disponibles (Charfield y Prothero, 1973).

Cabe mencionar que estamos comparando el inverso del pronóstico $\{Z_t/\lambda_t\}$ con las observaciones reales y que no se ha hecho ningún tipo de corrección por el sesgo introducido por la transformación potencia. Este tipo de corrección se hará en el capítulo siguiente.

CAPITULO 4

**Presentación y comparación de los diferentes métodos
para corregir pronósticos por sesgo.**

En el capítulo 1 se hizo mención a las etapas que Box y Jenkins sugieren seguir como estrategia de construcción de modelos: La última etapa de esta metodología trata del uso del modelo para los fines por el cual se haya realizado el estudio. Esta etapa esta íntimamente ligada con la realización de pronósticos.

En la mayoría de los casos, los pronósticos serán requeridos para la serie observada $\{Z_t\}$ y no para la transformada $\{T(Z_t)\}$, la cual puede carecer de interpretación en la práctica.

En una primera instancia, resultaría lógico pensar que, mediante aplicación de la transformación inversa T^{-1} (suponiendo que esta exista) se podrían obtener los pronósticos para la serie original $\{Z_t\}$; sin embargo, las propiedades óptimas del pronóstico $T(Z_t)(h)$ no se preservan necesariamente si la transformación $T(\cdot)$ es no-lineal y se usa a:

$$T^{-1}[T(Z_t)(h)] \quad (4.1)$$

como pronóstico de Z_{t+h} .

Algunas formas alternativas para obtener pronósticos insesgados han sido propuestas por Neyman y Scott (1960), Granger y Newbold (1976), Miller (1984) y Taylor (1986). Estos métodos serán presentados a continuación y se introducirá el método propuesto por Guerrero (1983).

4.1.- Método de Neyman y Scott.

Neyman y Scott (1960) consideran que el problema de la "Normalización" tiene dos diferentes facetas: la primera es referente a la identificación de la transformación apropiada para la variable en estudio, siguiendo una distribución con forma y propiedades particulares y la segunda con la naturaleza de los estadísticos capaces de servir como estimadores insesgados en caso de que la transformación elegida sea adecuada o exitosa para el experimento.

El propósito de estos dos autores, es el de obtener el estimador insesgado con varianza mínima de los resultados del experimento, expresado en las unidades originales de la variable.

Los estimadores expresados en las unidades originales de una serie, se acostumbra obtener con la transformación inversa de los estimadores en unidades transformadas. Como es sabido, este estimador es sesgado.

La suposición básica de la investigación de Neyman y Scott es que la transformación usada en el análisis de un experimento no es errónea, por lo que las variables transformadas siguen exactamente una distribución Normal con una media dada y con la misma varianza σ^2 desconocida.

Denótese como $\xi(\psi)$ a la variable transformada. La variable que es directamente observada será denotada por $x(\psi)$. se supondrá entonces que :

$$x(\psi) = f [\xi(\psi)] \quad (4.1.1)$$

donde f es una función estrictamente creciente definida para los reales como su argumento. Después se introducirán más

limitaciones a f . Nótese que f es la función inversa de la transformación usada en la variable observada X .

El problema tiene que ver con un par de variables específicas de la familia considerada, nombradas $\xi(\mu)$ y $x(\mu)$ donde μ es la media de $\xi(\mu)$ que está bien definida, pero se desconoce su valor. Específicamente interesa estimar

$$\theta = E[x(\mu)] \quad (4.1.2)$$

El problema surge cuando las variables $\xi(\mu)$ y $x(\mu)$ no son directamente observables; por otro lado, las variables que son observadas en el experimento nos llevan a los estimadores $\hat{\mu}$ y S^2 , mutuamente independientes y que conjuntamente son suficientes para μ y σ^2 . La primera es una variable con distribución Normal con media μ y varianza $\lambda^2 \sigma^2$, donde λ^2 es un número conocido. El segundo estadístico, S^2 , es la suma de los cuadrados de los residuales y , dividido por σ^2 , se distribuye como una χ^2 con un cierto número ν de grados de libertad.

El objetivo de Neyman y Scott es el de proyectar una función, dígame $\hat{\theta}(\hat{\mu}, S^2)$ tal que su esperanza sea igual a θ . Debido al resultado obtenido por Lehmann y Scheffe (1950) de que el conjunto de estadísticos $(\hat{\mu}, S^2)$ está totalmente limitado, se sigue que la función $\hat{\theta}(\hat{\mu}, S^2)$ es única y es el estimador insesgado con varianza mínima para θ .

Puesto que las transformaciones Normalizantes se suponen como un cambio de escala de la variable aleatoria observada, es natural suponer que la función f que determina la variable aleatoria X en términos de la variable Normal ξ es más o menos regular.

El método de Neyman y Scott para construir $\hat{\theta}(\hat{\mu}, S^2)$ está limitado al caso donde :

- i) $\theta = E\{f[\xi(\mu)]\}$ exista
- ii) f es una función continua y
- iii) La esperanza de θ puede ser obtenida a partir de las esperanzas de cada término de la expansión de Taylor para f , entonces:

$$\theta = E\{f[\xi(\mu)]\} = f(0) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)} E[\xi^n(\mu)], \quad (4.1.3)$$

donde $f^{(n)}$ existe para la n -ésima derivada de f evaluada en cero. Entonces, para cada n , se determina una combinación homogénea

$$T_n = \sum_{k=0}^n A_{n,k} \hat{\mu}^k S^{n-k} \quad (4.1.4)$$

tal que

$$E(T_n) = E[\xi^n(\mu)] \quad (4.1.5)$$

por lo cual :

$$\hat{\theta}(\hat{\mu}, S^2) = f(0) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)} T_n \quad (4.1.6)$$

es la solución al problema.

También, para una familia f en particular, Neyman y Scott dan un método alternativo para obtener $\hat{\theta}(\hat{\mu}, S^2)$. Para tal efecto hacen uso de ciertas fórmulas y calculan límites. Así, para $m=1,2$ se tiene

$$E[S^{2m}] = (2\sigma^2)^m \left(\frac{1}{2} \nu + m\right) / \Gamma\left(\frac{1}{2} \nu\right) \quad (4.1.7)$$

también

$$E[\xi^{2m}(\mu)] = \sum_{k=0}^m \frac{(2m)!}{(2k)!(m-k)!} \mu^{2k} (\sigma^2/2)^{m-k} \quad (4.1.8)$$

de donde se sigue que

$$\frac{(2m)!}{m!} (\sigma^2/2)^m \leq E[\xi^{2m}(\mu)] \leq \frac{(2m)!}{m!} [(\mu^2 + \sigma^2)/2]^m \quad (4.1.9)$$

similarmente

$$E[\xi^{2m+1}(\mu)] = \mu \sum_{k=0}^m \frac{(2m+1)!}{(2k+1)!(m-k)!} \mu^{2k} (\sigma^2/2)^{m-k} \quad (4.1.10)$$

para obtener cotas de $E[|\xi^{2m+1}|]$, primero consideraron que

$$|\mu| \frac{(2m+1)!}{m!} (\sigma^2/2)^m \leq |E[\xi^{2m+1}(\mu)]| \leq E[|\xi^{2m+1}(\mu)|] \quad (4.1.11)$$

Además, por la desigualdad de Schwarz y por (4.1.9)

$$\begin{aligned} E[|\xi^{2m+1}(\mu)|] &\leq (E[\xi^2(\mu)] E[\xi^{4m}(\mu)])^{1/2} \quad (4.1.12) \\ &\leq (\mu^2 + \sigma^2)^{1/2} \left[\frac{(4m)!}{(2m)!} \right]^{1/2} [(\mu^2 + \sigma^2)/2]^m. \end{aligned}$$

Por lo cual se deduce que

$$\frac{m!}{2^m (2m+1)!} \left[\frac{(4m)!}{(2m)!} \right]^{1/2} = \left[\prod_{k=1}^m \frac{(4k-3)(4k-1)}{(4k+2)^2} \right]^{1/2} \quad (4.1.13)$$

consecuentemente se puede escribir

$$|\mu| \frac{(2m+1)!}{m!} (\sigma^2/2)^m \leq E[|\xi^{2m+1}(\mu)|] \leq \frac{(2m+1)!}{m!} [\mu^2 + \sigma^2]^m \quad (4.1.14)$$

para toda m .

Para poder evaluar la esperanza término a término, de la serie de Taylor, Neyman y Scott usan las cotas anteriormente expuestas para probar los siguientes teoremas.

TEOREMA 1.- Para que la serie del término derecho de (4.1.3), sea convergente irrespectivamente a los valores de μ y σ^2 , es necesario y suficiente que el radio de convergencia de las dos series

$$\sum_{n=0}^{\infty} \frac{1}{n!} f(2n) Z^n \quad \text{y} \quad \sum_{n=0}^{\infty} \frac{1}{n!} f(2n+1) Z^n \quad (4.1.15)$$

sea infinito, por lo tanto (4.1.16)

$$\begin{aligned} \lim_{n \rightarrow \infty} \left[\frac{1}{n!} |f(2n)| \right]^{1/n} &= \lim_{n \rightarrow \infty} \left[\frac{1}{n!} |f(2n+1)| \right]^{1/n} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n \cdot \infty n} \left[|f(2n)| \right]^{1/n} = \lim_{n \rightarrow \infty} \frac{1}{n \cdot \infty n} \left[|f(2n+1)| \right]^{1/n} = 0. \end{aligned}$$

Nótese que la condición anterior es más fuerte que la suposición de que la expansión de Taylor para f

$$f(\xi) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)} \xi^n \quad (4.1.17)$$

sea convergente para todos los reales. Es decir, la condición suficiente y necesaria para que esto suceda puede ser escrita como (4.1.18)

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} (f^{(2n)})^{1/n} = \lim_{n \rightarrow \infty} \frac{1}{n^2} (f^{(2n+1)})^{1/n} = 0$$

Entonces, si el radio de convergencia de (4.1.15) es infinito, será infinito también para (4.1.17), pero lo contrario no necesariamente es verdadero. (Para profundizar remítase a Neyman y Scott, 1960).

TEOREMA 2.- Bajo las condiciones del **TEOREMA 1**, esto es, de acuerdo a la condición (4.1.16)

$$\theta = E[f(\xi(\mu))] = f(0) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)} E[\xi^n(\mu)], \quad (4.1.19)$$

para toda μ y σ^2 .

En otras palabras, si la condición (4.1.16) se satisface, entonces la esperanza de f se puede obtener mediante la esperanza de cada uno de los términos de la expansión de Taylor de esta función.

Por conveniencia de referencia se adopta la siguiente definición:

DEFINICION.- Una función f es llamada de segundo grado si satisface la condición (4.1.16).

Se verá que para toda función diferenciable infinita, cuyas derivadas están acotadas en un punto, es necesaria una función de segundo grado considerando que la suma de dos funciones de segundo grado es también una función de segundo grado.

De ahora en adelante se hará referencia exclusivamente a funciones $f(\xi)$, las cuales son completamente de segundo grado.

Sea $\hat{\mu}$ una variable con distribución Normal con media μ y varianza $\lambda^2 \sigma^2$ con λ^2 conocida. También, sea S^2 una variable independiente de $\hat{\mu}$ y tal que S^2/σ^2 se distribuye como una χ^2 con ν grados de libertad. Finalmente, para $n=0,1,2,\dots$ sea

(4.1.20)

$$T_{2n} = \sum_{k=0}^n \frac{(2n)!}{((2k)!(n-k)!)} \hat{\mu}^{2k} [1/4S^2(1-\lambda^2)]^{n-k} \frac{\Gamma(1/2\nu)}{\Gamma(1/2\nu+n-k)}$$

y

$$T_{2n+1} = \sum_{k=0}^n \frac{(2n+1)}{(2k+1)!(n-k)!} \hat{\mu}^{2k+1} [1/4S^2(1-\lambda^2)]^{n-k} \frac{\Gamma(1/2\nu)}{\Gamma(1/2\nu+n-k)}$$

Mediante cálculos directos se verifica que, para toda

En otras palabras, si la condición (4.1.16) se satisface, entonces la esperanza de f se puede obtener mediante la esperanza de cada uno de los términos de la expansión de Taylor de esta función.

Por conveniencia de referencia se adopta la siguiente definición:

DEFINICION.- Una función f es llamada de segundo grado si satisface la condición (4.1.16).

Se verá que para toda función diferenciable infinita, cuyas derivadas están acotadas en un punto, es necesaria una función de segundo grado considerando que la suma de dos funciones de segundo grado es también una función de segundo grado.

De ahora en adelante se hará referencia exclusivamente a funciones $f(\xi)$, las cuales son completamente de segundo grado.

Sea $\hat{\mu}$ una variable con distribución Normal con media μ y varianza $\lambda^2 \sigma^2$ con λ^2 conocida. También, sea S^2 una variable independiente de $\hat{\mu}$ y tal que S^2/σ^2 se distribuye como una χ^2 con ν grados de libertad. Finalmente, para $n=0,1,2,\dots$, sea

(4.1.20)

$$T_{2n} = \sum_{k=0}^n \frac{(2n)!}{((2k)! (n-k)!)} \hat{\mu}^{2k} [1/4S^2(1-\lambda^2)]^{n-k} \frac{\Gamma(1/2\nu)}{\Gamma(1/2\nu+n-k)}$$

y

$$T_{2n+1} = \sum_{k=0}^n \frac{(2n+1)!}{(2k+1)! (n-k)!} \hat{\mu}^{2k+1} [1/4S^2(1-\lambda^2)]^{n-k} \frac{\Gamma(1/2\nu)}{\Gamma(1/2\nu+n-k)}$$

(4.1.21)

Mediante cálculos directos se verifica que, para toda

$m=1,2,\dots$, se tiene

$$E(T_m) \cong E[\xi^m(\mu)] \quad (4.1.22)$$

Neyman y Scott derivan un tercer teorema.

TEOREMA 3.— Si f es una función de segundo grado, entonces

$$\theta(\hat{\mu}, S^2) = f(0) + \sum_{m=1}^{\infty} \frac{1}{n!} f^{(m)} T_m \quad (4.1.23)$$

es convergente para todos los valores de $\hat{\mu}$ y S^2 , y es un estimador insesgado de $\theta = E\{f[\xi(\mu)]\}$.

Comparando (4.1.20) y (4.1.21) con (4.1.8) y (4.1.10) nótese que

$$\prod (1/2v) / \prod (1/2v + n - k) \ll (2/v)^{n-k} \quad (4.1.24)$$

haciendo referencia a (4.1.9), se tiene que

$$|T_{2n}| \ll \frac{(2n)!}{n!} Y^n \quad (4.1.25)$$

De manera similar

$$|T_{2n+1}| \ll |\hat{\mu}| \frac{(2n+1)!}{n!} Y^n \quad (4.1.26)$$

con

$$Y = [v\hat{\mu}^2 + S^2(1+\lambda^2)]/2v \quad (4.1.27)$$

Debido a que f es una función de segundo grado, se sigue que la serie (4.1.23) es absolutamente convergente para todos los valores de $\hat{\mu}$ y S^2 . Para probar que la esperanza de $\theta(\hat{\mu}, S^2)$ definida en (4.1.23) puede ser calculada a partir de las esperanzas de cada término, es suficiente mostrar la convergencia de la serie obtenida de (4.1.23) sustituyendo cada T_m por la esperanza de su valor absoluto. El desarrollo se facilita si se advierte que $|T_m|$ no puede exceder la expresión obtenida al sustituir en (4.1.20) y (4.1.21) el valor de $\hat{\mu}$ por $|\hat{\mu}|$ y $1-\lambda^2$,

el cual puede ser negativo, por $1+\lambda^2$. Cálculos similares a los anteriores usando (4.1.9) y (4.1.14) indican que:

$$\sum_{n=1}^{\infty} \frac{1}{n!} |f^{(n)}| E|T_n| < +\infty \quad (4.1.28)$$

Debido a (4.1.22) se sigue que $\hat{\theta}(\hat{\mu}, S^2)$, según se define en (4.1.23), tiene la propiedad deseada de ser un estimador insesgado con varianzá mínima de θ , cualquiera que sea la función continua de segundo grado.

Neyman y Scott en su artículo dan una solución alternativa, aplicable a funciones continuas de segundo grado, la cual será descrita a continuación.

Definen a f como una función de "tipo recursivo" si satisface la ecuación diferencial de segundo grado

$$f''(x) = A + Bf(x) \quad (4.1.29)$$

donde A y B son constantes. Para eliminar el caso trivial donde f es lineal, ellos asumen que al menos una de las constantes es diferente de cero. Se puede verificar que cada función recursiva es necesariamente una función de segundo grado.

Neyman y Scott están particularmente interesados en la esperanza de la expansión de Taylor en μ . Debido a que los momentos centrales impares de la variable Normal son iguales a cero, consideran solamente las derivadas de f de orden par. Se tiene para toda n , que

$$f^{(2n)}(x) = AB^{n-1} + B^n f(x) \quad (4.1.30)$$

y si $B \neq 0$,

$$\theta = E\{f(\xi(\mu))\} = f(\mu) + \sum_{n=1}^{\infty} \frac{1}{n!} [AB^{n-1} + B^n f(\mu)] (\sigma^2/2)^n \quad (4.1.31)$$

$$= f(\mu)e^{B\sigma^2/2} + \frac{A}{B} (e^{B\sigma^2/2} - 1)$$

Alternativamente, si $B=0$, esto es si f es cuadrática

$$B = f(\mu) + A\sigma^2/2. \quad (4.1.32)$$

Similarmente para $B \neq 0$

$$E[f(\hat{\mu})] = f(\mu)e^{B\lambda^2\sigma^2/2} + (A/B)(e^{B\lambda^2\sigma^2/2} - 1) \quad (4.1.33)$$

y para $B = 0$,

$$E[f(\hat{\mu})] = f(\mu) + A\lambda^2\sigma^2/2 \quad (4.1.34)$$

Eliminando $f(\mu)$ de (4.1.31), (4.1.33), (4.1.32) y (4.1.34), se tiene

$$B = e^{B(1-\lambda^2)\sigma^2/2} E[f(\hat{\mu})] + (A/B)[e^{B(1-\lambda^2)\sigma^2/2} - 1] \quad (4.1.35)$$

para $B \neq 0$, y

$$B = E[f(\hat{\mu})] + A(1-\lambda^2)\sigma^2/2 \quad (4.1.36)$$

para $B=0$.

Esta última fórmula indica que, cuando $B=0$, el estimador insesgado con varianza mínima de B está dado por

$$\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu}) + A(1-\lambda^2)S^2/2v \quad (4.1.37)$$

Si $B \neq 0$, entonces para obtener $\hat{\theta}(\hat{\mu}, S^2)$ es suficiente con determinar una función, dígase $\mathbb{Q}(aS^2, v)$ independiente de $\hat{\mu}$, tal que su esperanza sea $\exp(a\sigma^2/2)$. Tomando en cuenta la expresión

$$e^{a\sigma^2/2} = \sum_{n=0}^{\infty} \frac{1}{n!} (a\sigma^2/2)^n \quad (4.1.38)$$

se tiene que

$$\begin{aligned} \mathbb{Q}(aS^2, v) &= \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\Gamma(1/2v)}{\Gamma(1/2v+n)} (aS^2/4)^n \quad (4.1.39) \\ &= \left[\frac{2}{S\sqrt{a}} \right]^{1/2v-1} \Gamma(1/2v) {}_1F_1(1/2v-1; S\sqrt{a}) \end{aligned}$$

donde $I_\lambda(x)$ es la función de Bessel con argumento imaginario.

Se sigue de (4.1.35) que

$$\hat{\theta}(\hat{\mu}, S^2) = \theta [B(1-\lambda^2)S^2, \nu] [f(\hat{\mu}) + (A/B)] - (A/B) \quad (4.1.40)$$

la cual es una fórmula general para el estimador insesgado con varianza mínima, correspondiente al caso donde f es una función recursiva. Se verá que $\hat{\theta}(\hat{\mu}, S^2)$ en general, es una función lineal del estimador tradicional $f(\hat{\mu})$ de θ , con coeficientes dependientes de S^2 , λ^2 y ν .

Si $\lambda^2=1$, esto es, si la varianza de $\hat{\mu}$ coincide con la de $f(\hat{\mu})$, entonces $\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu})$. De otra forma $f(\hat{\mu})$ es sesgado. En el caso particular de $B=0$, la corrección de sesgo es aditiva como se indica en (4.1.37). Si $A=0$ pero $B \neq 0$, entonces la corrección por sesgo en $f(\hat{\mu})$ es multiplicativa. Finalmente, si A y B difieren de cero, se tendrá una corrección combinada multiplicativa y aditiva.

La importancia del sesgo en estimadores tradicionales de θ puede ser evaluada resolviendo las ecuaciones (4.1.35) y (4.1.36) para $E[f(\hat{\mu})]$. Así se tiene que

$$E[f(\hat{\mu})] = [B + (A/B)]e^{-B(1-\lambda^2)\sigma^2/2} - (A/B) \quad (4.1.41)$$

para $B \neq 0$, y

$$E[f(\hat{\mu})] = B - A(1-\lambda^2)\sigma^2/2 \quad (4.1.42)$$

para $B=0$.

Neyman y Scott usan los resultados obtenidos en su artículo para deducir algunas fórmulas particulares. Obtienen el estimador insesgado de θ con varianza mínima y la esperanza de $f(\hat{\mu})$ referentes a la transformación logarítmica y a la raíz cuadrada.

En el caso de la raíz cuadrada, la variable transformada es

$$\xi = (x - a)^{1/2} \quad (4.1.43)$$

donde a es una constante conocida. Ellos no toman en cuenta la ambigüedad del hecho de que para que ξ sea una variable Normal, debe ser capaz de manejar valores negativos.

La función f es

$$x = f(\xi) = \xi^2 + a \quad (4.1.44)$$

la cual es una función recursiva con $A=2$ y $B=0$. Consecuentemente la fórmula (4.1.37) nos lleva a:

$$\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu}) + (1-\lambda^2)S^2/v \quad (4.1.45)$$

$$= \hat{\mu}^2 + a + (1-\lambda^2)S^2/v$$

El sesgo del estimador tradicional $f(\hat{\mu}) = \hat{\mu}^2 + a$ se obtiene de (4.1.42), esto es

$$E[f(\hat{\mu})] = \theta - (1 - \lambda^2)\sigma^2 \quad (4.1.46)$$

Por lo tanto, a menos que $\lambda^2 > 1$, la varianza de $\hat{\mu}$ es por lo menos igual a $\xi(\mu)$; el uso sistemático de $f(\mu)$ puede sub-estimar θ . Más aún, entre mejor se estime $\hat{\mu}$, es decir, entre más pequeño sea el valor de λ^2 , mayor será el sesgo.

En el caso de la transformación logarítmica, se tienen

$$\xi = \text{LOG}_{10} x \quad (4.1.47)$$

y por lo tanto

$$x = f(\xi) = 10^\xi = e^{m\xi} \quad (4.1.48)$$

Nótese que aquí también la función f es de tipo recursivo con $A=0$ y $B=m^2$. La fórmula (4.1.40) da como resultado

$$\hat{\theta}(\hat{\mu}, S^2) = \Phi[m^2(1-\lambda^2)S^2, v] f(\hat{\mu}) \quad (4.1.49)$$

$$= \Phi[m^2(1-\lambda^2)S^2, v] 10^{\hat{\mu}^2}$$

Substituyendo $A=0$ y $B=m^2$ en (4.1.41) se tiene

$$E[f(\hat{\mu})] = E[\hat{\mu}] = \theta e^{-m^2(1-\lambda^2)\sigma^2/2} \quad (4.1.50)$$

Entonces, con la Transformación Logarítmica el sesgo del estimador tradicional es multiplicativo. Si la varianza de $\hat{\mu}$ es menor que σ^2 entonces el uso de $\hat{\mu}$ sistemáticamente sub-estimaré θ y viceversa. El sesgo crece cuando crece $|1-\lambda^2|$.

Neyman y Scott concluyen su artículo explicando que la fórmula (4.1.39) en donde se define \mathbb{D} , parece ser complicada, pero de hecho la serie del término derecho de la fórmula, converge más o menos rápido, por lo que se puede tener mucha precisión con pocos términos. Además, consideran que la fórmula (4.1.39) puede reescribirse como sigue

$$\begin{aligned} \mathbb{D}(a\sigma^2, \nu) &= \sum_{n=0}^{\infty} \frac{(a\sigma^2/2)^n}{n! \prod_{k=1}^n (\nu+2k-2)} \\ &= \sum_{n=0}^{\infty} \frac{(a\hat{\sigma}^2/2)^n}{n! \prod_{k=1}^n \left(1 + \frac{(2k-2)}{\nu}\right)} \end{aligned} \quad (4.1.51)$$

donde

$$\hat{\sigma}^2 = \sigma^2/\nu \quad (4.1.52)$$

es el estimador insesgado de σ^2 . Se considera que para $n \gg 1$, el valor absoluto de cada término en el lado derecho de la fórmula, es menor que su término correspondiente en la expansión $\exp(a\sigma^2/2)$. En otras palabras, las series (4.1.51) convergen más rápido que las series para la función exponencial.

4.2.- Método de Granger y Newbold.

Nuevamente, bajo la suposición de que se ha identificado un modelo para el proceso X_t , al cual se le ha aplicado una transformación $Y_t = T(X_t)$ y suponiendo que X_t es estacionaria, o que puede ser estacionaria aplicando diferencias, Granger Y Newbold analizan, en su artículo publicado en 1976, la estructura de las autocovarianzas de la serie $\{Y_t\}$ y sugieren 4 métodos alternativos para pronosticar las series transformadas y su expresión obtenida para el sesgo cuando no se utiliza el pronóstico óptimo.

Su análisis emplea polinomios Hermitianos (consúltese el apéndice para sus propiedades básicas), de esa manera se obtienen resultados derivados de una clase muy general de transformaciones instantáneas.

Sea $\{X_t\}$ una serie estacionaria con media μ , varianza σ^2 y una estructura de autocorrelaciones $\text{corr}(X_t, X_{t-r}) = \rho_r$. Sea

$$Z_t = \frac{X_t - \mu}{\sigma} \quad \text{para toda } t.$$

Entonces Z_t Y Z_{t-r} serán conjuntamente distribuidas como una Normal bivariada con media cero, varianza uno y función de autocorrelación $\{\rho_r\}$.

Considérese ahora una transformación instantánea de la forma $Y_t = T(Z_t)$ con una expansión $T(Z) = \sum_{j=0}^M \alpha_j H_j(Z)$. Entonces

$$E(Y_{t-r}) = E\left\{ \sum_{j=0}^M \alpha_j H_j(Z_{t-r}) \right\} = \alpha_0 \quad (4.2.1)$$

por (A.2) se tiene que la serie transformada tiene media α_0 . Más aún, usando (A.4)

$$E(Z_t Y_{t-r}) = E\left\{H_1(Z_t) \sum_{j=0}^M \alpha_j H_j(Z_{t-r})\right\} = \alpha_1 \rho_r \quad (4.2.2)$$

y por lo tanto la covarianza entre la serie transformada y la serie original es $\text{cov}(X_t Y_{t-r}) = \alpha_1 \rho_r \sigma$. Usando (4.2.2), la secuencia de autocovarianzas de la serie transformada es dada por

$$C_{Y,r} = \text{cov}(Y_t Y_{t-r}) = E\left\{\sum_{j=1}^M \alpha_j H_j(Z_t) \sum_{l=1}^M \alpha_l H_l(Z_{t-r})\right\} = \sum_{j=1}^M \alpha_j^2 j! \rho_r^j$$

por (A.1). En particular, fijando $r=0$ en esta expresión se tiene

$$\text{var}(Y_t) = \sum_{j=1}^M \alpha_j^2 j!$$

Entonces la secuencia de autocorrelaciones de la serie transformada es

$$\rho_{Y,r} = \sum_{j=1}^M \alpha_j^2 j! \rho_r^j / \sum_{j=1}^M \alpha_j^2 j!$$

Por lo tanto, si $\rho_r \neq 0$ para alguna $r = 0$, se sigue que

$$|\rho_{Y,r}| < |\rho_r| \quad \text{para } M > 1.$$

en este sentido, la serie transformada estará más cerca al ruido blanco que la serie original.

Ejemplificando lo anterior, considérese la transformación cuadrática $Y_t = a + bX_t + cX_t^2$, que puede ser escrita como

$$Y_t = T(Z_t) = \alpha_0 H_0(Z_t) + \alpha_1 H_1(Z_t) + \alpha_2 H_2(Z_t) = \alpha_0 + \alpha_1 Z_t + \alpha_2 (Z_t^2 - 1) \quad (4.2.4)$$

donde $\alpha_0 = a + b\mu + c(\mu^2 + \sigma^2)$, $\alpha_1 = (b + 2c\mu)\sigma$, $\alpha_2 = c\sigma^2$

se sigue de (4.2.3) que la secuencia de autocovarianzas para la serie transformada es

$$C_{Y,r} = \alpha_1^2 \rho_r + 2\alpha_2^2 \rho_{2r} \quad (4.2.5)$$

donde α_1 y α_2 están dadas en (4.2.4). Como se ha visto en

(4.2.1), la media de la serie transformada es α_0 .

El propósito de Granger Y Newbold con lo anterior, fue el de relacionar las propiedades de las autocovarianzas de una serie de tiempo en particular, a aquellas de una cierta transformación instantánea de esa serie. Ellos consideran que esto es importante debido a que la transformación potencia de Box-Cox se utiliza frecuentemente en el análisis de series de tiempo univariadas, y más aún, a que el método de Box y Jenkins (1970) en la etapa de identificación de modelos, se basa en la identificación de patrones típicos de comportamiento en la función de autocorrelación (y en la función de autocorrelación parcial) .

Sus resultados indican que si la serie a ser ajustada es sujeta a una transformación instantánea, puede ser el caso (excepto por los procesos MA) que la función de autocorrelación de la serie transformada muestre marcadas diferencias en los patrones de comportamiento comparadas con aquellas de la serie original.

Granger Y Newbold aplican el método anterior a series que requieren una diferencia para obtener estacionariedad.

Sea X_t una serie de tiempo tal que :

$$\nabla X_t = u_t \quad (4.2.6)$$

donde u_t es una serie estacionaria Gaussiana con media cero y varianza $\sigma^2 u$. Se considera que (4.2.6) ha sido generada por el modelo de la forma

$$X_t = \mu + \sum_{j=1}^t u_j \quad (4.2.7)$$

i.e. empezando en $t=0$ con $X_0 = \mu$ escribáse:

$$C_{t,r} = E \left(\sum_{j=1}^t u_j \cdot \sum_{i=1}^r u_i \right) \quad (4.2.8)$$

Entonces por (4.2.2) y (4.2.3) X_t es normalmente distribuida con media μ y Varianza $C_{t,r}$. Ahora definamos a $Z_t = (X_t - \mu) / \sqrt{C_{t,r}}$.

Entonces Z_t y Z_{t-r} serán distribuidas conjuntamente como una Normal bivariada con media cero, varianza uno y correlación

$$\rho_{t,t-r} = C_{t,t-r} / \sqrt{(C_{t,t} C_{t-r,t-r})}.$$

Se considera una transformación instantánea de la forma

$$Y_t = T(Z_t) = \sum_{j=0}^M \alpha_j^{(t)} H_j(Z_t) \quad (4.2.9)$$

Nótese que $\alpha_j^{(t)}$ depende de t . Usando los argumentos anteriores (respecto a la función de autocorrelación), se sigue que

$$E(Y_t) = \alpha_0^{(t)} \quad (4.2.10)$$

Más aún, un argumento análogo al visto anteriormente, nos lleva a

$$\text{cov}(Y_t, Y_{t-r}) = \sum_{j=1}^M \alpha_j^{(t)} \alpha_j^{(t-r)} j! \rho_{t,t-r}^j$$

por lo tanto

$$E(Y_t Y_{t-r}) = \sum_{j=1}^M \alpha_j^{(t)} \alpha_j^{(t-r)} j! \rho_{t,t-r}^j + \alpha_0^{(t)} \alpha_0^{(t-r)} \quad (4.2.11)$$

El interés de Granger y Newbold está en ∇Y_t más que en Y_t , y se sigue de (4.2.10) y (4.2.11) que

$$E(\nabla Y_t) = \alpha_0^{(t)} - \alpha_0^{(t-1)} \quad (4.2.12)$$

y

$$\text{cov}(\nabla Y_t, \nabla Y_{t-r}) = \sum_{j=1}^M \alpha_j^{(t)} \alpha_j^{(t-r)} j! \rho_{t,t-r}^j \quad (4.2.13)$$

$$\begin{aligned}
& + \sum_{j=1}^M \alpha_j^{(t-1)} \alpha_j^{(t-r-1)} j! \rho_{t-1, t-r-1}^j \\
& - \sum_{j=1}^M \alpha_j^{(t)} \alpha_j^{(t-r-1)} j! \rho_{t, t-r-1}^j \\
& - \sum_{j=1}^M \alpha_j^{(t-1)} \alpha_j^{(t-r)} j! \rho_{t-1, t-r}^j
\end{aligned}$$

Las fórmulas (4.2.12) y (4.2.13) son resultados totalmente generales. Como ejemplo de lo anterior, considérese nuevamente la transformación cuadrática que puede ser escrita como

$$Y_t = T(Z_t) = \alpha_0^{(t)} + \alpha_1^{(t)} Z_t + \alpha_2^{(t)} Z_t^2$$

donde (4.2.14)

$$\alpha_0^{(t)} = a + b\mu + c(\mu^2 + C_{t,t}), \alpha_1^{(t)} = (b + 2c\mu) C_{t,t}^{1/2}, \alpha_2^{(t)} = c C_{t,t}$$

Se sigue de (4.2.12) que la media de ∇Y_t está dada por

$$E(\nabla Y_t) = c(C_{t,t} - C_{t-1, t-1}) = cE(u_t^2 + 2u_t \sum_{j=1}^{t-1} u_j) \quad (4.2.15)$$

Granger y Newbold notaron que para valores grandes de t , esto es, para la serie empezada en $t=1$ en (4.2.2), $E(\nabla Y_t)$ se acerca a una constante si u_t constituye un proceso estacionario. (de echo es exactamente constante para toda t) y si u_t es un Proceso de Promedios Móviles de orden q th).

Substituyendo $\mu=2$ y (4.2.14) en (4.2.13), las autocovarianzas se obtienen como

$$\begin{aligned}
\text{cov}(\nabla Y_t \nabla Y_{t-r}) &= (b+2c\mu)^2 (C_{t, t-r} + C_{t-1, t-r-1} - C_{t, t-r-1} - C_{t-1, t-r}) \\
&\quad - 2c^2 (C_{t, t-r}^2 + C_{t-1, t-r-1}^2 - C_{t, t-r-1}^2 - C_{t-1, t-r}^2)
\end{aligned} \quad (4.2.16)$$

Entonces, en particular, si $Y_t = X_{2t}$ se tiene por sustitución

$b=0$ y $c=1$ en (4.2.15) y (4.2.16), que

$$E(\nabla Y_t) = E(u_t^2 + 2u_t \sum_{j=1}^{t-1} u_j) \quad (4.2.17)$$

y

$$(4.2.18)$$

$$\text{cov}(\nabla Y_t \nabla Y_{t-r}) = 4\mu^2(C_{t,t-r} + C_{t-1,t-r-1} - C_{t,t-r-1} - C_{t-1,t-r}) \\ + 2(C_{t,t-r}^2 + C_{t-1,t-r-1}^2 - C_{t,t-r-1}^2 - C_{t-1,t-r}^2)$$

Finalmente, Granger y Newbold discuten el problema de pronosticar la variable transformada, considerando la situación en que estando en el tiempo n , se ha obtenido el pronóstico $f_{n,h}$ de X_{n+h} , el cual es óptimo en el sentido de la función de "pérdida cuadrática" que sufre, dado un conjunto de datos I_n (i.e. $f_{n,h} = E(X_{n+h}|I_n)$).

En el análisis que ellos hacen, suponen que el pronóstico óptimo $f_{n,h}$ tiene una distribución Normal. Suponen que X_{n+h} tiene media μ y varianza σ^2 y en todo lo expuesto a continuación, consideran que X_t es un proceso Gaussiano. Considerando una transformación instantánea de la forma $Y_{n+h} = T(Z_{n+h})$, donde

$$Z_{n+h} = \frac{X_{n+h} - \mu}{\sigma} \quad (4.2.19)$$

Granger y Newbold exponen cuatro tipos de pronóstico de Y_{n+h} :

i) El pronóstico óptimo según una función de pérdida cuadrática de Y_{n+h} , dado un conjunto de información I_n , denotándolo como $g^{(1)}_{n,h}$.

ii) El pronóstico simple (Naive forecast).

$$g^{(2)}_{n,h} = T((f_{n,h} - \mu)/\sigma)$$

iii) El pronóstico $g^{(3)}_{n,h}$ de Y_{n+h} el cual es óptimo en el sentido de pérdida cuadrática en la clase de pronósticos que son

lineales en X_{n-j} ; $j \geq 0$.

(iv) El pronóstico $g^{(1)}_{n,h}$ de Y_{n+h} el cual es similarmente óptimo en la clase de pronósticos que son lineales en Y_{n-j} , $j \geq 0$.

El Pronóstico $g^{(1)}_{n,h}$.

Sea $X_{n+h} = f_{n,h} + e_{n,h}$, en donde $e_{n,h}$ es el error del pronóstico h periodos hacia adelante, cuando X_{n+h} es pronosticado óptimamente tomando como base el conjunto de información I_n . Entonces, dada la suposición de Normalidad, se sigue que la distribución condicional de X_{n+h} , dada la información disponible, es Normal con media $f_{n,h}$ y varianza $S^2(h)$, donde $S^2(h) = \text{var}(e_{n,h})$. Escribase entonces

$$W_{n+h} = \frac{X_{n+h} - f_{n,h}}{S(h)} \quad \text{y} \quad Y_{n+h} = \sum_{i=0}^M \gamma_i H_i(W_{n+h})$$

Esta es una transformación instantánea puesto que $f_{n,h}$ y $S(h)$ son consideradas como constantes, en general las γ 's serán funciones de $f_{n,h}$ y $S(h)$. Entonces, como W_{n+h} esta distribuida condicionalmente como una variable Normal estándar, se sigue que el pronóstico óptimo con pérdida cuadrática, es decir, la esperanza condicional de $Y_{n,h}$ es, por (4.2.1)

$$g^{(1)}_{n,h} = \gamma_0 \quad (4.2.20)$$

Para evaluar el error cuadrático medio de la expresión anterior y subsecuentemente del pronóstico, dos cantidades son de interés, la Esperanza Condicional dada la información y la Esperanza Incondicional, que es simplemente la media de la anterior entre todas las posibilidades de realización del conjunto de datos. De (4.2.3) se sigue que el error cuadrático medio del pronóstico, dados los datos, es

$$V_c^{(1)}(h) = \sum_{j=1}^M \alpha_j^2 j! \quad (4.2.21)$$

El error cuadrático medio incondicional puede ser encontrado mediante

$$Y_{n+h} = \sum_{j=0}^M \alpha_j H_j(Z_{n+h})$$

donde Z_{n+h} esta dado por (4.2.1). En esta fórmula se puede escribir

$$g^{(1)}_{n,h} = \sum_{j=0}^M \alpha_j A^j H_j(P),$$

donde $A = \{(1-S^2(h)/\sigma^2)\}^{1/2}$, $P = (f_{n,h} - \mu)/\{\sigma^2 - S^2(h)\}^{1/2}$. Con un poco de álgebra y (A.5), el error cuadrático medio del pronóstico incondicional será

$$V_c^{(1)}(h) = \sum_{j=1}^M \alpha_j^2 j! (1-A^{2j}) \quad (4.2.22)$$

En discusiones subsiguientes de pronósticos sub-óptimos, el error cuadrático medio condicional puede ser utilizado para obtener una medición de la pérdida en eficiencia de una serie específica, mientras que el "incondicional" puede ser usado para estimar la pérdida.

Se define entonces, como una medida de que tan bien puede ser pronosticada una serie, dado un conjunto de datos, a

$$R^2_{n,y} = \frac{\text{var}(g^{(1)}_{n,h})}{\text{var}(Y_{n+h})}$$

donde $g^{(1)}_{n,h}$ es el pronóstico óptimo de Y_{n+h} y $\text{var}(Y_{n+h})$ es la varianza incondicional. Si el error de pronóstico es $\eta^{(1)}_{n,h}$ Entonces: $\text{Var}(Y_{n+h}) = \text{Var}(g^{(1)}_{n,h}) + \text{Var}(\eta^{(1)}_{n,h})$ y por lo tanto de (4.2.4)

$$R^2_{h,y} = \frac{\sum_{j=1}^M \alpha^2 j! A^{2j}}{\sum_{j=1}^M \alpha^2 j!}$$

Ya que $0 < A^2 < 1$, y $R^2_{h,x} = A^2$, se sigue $R^2_{h,y} < R^2_{h,x}$ para $\mu > 1$. Entonces la serie transformada Y es siempre "menos pronosticable" que la serie original X y en este sentido está "más cerca de ser ruido blanco".

Pronóstico $g^{(2)}_{n,h}$.

Considere el pronóstico simple de Y_{n+h} , obtenido de la sustitución de $f_{n,h}$ por X_{n+h} en la transformación $g^{(2)}_{n,h} = T\{(f_{n,h} - \mu)/\sigma\}$. El error cuadrático medio del pronóstico condicional es entonces la media condicional de

$$(Y_{n+h} - g^{(2)}_{n,h})^2 = \{(Y_{n+h} - g^{(1)}_{n,h}) + (g^{(1)}_{n,h} - g^{(2)}_{n,h})\}^2$$

así que

$$V^{(2)}_c(h) = \sum_{j=1}^M \gamma^2 j! + (g^{(1)}_{n,h} - g^{(2)}_{n,h})^2$$

escribáse entonces

$$g^{(2)}_{n,h} = \sum_{j=0}^M \alpha_j H_j \{ (f_{n,h} - \mu)/\sigma \}$$

Después de un tedioso ejercicio de álgebra, usando (A.5) se tiene

$$V^{(2)}(h) = \sum_{j=1}^M \alpha^2 j! (1 - A^{2j}) \tag{4.2.23}$$

$$+ \sum_{j=0}^M A^{2j} j! \times \left\{ \frac{1/2(M-j)}{\sum_{k=1}^{M-j} \alpha_{j+2k}} \frac{(j+2k)!}{k!} (-1/2B^2)^k \right\}^2$$

donde A se define como en (4.2.22) y $B^2 = 1 - A^2$.

El primer término de esta expresión es $V^{(1)}(h)$ como está dada en (4.2.22). El segundo término representa la cantidad promedio de pérdida en el error cuadrático si es que se ha utilizado un pronóstico simple.

El pronóstico $g^{(3)}_{n,h}$.

Supóngase ahora que el conjunto de datos dados consiste en observaciones pasadas y actuales de la serie X, esto es

$$I_n = (X_{n-j}, j \geq 0) \quad (4.2.24)$$

Se considera el pronóstico de la variable transformada como una combinación lineal de los miembros del conjunto de datos. La covarianza entre X_t y Y_{t-r} se sigue de (4.2.2).

Granger y Newbold, en este desarrollo, hacen uso de la teoría clásica de Kolmogorov (Whittie, 1963) llegando al resultado de que la pérdida cuadrática del pronóstico óptimo de Y_{n+h} , el cual es lineal en el conjunto de datos de I_n , es:

$$g^{(3)}_{n,h} = \alpha_0 + \alpha_1 (f_{n,h} - \mu / \sigma)$$

El error de pronóstico es entonces:

$$\begin{aligned} \eta^{(3)}_{n,h} &= Y_{n+h} - g^{(3)}_{n,h} \\ &= \sum_{j=0}^M \alpha_j H_j (X_{n+h} - \mu / \sigma) - (\alpha_0 + \alpha_1 (f_{n,h} - \mu / \sigma)) \end{aligned}$$

Se puede verificar que el error cuadrático del pronóstico esperado condicional es:

$$V^{(3)}_c(h) = V^{(1)}_c(h) + (g^{(1)}_{n,h} - g^{(3)}_{n,h})^2$$

El pronóstico $g^{(4)}_{n,h}$.

Nuevamente se considera el conjunto de información (4.2.24). Considérese también el problema de pronosticar a Y_{n+h} como una combinación lineal de $Y_{n-j}; j \geq 0$. La estructura de la media y las autocovarianzas de la serie transformada, están dadas por (4.2.1) y (4.2.3), y los resultados correspondientes para procesos integrados por (4.2.12) y (4.2.13). Entonces, Granger y

Newbold están de acuerdo que en principio el problema del pronóstico lineal óptimo está resuelto.

Estos dos autores comentan que en la práctica los resultados se facilitan cuando puede identificarse un modelo lineal estocástico con estructuras de autocovarianzas apropiadas. Para ilustrar esto, ellos consideran un caso estacionario suponiendo que un modelo apropiado es de la forma

$$\phi(B)(Y_t - \alpha_0) = \theta(B)e_t \quad (4.2.25)$$

donde

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

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

y e_t es un proceso de ruido blanco. La media α_0 es obtenida como se indica en (4.2.1). Entonces el pronóstico óptimo puede ser calculado a partir del modelo (4.2.25) de acuerdo con Box y Jenkins (1970). Más aún, la varianza del pronóstico lineal óptimo está dada por

$$v^{(h)}(h) = \sigma^2 e (1 + \psi^2 + \dots + \psi^{2h-1}) \quad (4.2.26)$$

donde $\sigma^2 e$ es la varianza de e_t y $\phi(B)(1 + \psi_1 B + \psi_2 B^2 + \dots) = \theta(B)$, pero no siempre puede obtenerse de manera directa un modelo lineal con la estructura de autocovarianzas discutida anteriormente.

4.3.- Método de Miller.

El propósito de Miller (1984) al escribir su artículo, fue el de hacer notar el problema de sesgo y dar "soluciones simples" la corrección por sesgo introducido por un conjunto de transformaciones de las que más comunmente se aplican a los datos, sin considerar específicamente el caso de series de

tiempo, sino el de regresión lineal.

Miller presenta un ejemplo que ilustra la necesidad de un estimador de la "respuesta media" en el espacio original de la variable, explicando la razón del sesgo y dando remedios apropiados al problema.

Un ejemplo ilustra la necesidad de estimar la media de la curva o variable transformada (mean response) en términos de la variable original, así como la magnitud del problema que esto conlleva. Miller supone en su ejemplo que una compañía desea utilizar los datos obtenidos de su investigación de mercado para modelar la relación entre el volumen de ventas de cierto tipo de establecimientos o negocios (Y) y un indicador del volumen de ventas (X), tal como el número de empleados del negocio. El modelo es utilizado para estimar volúmenes totales de venta para un sector del mercado que corresponden a valores representativos de X.

Para hacer esto, es necesario un modelo para la media de Y, más que para la mediana u otro estimador de tendencia central, porque esta puede ser multiplicada por el número de negocios en el sector seleccionado y así proporcionar un estimador válido de las ventas totales.

Se aplicó la transformación logarítmica a ambas variables y se ajustó un modelo por mínimos cuadrados resultando

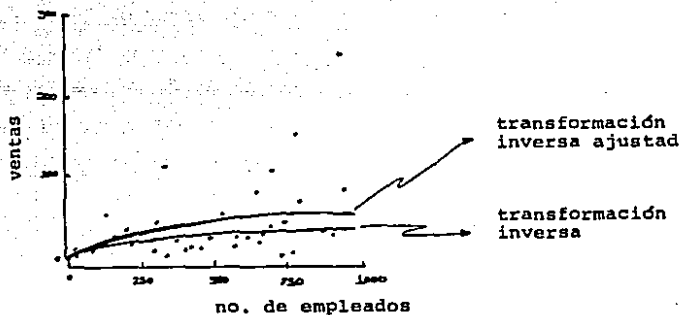
$$\ln(Y) = 7.3 + 0.5 \ln(X) \quad (4.3.1)$$

con un conjunto satisfactorio de residuales. La transformación inversa resulta ser

$$Y = 1500 X^{0.5} \quad (4.3.2)$$

La curva que resulta de la transformación inversa aparece graficada en la gráfica 8, Junto con los datos originales.

GRAFICA 8



Nótese que esta curva es representativa en el sentido de que aproximadamente la mitad de los puntos están abajo de ella y la otra mitad están arriba (como ocurre con la mediana).

Ahora, nótese que el estimador de la media de la curva transformada (mean response) parece estar altamente sesgada hacia abajo. Esto es en realidad el punto a tratar: la verdadera media de la curva es aproximadamente 50% más arriba que la curva graficada. Cuando se aplica un ajuste para reducir el sesgo, se obtiene una curva aproximadamente insesgada más arriba.

Es fácil demostrar que para cualquier transformación que es una función monótona, la transformación de la mediana de la

distribución original viene a ser la mediana de la distribución transformada. Es decir, $P(X < \text{Med}) = 0.5$ por lo tanto $P(X > \text{Med}) = 0.5$.

Ahora, ya que se supone que los residuales del modelo lineal tienen una distribución Normal, la línea recta ajustada representa la mediana de la curva resultante de la variable transformada, dado X , así como su media.

Consecuentemente, la curva que resulta de aplicar la transformación inversa al modelo lineal ajustado, con el término de error fijado en cero, representa el valor de la mediana de Y dado X , más que la media.

El sesgo producido al estimar la media de la curva obtenida aplicándole la transformación inversa, tiene las mismas bases teóricas que resultan de usar la mediana muestral de un conjunto de variables sesgadas univariadas como un estimador de la media poblacional.

Las soluciones que Miller propone, son para un conjunto de transformaciones sencillas que proporcionan una linealización satisfactoria a un gran número de aplicaciones. El conjunto está formado por la transformación logarítmica $\ln(Y)$ y algunas transformaciones potencia de fracciones de la forma $Y^{1/N}$ y $Y^{-1/N}$, donde N es un entero positivo. Se supone que en cada una de estas transformaciones, se manejan conjuntos de datos positivos.

Cuando la transformación logarítmica es apropiada, el modelo lineal ajustado será de la forma:

$$\ln(\bar{Y}) = \hat{\beta}_0 + \hat{\beta}_1 X \quad (4.3.3)$$

donde $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores de mínimos cuadrados de los coeficientes del modelo lineal. Destransformando se llega a :

$$Y = \hat{\beta}_0^* e^{\hat{\beta}_1 X} \quad (4.3.4)$$

donde $\hat{\beta}_0^* = e^{\hat{\beta}_0}$. Para demostrar el sesgo de la transformación inversa de este modelo, primero se considera el verdadero modelo lineal.

$$\ln(Y) = \beta_0 + \beta_1 X + \epsilon \quad (4.3.5)$$

donde, aplicando la transformación inversa, se obtiene el siguiente modelo multiplicativo.

$$Y = \beta_0^* e^{\beta_1 X} \cdot \epsilon^* \quad (4.3.6)$$

donde $\beta_0^* = e^{\beta_0}$ y $\epsilon^* = e^{\epsilon}$ tiene una distribución LogNormal.

La media de Y es

$$\begin{aligned} E(Y) &= \beta_0^* e^{\beta_1 X} E(\epsilon^*) \\ &= \beta_0^* e^{\beta_1 X} \cdot e^{1/2 \sigma^2} \end{aligned} \quad (4.3.7)$$

Ahora, la mediana de la transformación inversa (median response) es dada por

$$\begin{aligned} MD(Y) &= \beta_0^* e^{\beta_1 X} MD(\epsilon^*) \\ &= \beta_0^* e^{\beta_1 X} \end{aligned} \quad (4.3.8)$$

puesto que $MD(\epsilon^*)$ es igual a 1 en este caso. (Hald, 1952).

Ahora, $\ln(Y)$ tiene una distribución Normal con media $\beta_0 + \beta_1 X$ y varianza

$$\sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

por lo tanto \hat{Y} tiene una Distribución Lognormal con media

$$E(\hat{Y}) = \left[\exp \left(\beta_0 + \beta_1 x + \frac{1}{2} \sigma^2 \right) \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \right]$$

Conforme crezca la muestra, de tamaño n , se acercará más a esta aproximación (4.3.8); por lo tanto, el estimador de la destransformación (transformación inversa) (4.3.4) proporciona un estimador consistente de la mediana de la transformación inversa, pero sistemáticamente subestima la media de la transformación (4.3.7). Nótese que el factor de sesgo, $e^{1/2\sigma^2}$, es multiplicativo y crece exponencialmente con σ^2 . Un remedio simple, es aplicar una estimación de este factor al estimador de la transformación inversa (4.3.4), O sea:

$$\hat{E}(Y) = \hat{\beta}_0^* e^{\hat{\beta}_1 x} \cdot e^{1/2 \hat{\sigma}^2}$$

donde $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\sigma}^2$ son los estimadores de mínimos cuadrados para el modelo (4.3.5). Este modelo todavía se mantiene un poco sesgado puesto que $\hat{\beta}_0^*$ y $\exp(\hat{\beta}_1)$ son estimadores sesgados para β_0^* y $\exp(\beta_1)$ respectivamente, sin embargo, al aplicar el factor de ajuste se elimina la mayor porción de sesgo.

Para las transformaciones potencia fraccionadas $Y^{1/N}$, la transformación inversa del modelo antes expuesto es:

$$Y = (\beta_0 + \beta_1 X + \epsilon)^N \quad (4.3.9)$$

si ϵ tiene una Distribución Normal $(0, \sigma^2)$, entonces $E(Y)$ es el n -ésimo momento de una variable aleatoria Normal con media $\beta_0 + \beta_1 X$ y varianza σ^2 . Los momentos de la distribución Normal ya son conocidos y proporcionan el valor esperado de Y dado X , para cualquier valor de N . Nótese que : a) el que Y sea positivo asegura que (4.3.9) es monótona y b) los momentos son aproximados ya que ϵ sólo puede aproximarse a una Normal.

Para la transformación raíz cuadrada, $N=2$, se tiene

$$E(Y) = (\beta_0 + \beta_1 X)^2 + \sigma^2 \quad (4.3.10)$$

por lo tanto, el estimador de "bajo-sesgo" se obtiene de sustituir $\hat{\beta}_0, \hat{\beta}_1$ y $\hat{\sigma}^2$ del modelo lineal ajustado:

$$\hat{E}(Y) = (\hat{\beta}_0 + \hat{\beta}_1 X)^2 + \hat{\sigma}^2 \quad (4.3.11)$$

Nótese que el primer término es el modelo ajustado destransformado y el segundo término es el factor de ajuste para reducir el sesgo.

Si la transformación inversa $1/Y$ es la apropiada, el modelo sera

$$1/Y = \beta_0 + \beta_1 X + \epsilon \quad (4.3.12)$$

note que si Y es siempre positiva, se asegura que $E(Y)$ exista. Para encontrar $E(Y)$, Miller recurrió a la identidad.

$$1/W = 1/E(W) [1 + [W-E(W)] / E(W)]^{-1} \quad (4.3.13)$$

expresando el término de los corchetes como una serie geométrica y cortando después de los tres primeros términos, se obtiene la

siguiente aproximación:

$$E(Y) = \frac{1}{\beta_0 + \beta_1 X} \left[1 + \frac{\sigma^2}{(\beta_0 + \beta_1 X)^2} \right] \quad (4.3.14)$$

Un estimador de "bajo sesgo" se obtiene sustituyendo $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\sigma}^2$. El primer factor será entonces el modelo destransformado, mientras que el factor entre paréntesis es el de ajuste.

Cuando la transformación inversa apropiada es $Y^{-1/N}$ la media de Y puede encontrarse combinando los métodos previamente expuestos. Para el inverso de la raíz cuadrada se tiene

$$E(Y) = \frac{1}{(\hat{\beta}_0 + \hat{\beta}_1 X)^2 + \hat{\sigma}^2} \times \left[1 + \frac{2\sigma^4 + 4(\hat{\beta}_0 + \hat{\beta}_1 X)^2 \sigma^2}{[(\hat{\beta}_0 + \hat{\beta}_1 X)^2 + \sigma^2]^2} \right] \quad (4.3.15)$$

El estimador sugerido por Miller para reducir el sesgo introducido por la transformación inversa consiste en una simple inclusión de un término adicional o factor para compensar los valores esperados. Estos todavía tienen un pequeño grado de sesgo, porque consisten en funciones no-lineales de estimadores de mínimos cuadrados.

4.4 .- Método de Taylor.

Taylor (1986) proporciona un método aproximado para estimar la media de la variable dependiente después de que un modelo lineal ha sido ajustado a la variable transformada, estudiando también sus propiedades y mostrando que hay un cierto "costo", debido a la estimación de λ , comparado con la situación que presenta una λ conocida a priori, aunque en general este costo es muy pequeño.

Considérese la función

$$h(Y, \lambda) = A\beta + \sigma_e \quad (4.4.1)$$

donde σ_e se supone que tiene una función de densidad Normal estándar, A es una matriz dada, β es un vector de parámetros y

$$\begin{aligned} h(Y, \lambda) &= (Y\lambda - 1)/\lambda & \lambda \neq 0 \\ &= \text{Log}(Y) & \lambda = 0 \end{aligned}$$

para toda $Y > 0$.

La ecuación (4.4.1) presupone que sólo una transformación puede alcanzar linealidad, varianza constante y errores con distribución Normal. Taylor hace notar que en la práctica esto es muy raro, sin embargo la exactitud de (4.4.1) no le parece muy importante.

Taylor utiliza el método de Máxima Verosimilitud (Box y Cox, 1964) para estimar los parámetros del modelo y hace mención a algunos comentarios de estadísticos sobre la media de β . En particular, hace referencia a Bickel y Doksum (1981), quienes muestran que la varianza de $\hat{\beta}$ se ve afectada cuando λ es estimada a partir de los datos, comparada con la λ dada a priori, y a los de Carroll y Ruppert (1981), los cuales se inclinan por los parámetros definidos independientes de la escala, sugiriendo el uso de la transformación inversa y haciendo inferencia a la escala original de las variables. Estos últimos estudian las propiedades de la mediana condicional de la distribución de Y dado A , esto es, la mediana retransformada después de ajustar un modelo lineal a las observaciones transformadas.

Supóngase que Y_1, \dots, Y_n han sido generadas de acuerdo a

(4.4.1) con f una función de densidad general que satisface las condiciones $E_f(e)=0$ y $E_f(e^2)=1$. Entonces la media condicional de un valor futuro de Y , cuando $A=A_0$, está dada por

$$E(Y|A_0) = \int (1 + \lambda A_0 \beta + \lambda \sigma e)^{1/\lambda} dF(e).$$

Un estimador de lo anterior es

$$\int (1 + \hat{\lambda} A_0 \hat{\beta} + \hat{\lambda} \hat{\sigma} e)^{1/\hat{\lambda}} dF(e). \quad (4.4.2)$$

Para evaluar esto se requerirá estimar los parámetros $\xi(\lambda, \beta, \sigma)$, estimando la función de densidad f para después evaluar numéricamente la integral. Claramente, no es un método satisfactorio desde el punto de vista práctico.

Para obtener una aproximación de esta cantidad, Taylor utilizó el método de " σ -pequeña" dado por Draper y Cox (1969).

Sea $\theta(A_0) = \lambda \sigma / (1 + \lambda A_0 \beta)$, $\lambda \neq 0$. Entonces la condición es que $|\theta(A_0)| \ll 1$ para toda A_0 de interés. Es necesario asegurarse que las observaciones Y sean todas positivas, para lo cual

$$\begin{aligned} E(Y|A_0) &= (1 + \lambda A_0 \beta)^{1/\lambda} \int (1 + \theta(A_0)e)^{1/\lambda} dF(e) \\ &= (1 + \lambda A_0 \beta)^{1/\lambda} \int \end{aligned}$$

Haciendo la expansión de la integral, se tiene

$$1 = \int \left[1 + \sum_{j=1}^{\infty} (\theta(A_0)e)^j \frac{j-1}{j!} \left(\frac{1}{\lambda} - j \right) \right] dF(e)$$

Si la función de densidad f es simétrica y el orden de la integración y la suma pueden ser intercambiadas, entonces al suprimir términos de grado mayor que $(\theta(A_0))^2$, se tiene que

$$E(Y|A_0) \approx Y_a = (1 + \lambda A_0 \beta)^{1/\lambda} \left\{ 1 + \frac{\sigma^2(1-\lambda)}{2(1 + \lambda A_0 \beta)^2} \right\}$$

Entonces un estimador a usar es

$$\hat{Y}_a = (1 + \hat{\lambda} A_0 \hat{\beta})^{1/\hat{\lambda}} [1 + \hat{\sigma}^2 (1 - \hat{\lambda}) / 2 (1 + \hat{\lambda} A_0 \hat{\beta})^2]$$

Nótese lo siguiente: a) \hat{Y}_a es igualmente variante a los cambios en la escala de Y . b) Y_a puede considerarse como la mediana multiplicada por un factor de corrección, donde el factor de corrección es mayor que 1 si $\lambda < 1$ y menor que uno si $\lambda > 1$.

Taylor hace notar dos puntos importantes:

1) La cercanía de Y_a de $E(Y|A_0)$, para λ cerca de cero, $Y_a \approx \exp(A_0 \beta) (1 + \sigma^2 / 2)$; ahora, en caso de que f sea Normal, entonces $E(Y|A_0) \approx \exp(A_0 \beta) (1 + \sigma^2 / 2)$. Por lo tanto, la aproximación no es exacta en la situación donde λ está cerca de cero y σ es grande. Taylor define vagamente a esta región del espacio del parámetro como S' y su complemento como S ; esto es, $S' = \{\xi : \lambda \text{ cerca de cero y } \sigma \text{ grande}\}$, donde $\xi = (\lambda, \beta, \sigma)$.

Desafortunadamente, S' es una región importante, pero como se puede ver, es también una región donde es muy difícil obtener un estimador razonable. De hecho Taylor demuestra en su artículo que S' parece ser la única región donde la aproximación no es válida.

2) Propiedades de \hat{Y}_a considerado como un estimador de Y_a . Una expansión en serie de Taylor da:

$$\hat{Y}_a = Y_a + P(\hat{\lambda} - \lambda) + Q(\hat{\beta} - \beta) + R(\hat{\sigma} - \sigma) + o(\hat{\lambda} - \lambda, \hat{\beta} - \beta, \hat{\sigma} - \sigma).$$

Por lo que, si los parámetros ξ son estimados consistentemente, entonces $\hat{Y}_a \xrightarrow{p} Y_a$ cuando $n \rightarrow \infty$.

Similarmente, la varianza límite de $\eta^{1/2}(\hat{Y}_a - Y_a)$ puede ser obtenida a partir de la covarianza asintótica de ξ .

4.5.- Método de Guerrero.

Como se ha comentado a través de esta tesis, en la práctica la transformación que se emplea con frecuencia es la transformación potencia en sus dos versiones (2.1.2) y (2.2.1). Para esta transformación es posible realizar el cálculo de un factor que permita corregir aproximadamente el sesgo que introduce la aplicación de la transformación inversa (para regresar a la escala original). Dicho factor se presenta a continuación.

Para empezar, nótese que para un valor λ dado, $Z_{t+h}^{(\lambda)}$ puede escribirse mediante la siguiente expansión en serie de Taylor alrededor de $Z_0 > 0$

$$Z_{t+h}^{(\lambda)} = \sum_{k=0}^{\infty} \frac{D^k Z_0^{(\lambda)}}{k!} (Z_{t+h} - Z_0)^k \quad (4.5.1)$$

con

$$D^k Z_0^{(\lambda)} = \frac{d^k Z_t^{(\lambda)}}{d Z_t^k} \Big|_{Z_t = Z_0} = \begin{cases} Z_0^{(\lambda)} & k=0 \\ Z_0^{\lambda-1} & k=1 \\ (\lambda-1)(\lambda-2)\dots(\lambda-k+1)Z_0^{\lambda-k} & k > 2 \end{cases}$$

de tal forma que, si se elige a Z_0 como

$$Z_0 = E_t(Z_{t+h}) = E(Z_{t+h} | Z_t, Z_{t-1}, \dots)$$

en donde E_t denota a la esperanza condicional, dada toda la información hasta el momento t , entonces se tendrá

$$Z_{t+h}^{(\lambda)} = [E_t(Z_{t+h})]^{(\lambda)} + [E_t(Z_{t+h})]^{\lambda-1} [Z_{t+h} - E(Z_{t+h})]$$

$$+ \frac{\lambda-1}{2} [E_t(Z_{t+h})]^{\lambda-2} [Z_{t+h}-E_t(Z_{t+h})]^2 + \dots$$

así que

$$E_t^{(\lambda)}(Z_{t+h}) = [E_t(Z_{t+h})]^{(\lambda)} + \frac{\lambda-1}{2} [E_t(Z_{t+h})]^{\lambda-2} \text{Var}_t(Z_{t+h}) \quad (4.5.2)$$

de manera similar, ya que

$$\begin{aligned} (Z_{t+h})^{(\lambda)} = & \left[\frac{[E_t(Z_{t+h})]^{(\lambda)}}{t} \right]^2 + 2 \left[\frac{[E_t(Z_{t+h})]^{(\lambda)}}{t} \right] \left[\frac{[E_t(Z_{t+h})]^{\lambda-1} [Z_{t+h}-E_t(Z_{t+h})]}{t} \right] \\ & + \left\{ (\lambda-1) \left[\frac{[E_t(Z_{t+h})]^{(\lambda)}}{t} \right] \left[\frac{[E_t(Z_{t+h})]^{\lambda-2}}{t} \right] \right. \\ & \left. + \left[\frac{[E_t(Z_{t+h})]^{2\lambda-2}}{t} \right] [Z_{t+h}-E_t(Z_{t+h})]^2 + \dots \right\} \quad (4.5.3) \end{aligned}$$

entonces

$$\begin{aligned} E_t^{(\lambda)}[(Z_{t+h})^2] = & \left[\frac{[E_t(Z_{t+h})]^{(\lambda)}}{t} \right]^2 + \left\{ (\lambda-1) \left[\frac{[E_t(Z_{t+h})]^{(\lambda)}}{t} \right] \left[\frac{[E_t(Z_{t+h})]^{\lambda-2}}{t} \right] \right. \\ & \left. + \left[\frac{[E_t(Z_{t+h})]^{2\lambda-2}}{t} \right] \right\} \text{Var}_t(Z_{t+h}) \end{aligned}$$

por lo tanto, de (4.5.2) y (4.5.3) se obtiene

$$\text{var}_t^{(\lambda)}(Z_{t+h}) = \left[\frac{[E_t(Z_{t+h})]^{2\lambda-2}}{t} \text{var}_t(Z_{t+h}) \right] \left\{ 1 - \frac{(\lambda-1)^2}{4} \left[\frac{[E_t(Z_{t+h})]^{\lambda-2}}{t} \text{var}_t(Z_{t+h}) \right] \right\}$$

que puede reducirse, mediante una segunda aproximación, a

$$\text{var}_t^{(\lambda)}(Z_{t+h}) = \left[\frac{[E_t(Z_{t+h})]^{2\lambda-2}}{t} \text{var}_t(Z_{t+h}) \right] \quad (4.5.4)$$

Esta expresión implica en particular que

$$\left[\frac{[E_t(Z_{t+h})]^{\lambda-2}}{t} \text{var}_t(Z_{t+h}) \right] = \left[\frac{[E_t(Z_{t+h})]^{-\lambda}}{t} \text{var}_t^{(\lambda)}(Z_{t+h}) \right]$$

lo cual aplicado en (4.5.2) produce

$$E_t^{(\lambda)}(Z_{t+h}) = \left[\frac{[E_t(Z_{t+h})]^{(\lambda)}}{t} \right] + \frac{(\lambda-1)}{2} \left[\frac{[E_t(Z_{t+h})]^{-\lambda}}{t} \text{var}_t^{(\lambda)}(Z_{t+h}) \right] \quad (4.5.5)$$

La relación (4.5.5) es fundamental para obtener factores de corrección por sesgo ya que, en el caso de $\lambda=0$, se tendrá

$$E_t [\log(Z_{t+h})] = \log[E_t(Z_{t+h})] - \frac{\text{Var}_t[\log(Z_{t+h})]}{2}$$

de donde

$$E_t(Z_{t+h}) = \exp[E_t(\log(Z_{t+h}))] \exp\{\frac{\text{Var}_t[\log(Z_{t+h})]}{2}\} \quad (4.5.6)$$

mientras que, para $\lambda \neq 0$, (4.5.5) implica

$$[E_t(Z_{t+h})]^{2\lambda} - [1+\lambda E_t(Z_{t+h})] [E_t(Z_{t+h})]^\lambda + \frac{\lambda(\lambda-1)}{2} \text{Var}_t(Z_{t+h})^\lambda = 0$$

la cual es una ecuación de segundo grado, cuya solución viene dada por

$$[E_t(Z_{t+h})]^\lambda = [1+\lambda E_t(Z_{t+h})] \pm \sqrt{[1+\lambda E_t(Z_{t+h})]^2 - 2\lambda(\lambda-1) \text{Var}_t(Z_{t+h})^\lambda} \quad (4.5.7)$$

es decir, si $\lambda \neq 0$

$$E_t(Z_{t+h}) = [1+\lambda E_t(Z_{t+h})]^{1/\lambda} \{1/2 \pm \sqrt{1-2\lambda(\lambda-1)[1+\lambda E_t(Z_{t+h})]^{-2} \text{Var}_t(Z_{t+h})^\lambda / 2}\}^{1/\lambda}$$

La elección del signo del radical en (4.5.7) debe efectuarse de tal manera que se cumpla con la siguiente propiedad

$$E_t(Z_{t+h}) \geq [1+\lambda E_t(Z_{t+h})]^{1/\lambda} \quad \text{según si } \lambda \geq 1 \quad (4.5.8)$$

la cual podría demostrarse a partir del hecho de que la transformación de Box y Cox es una función monótona del parámetro λ . Así pues, ya que (4.5.7) puede expresarse como

$$E_t(Z_{t+h}) = [1+\lambda E_t(Z_{t+h})]^{1/\lambda} c_{t,\lambda}(h) \quad (4.5.9)$$

entonces debe satisfacer que $c_{t,\lambda}(h) = 1$, de acuerdo con $\lambda \geq 1$, para toda h . En particular, si $\lambda < 1$, el signo a elegir deberá ser el positivo.

En consecuencia, de (4.5.6) y (4.5.7) se sigue que, si $T(Z_t) = Z_t^{(\lambda)}$, entonces

$$E_t(Z_{t+h}) = T^{-1}\{E_t[T(Z_{t+h})]\} \times c_{t,\lambda}(h) \quad (4.5.10)$$

con

$$c_{t,\lambda}(h) = \begin{cases} (1/2) \sqrt{1 - 2\lambda(\lambda-1) \frac{[E_t(Z_{t+h})]^{(\lambda)-2}}{t} \frac{\text{Var}(Z_{t+h})/2}{t}}^{1/\lambda} & \text{si } \lambda < 1, \lambda \neq 0 \\ \exp(\text{Var}[\log(Z_{t+h})]/2) & \text{si } \lambda = 0 \end{cases} \quad (4.5.11)$$

Por otro lado, si la transformación utilizada fue la versión original (2.1.2) y no la versión de Box y Cox, el desarrollo anterior también sería válido, pero ahora el factor (4.5.11) se convertirá en

$$c_{t,\lambda}(h) = \begin{cases} (1/2) \sqrt{1 - \frac{2(\lambda-1)}{\lambda} \frac{[E_t(Z_{t+h})]^{(\lambda)-2}}{t} \frac{\text{Var}(Z_{t+h})/2}{t}}^{1/\lambda} & \text{si } \lambda < 1, \lambda \neq 0 \\ \exp(\text{Var}[\log(Z_{t+h})]/2) & \text{si } \lambda = 0 \end{cases} \quad (4.5.12)$$

Es necesario también, obtener intervalos de confianza para los pronósticos de Z_{t+h} ; para esto se sugiere simplemente traer a la escala original el intervalo que se haya construido (para $T(Z_{t+h})$) en la escala transformada, el cual resultará

centrado en $T^{-1}\{\hat{T}(Z_t)(h)\}$ y vendrá dado por

$$T^{-1}\left\{E\left[\hat{T}(Z_{t+h})\right] \pm z_{\alpha/2} \sqrt{\text{Var}\left\{\frac{E[\hat{T}(Z_{t+h})]}{t}\right\}}\right\} \quad (4.5.13)$$

dicho intervalo puede entonces corregirse por sesgo al multiplicarse por el factor $c_{t,\lambda}(h)$ dado por (4.5.11) o, en su caso, por (4.5.12).

Guerrero (1983) señala que: "conviene advertir que en el desarrollo anterior se supuso que λ era un valor fijo y conocido; si, por el contrario, λ se estima con base en los mismos datos de la serie de tiempo en estudio, la varianza del estimador $\hat{\lambda}$ que se utilice, afectará a los resultados, principalmente al intervalo de confianza. Esto no ocurre si el valor λ es dado a priori (antes del análisis de la serie) o bien si dicha λ surge de consideraciones ajenas a un procedimiento de estimación formal; en este sentido sería preferible emplear el método de selección de la potencia que establezca la varianza recién presentado y no emplear el método de máxima verosimilitud."

4.6 .- Comparaciones Teóricas.

Uno de los puntos fuertemente discutidos por Chatfield y Prothero en el análisis de su serie, fue el del sesgo introducido por la transformación logarítmica, que como se vió en el capítulo anterior, no es la más adecuada para la serie en estudio y por consecuencia el modelo A da pronósticos muy pobres. Sin embargo, para obtener pronósticos óptimos es necesario corregir el sesgo introducido por la transformación aplicada a la serie.

De los métodos presentados en este capítulo se puede observar que el de Neyman y Scott (1960) y el de Granger y

Newbold (1976), son difíciles de usar por su pesado desarrollo matemático. Granger y Newbold (1976) comentan que aunque "el método permite considerar una amplia gama de transformaciones y modelos, en la mayoría de los casos la cantidad de cálculos algebraicos para la obtención del pronóstico óptimo es muy grande"; sin embargo Miller (1984) opina que aunque son métodos muy complejos y tienen muy poca oportunidad de ser utilizados periódicamente en la práctica estadística, sus resultados son substanciales y exactos.

Los métodos de Guerrero (1983) y Miller (1984) reducen el sesgo introducido por la transformación inversa estimada con la "simple" inclusión de un término adicional o factor para compensar la pérdida del pronóstico. Ambos métodos son similares y pueden ser calculados con los resultados estándares de cualquier paquete estadístico. Su implementación no es difícil de llevarse a cabo, sin embargo el método de Miller sólo puede ser aplicado cuando $\lambda=1/N$ y N es un entero. En el apéndice A se presenta un programa escrito en Pascal que calcula el factor de corrección aplicando el método de Guerrero para la transformación potencia.

De acuerdo al autor, el método de Taylor "es un método que no requiere de integración numérica o estimación de densidades excepto para λ cerca de cero y σ grande". Este método proporciona aproximaciones muy exactas.

Por razones de cálculo, solo se aplicaron los tres últimos métodos a la serie de Chatfield y Prothero obteniendo los resultados mostrados en la tabla 11. En este caso, el método de

TABLA 11

PRONOSTICO EN LA ESCALA ORIGINAL

Serie: $Z_t = C \& P_t$					Transformación $T(Z_t) = \text{Log}(Z_t)$				
Obs.	Pronóstico	Guerrero	Miller	Taylor					
260	305	311	311	311					
304	482	495	495	495					
390	673	699	699	699					
614	990	1038	1038	1038					
783	1297	1374	1374	1374					
872	1387	1483	1483	1483					
	M_1	-362.8	-362.8	-362.8					
	M'_1	-62.71%	-62.71%	-62.71%					

Serie: $Z_t = C \& P_t$					Transformación $T(Z_t) = Z^{0.24}_t$				
Obs.	Pronóstico	Guerrero	Miller	Taylor					
260	284	287	287	284					
304	410	414	414	410					
390	529	536	536	529					
614	743	754	754	744					
783	935	950	950	936					
872	988	1006	1006	989					
	M_1	-120.6	-120.6	-111.5					
	M'_1	-23.91%	-23.91%	-22.31%					

Serie: $Z_t = C \& P_t$					Transformación $T(Z_t) = Z^{0.25}_t$				
Obs.	Pronóstico	Guerrero	Miller	Taylor					
260	286	288	286	286					
304	409	413	409	409					
390	511	518	511	511					
614	761	771	762	762					
783	966	980	967	967					
872	1091	1109	1092	1092					
	M_1	-142.6	-134	-134					
	M'_1	-26.22%	-19.50%	-19.50%					

Serie: $Z_t = C \& P_t$					Transformación $T(Z_t) = Z^{0.254}_t$				
Obs.	Pronóstico	Guerrero	Miller	Taylor					
260	284	286	286	284					
304	407	411	407	407					
390	523	530	523	523					
614	735	745	736	736					
783	924	938	925	925					
872	976	992	977	977					
	M_1	-113.1	-113.1	-104.8					
	M'_1	-22.66%	-22.66%	-21.21%					

Miller sólo se puede aplicar a la transformación logarítmica y $\lambda=1/4$, aunque se tiene conocimiento de su generalización, esta demostración no ha sido publicada a la fecha.

Como se puede ver en la tabla anterior, los pronósticos corregidos con el método de Taylor y Miller son iguales, inclusive Taylor (1986) hace notar que el factor de corrección es igual al de Miller cuando λ es positiva o de la forma $1/N$ donde N es un entero positivo, pero difieren cuando λ es negativa y menor que -1 o de la forma $-1/N$, donde N es un entero positivo.

En la tabla 11 se puede ver que la media de los errores de pronóstico son negativas, esto quiere decir que se esta sobreestimando los valores de la serie. Sin embargo para $\lambda=0.254$, tanto el método de Taylor como el de Guerrero sobreestiman en casi la misma proporción, 22.66% con el de Guerrero y 21.21% con el de Taylor.

Los tres métodos dan los mismos resultados para $\lambda=0$. En la tabla 12 se muestra el factor de corrección para los tres métodos cuando la λ es 1, $1/2$, 0, $-1/2$ y -1 .

TABLA 12

	GUERRERO (1983)	MILLER (1984)	TAYLOR (1986)
1	$1 + \mu$	$1 + \mu$	$1 + \mu$
1/2	$\left(1 + \frac{\mu}{2}\right)^2 \left[\frac{1}{2} + \sqrt{1 + \frac{\sigma^2}{2\left(1 + \frac{\mu}{2}\right)^2}} / 2 \right]^2$	$\left(1 + \frac{\mu}{2}\right)^2 \left[1 + \frac{\sigma^2}{4\left(1 + \frac{\mu}{2}\right)^2} \right]$	$\left(1 + \frac{\mu}{2}\right)^2 \left[1 + \frac{\sigma^2}{4\left(1 + \frac{\mu}{2}\right)^2} \right]$
0	$e^{\mu + \sigma^2/2}$	$e^{\mu + \sigma^2/2}$	$e^{\mu + \sigma^2/2}$
-1/2	$\left(1 - \frac{\mu}{2}\right)^{-2} \left[\frac{1}{2} + \sqrt{1 - \frac{3\sigma^2}{2\left(1 - \frac{\mu}{2}\right)^2}} / 2 \right]^{-2}$	$\left[\left(1 - \frac{\mu}{2}\right)^2 + \frac{\sigma^2}{4} \right]^{-1} \left[1 + \frac{14\sigma^4 + \left(1 - \frac{\mu}{2}\right)^2 \sigma^2}{\left[\left(1 - \frac{\mu}{2}\right)^2 + \frac{\sigma^2}{4}\right]^2} \right]$	$\left(1 - \frac{\mu}{2}\right)^{-2} \left[1 + \frac{3\sigma^2}{4\left(1 - \frac{\mu}{2}\right)^2} \right]$
-1	$\left(1 - \mu\right)^{-1} \left[\frac{1}{2} + \sqrt{1 - \frac{4\sigma^2}{\left(1 - \mu\right)^2}} / 2 \right]^{-1}$	$\left(1 - \mu\right)^{-1} \left[1 + \frac{\sigma^2}{\left(1 - \mu\right)^2} \right]$	$\left(1 - \mu\right)^{-1} \left[1 + \frac{\sigma^2}{\left(1 - \mu\right)^2} \right]$

CAPITULO 5

Algunas aplicaciones empíricas

En este capítulo presentaré los resultados obtenidos al aplicar los métodos de Guerrero introducidos en esta tesis, a las series de :

5.1) Ventas de la compañía X que Chatfield y Prothero utilizaron en su artículo (1973) previamente discutido. Los datos fueron tomados de dicho artículo y comprenden 6 años con 5 meses, empezando en 1965 y terminando en mayo de 1971, es decir, un total de 77 observaciones.

5.2) Índice Nacional de Precios al Consumidor (Base 1978=100), la cual resulta de la recopilación mensual de 90,000 cotizaciones directas en 35 ciudades, sobre los precios de aproximadamente 1,200 artículos y servicios específicos. Los promedios de dichas cotizaciones dan lugar a los índices de los 302 conceptos genéricos sobre bienes y servicios, que forman la canasta del índice general en cada una de las ciudades y a nivel Nacional en la República Mexicana. Los datos fueron tomados de la publicación anual de Índices de Precios, Salarios y Encuesta de Inversión Empresarial que hace el Banco de México. Estos comprenden de enero 1982 a diciembre de 1987, haciendo un total de 72 observaciones.

El análisis de estas series fué hecho en una Computadora Personal con el paquete STATISTIX II de NH Analytical software y que para sus cálculos utiliza los métodos descritos en el libro de Box y Jenkins (1970). En particular con el caso Chatfield y Prothero se compararán los resultados con los que han sido obtenidos por otros autores.

5.1) Caso Chatfield y Prothero.

Los datos considerados son:

TABLA 13
VENTAS DE LA COMPANIA X.

	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
1965	154	96	73	49	36	59	95	169	210	278	298	245
1966	200	118	90	79	78	91	167	169	289	347	375	203
1967	223	104	107	85	75	99	135	211	335	460	488	326
1968	346	261	224	141	148	145	223	272	445	560	612	467
1969	518	404	300	210	196	186	247	343	464	680	711	610
1970	613	392	273	322	189	257	324	404	677	858	895	664
1971	628	308	324	248	272							

Como se muestra en la gráfica 1, la serie original presenta una tendencia y un patrón estacional de un ciclo de 12 observaciones. El tamaño del efecto estacional se ve incrementado con el nivel de las ventas por lo que aplicaré una transformación estabilizadora, utilizando el método de Guerrero.

Para llevar a cabo los cálculos del método de Guerrero, se utilizó en primera instancia una hoja de cálculo en LOTUS 1-2-3, el cual no encontré muy conveniente ya que para utilizarlo sistemáticamente, hay que ajustar manualmente los cálculos de acuerdo al número de datos que la serie en estudio contenga. Posteriormente programé el método en lenguaje Pascal para evitar el ajuste manual, el programa se presenta en el Apéndice de esta tesis.

VENTAS DE LA COMPANIA X

Grafica 1

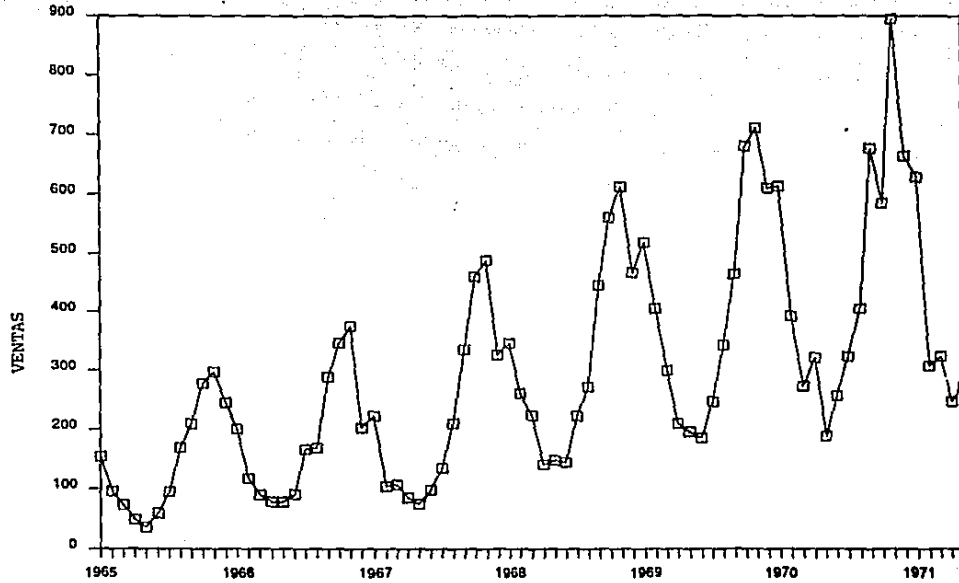


TABLA 14
DETERMINACION DE λ PARA LA SERIE DE VENTAS

Grupo	Potencia						
	0	0.25	0.253	0.254	0.255	0.34	1
1	.63130	2.19756	2.23070	2.24186	2.25308	3.44315	92.69582
2	.56520	2.08603	2.11891	2.12999	2.14112	3.33517	104.14486
3	.67120	2.58696	2.62918	2.64341	2.65771	4.20461	148.11257
4	.51333	2.17168	2.20960	2.22238	2.23524	3.65006	164.43641
5	.46823	2.10146	2.13967	2.15256	2.16552	3.60799	189.98331
6	.49503	2.32786	2.37151	2.38624	2.40106	4.06439	242.06874
$M(\lambda)$.55760	2.24526	2.28326	2.29607	2.30896	3.71756	156.90695
$DE(\lambda)$.08035	0.18831	0.19145	0.19252	0.19360	0.34514	55.47525
$CV(\lambda)$.14410	.08387	.0838497	.0838481	.0838490	.09284	0.35356

Obsérvese que en la tabla anterior el mínimo coeficiente de variación (dentro del grupo de λ elegido) se logra con $\lambda=0.254$ por lo que la transformación a considerar será : $T(Z_t)=Z_t^{0.254}$, siguiendo el proceso de Box y Jenkins (1970) de identificación de modelos, se analizarán las funciones de autocorrelación de las series $\nabla^d \nabla_{0.12} Z_t$ para $d=0,1,2$ y $0=0,1,2$ y poder determinar el grado de diferenciación para estacionarizar la serie. Las correspondientes desviaciones estándar son :

$$S(0,0) = 0.72047$$

$$S(0,1) = 0.21672$$

$$S(0,2) = 0.34521$$

$$S(1,0) = 0.33119$$

$$S(1,1) = 0.23960$$

$$S(1,2) = 0.38924$$

$$S(2,0) = 0.38302$$

$$S(2,1) = 0.42359$$

$$S(2,2) = 0.69557$$

Como se ha venido comentando a través de este trabajo, varios fueron las personas que estudiaron esta serie, por lo que en la tabla 15 se muestran las desviaciones estándar de la serie para $\lambda=0, 0.24, 0.25, 0.254$. Los resultados de la estimación de los modelos identificados para esta serie se muestran en las tablas 16,17,18,19.

TABLA 15
DESVIACION ESTANDAR DE LA SERIE T(Zt)

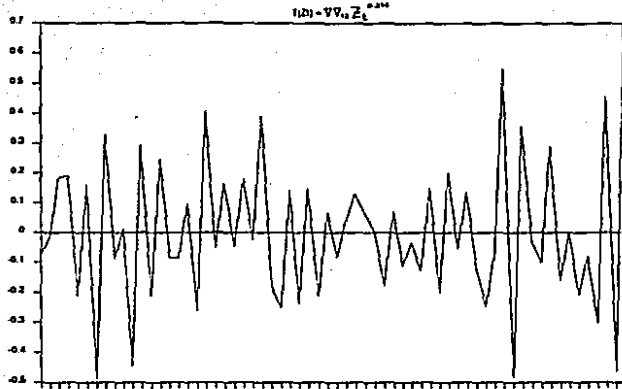
	$Z^{0.254t}$	$Z^{0.25t}$	$\text{Log}(Z_t)$	$Z^{0.24t}$
S(0,0)	0.72047	0.69372	0.71655	0.63042
S(0,1)	0.21672	0.20884	0.23152	0.19017
S(0,2)	0.34521	0.33282	0.37262	0.30342
S(1,0)	0.33119	0.31880	0.32559	0.28950
S(1,1)	0.23960	0.23078	0.24412	0.20990
S(1,2)	0.38924	0.37499	0.39957	0.34123
S(2,0)	0.38302	0.36861	0.37057	0.33454
S(2,1)	0.42359	0.40802	0.43050	0.37115
S(2,2)	0.69557	0.67013	0.71426	0.60984

Para $\lambda=0.25, \lambda=0.254, \text{Log}(Z_t)$ y $\lambda=0.24$, el mínimo se logra con S(0,1) sin embargo S(1,1) se encuentra muy cercano a S(0,1). A pesar de lo anterior, Chatfield y Prothero, Box y Jenkins analizan las serie $T(Z_t) = \nabla \nabla_{12} Z^{\lambda t}$ para $\lambda=0, 0.24$ y 0.25 respectivamente. En la gráfica 9 se muestra la serie $\{\nabla \nabla_{12} Z^{0.254t}\}$ y su respectiva FAC muestral.

Después de haber identificado varios modelos para la serie, Chatfield y Prothero, Wilson y Box y Jenkins consideraron como modelos final a $(1-\theta B) \nabla \nabla_{12} T(Z_t) = (1-\theta B^2)a_t$

Los resultados obtenidos por estos autores han sido discutidos en capítulos anteriores, sin embargo los pronósticos obtenidos para este modelo se presentan nuevamente en la

GRAFICA 9a
Serie Transformada



GRAFICA 9b
FAC Muestral

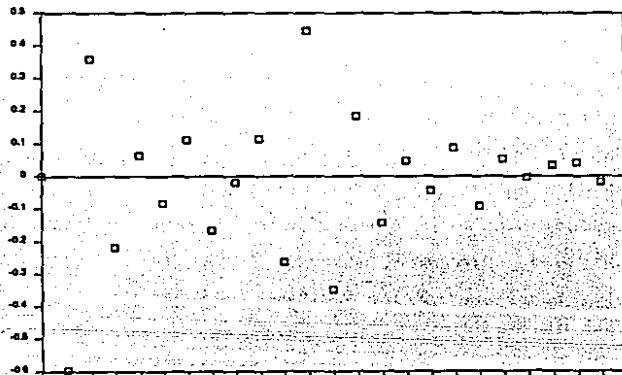


Tabla 16
Resumen de Resultados de la Estimación de Modelos

Periodo y Número de Observaciones	Modelo	Parámetros Estimados	Intervalos del 95% de Confianza	Correlación entre Parámetros (0.5 o (-0.5	análisis de residuales				
					$\hat{\mu}(\hat{a})$	$\hat{\sigma}_a$	$\hat{\sigma}^2_{0,1}$	$r(\hat{a})/0$	Residuales Grandes
Enero 1965 a Mayo 1971 N=77	$\nabla \nabla_{12} I(Z_t) =$ $(1-\theta B)(1-\theta B) \text{ at}$	$\hat{\theta}=0.47586$ $\hat{\theta}=-0.82031$	(0.2747, 0.6769) (0.7206, 0.9199)	--- ---	-0.005875 -0.32875	0.14653	38.46, 34	$r(\hat{a})=-0.256$ $r(1(\hat{a}))=0.395$	$\hat{\sigma}_{24} = 2.12 \hat{\sigma}_a$
Enero 1965 a Mayo 1971 N=77	$(1-\beta B) \nabla \nabla_{12} I(Z_t) =$ $(1-\theta B)^2 \text{ at}$	$\hat{\beta}=0.49185$ $\hat{\theta}=0.79062$	(-.6857, -.2979) (0.6911, 0.8900)	-- ---	-0.005775 0.313441	0.14624	38.19, 34	$r(\hat{a})=-0.253$ $r(1(\hat{a}))=0.399$ $r(2(\hat{a}))=-0.257$	$\hat{\sigma}_{23} = 2.18 \hat{\sigma}_a$
Enero 1965 a Nov. 1969 N=59	$\nabla \nabla_{12} I(Z_t) =$ $(1-\theta B)(1-\theta B) \text{ at}$	$\hat{\theta}=0.32634$ $\hat{\theta}=0.79744$	(0.0820, 0.5706) (0.6934, 0.9014)	--- ---	0.0004904 0.02588	0.12847	30.03, 34	$r(1(\hat{a}))=0.349$	$\hat{\sigma}_{24} = 2.44 \hat{\sigma}_a$
Enero 1965 a Nov. 1969 N=59	$(1-\beta B) \nabla \nabla_{12} I(Z_t) =$ $(1-\theta B)^2 \text{ at}$	$\hat{\beta}=0.37332$ $\hat{\theta}=0.78906$	(-.4135, -.1330) (0.6856, 0.8924)	--- ---	-0.0005327 -0.02795	0.12783	28.24, 34	$r(\hat{a})=-0.277$ $r(1(\hat{a}))=0.350$	$\hat{\sigma}_{19} = -2.20 \hat{\sigma}_a$ $\hat{\sigma}_{23} = -2.41 \hat{\sigma}_a$

Tabla 17
Resumen de Resultados de la Estimación de Modelos

Serie $Z_t = \text{C&PI}$		Transformación: $Y(Z_t) = Z_t^{0.25}$							
Período y Número de Observaciones	Modelo	Parámetros Estimados	Intervalos del 95% de Confianza	Correlación entre Parámetros >0.5 o <-0.5	Análisis de Residuales				
					$n(\hat{\alpha})$	$\hat{\sigma}_a$	$Q^*, 9.1$	$rk(\hat{\alpha})=0$	Residuales Grandes
Enero 1965 a Mayo 1971 N=77	$\nabla \nabla_{12} I(Z_t) =$ $(1-\theta B)(1-\theta B^{12})a_t$	$\hat{\theta} = 0.47630$ $\hat{\theta} = 0.82031$	(0.2753, 0.6772) (0.72021, 0.9204)	--- ---	-0.0054467 -0.32023	0.14105	38.30, 34	$r7(\hat{\alpha}) = -0.250$ $r11(\hat{\alpha}) = 0.394$	
Enero 1965 a Mayo 1971 N=77	$(1-\beta B) \nabla \nabla_{12} I(Z_t) =$ $(1-\theta B^{12})a_t$	$\hat{\beta} = -0.49250$ $\hat{\theta} = 0.79062$	(-0.6864, -0.298) (0.6908, 0.8903)	--- ---	0.0055641 0.313729	0.14077	38.23, 34	$r7(\hat{\alpha}) = -0.253$ $r11(\hat{\alpha}) = 0.399$ $r24(\hat{\alpha}) = -0.257$	$\hat{\sigma}_{23} = 2.17 \hat{\sigma}_a$
Enero 1965 a Nov. 1969 N=59	$\nabla \nabla_{12} I(Z_t) =$ $(1-\theta B)(1-\theta B^{12})a_t$	$\hat{\theta} = 0.32646$ $\hat{\theta} = 0.79744$	(0.0821, 0.5707) (0.6933, 0.9015)	--- ---	0.0004597 0.025168	0.12388	30.02, 34	$r11(\hat{\alpha}) = 0.350$	$\hat{\sigma}_{24} = 2.44 \hat{\sigma}_a$
Enero 1965 a Nov. 1969 N=59	$(1-\beta B) \nabla \nabla_{12} I(Z_t) =$ $(1-\theta B^{12})a_t$	$\hat{\beta} = -0.37308$ $\hat{\theta} = 0.78906$	(-0.6133, -0.1328) (0.6856, 0.8924)	--- ---	-0.0005366 -0.0292	0.12327	28.32, 34	$r11(\hat{\alpha}) = 0.351$	$\hat{\sigma}_{19} = -2.20 \hat{\sigma}_a$

Tabla 18
Resumen de Resultados de la Estimación de Modelos

Serie Zt = CPI		Transformación: $l(Z_t) = \log(Z_t)$							
Periodo y Número de Observaciones	Modelo	Parámetros Estimados	Intervalos del 95% de Confianza	Correlación entre Parámetros ≥ 0.5 o ≤ -0.5	análisis de residuales				
					$\hat{u}(\hat{\sigma})$	$\hat{\sigma}_a^2$	σ', σ, l	$r1(\hat{\sigma})=0$	Residuales Grandes
Enero 1965 a Mayo 1971 N=77	$\nabla \nabla_{12} l(Z_t) = (1-\theta B)(1-\theta B^{12})\alpha t$	$\hat{\theta} = -0.44675$ $\hat{\theta} = 0.81094$	$(-0.6471, -0.2463)$ $(0.7141, 0.9076)$	--- ---	0.0079047 0.410585	0.15281	32.34, 34	$r7(\hat{\sigma}) = -0.274$ $r11(\hat{\sigma}) = 0.369$	
Enero 1965 a Mayo 1971 N=77	$(1-\beta B)\nabla \nabla_{12} l(Z_t) = (1-\theta B^{12})\alpha t$	$\hat{\beta} = -0.38031$ $\hat{\theta} = 0.84375$	$(0.1709, 0.5896)$ $(0.75035, 0.9371)$	--- ---	-0.0037406 -0.19444	0.1539	30.98, 34	$r7(\hat{\sigma}) = -0.263$ $r11(\hat{\sigma}) = 0.347$	
Enero 1965 a Nov. 1969 N=59	$\nabla \nabla_{12} l(Z_t) = (1-\theta B)(1-\theta B^{12})\alpha t$	$\hat{\theta} = 0.25459$ $\hat{\theta} = 0.80469$	$(0.00555, 0.5036)$ $(0.7027, 0.9066)$	--- ---	0.0000054 0.00092499	0.14655	29.53, 34	$r11(\hat{\sigma}) = 0.366$	
Enero 1965 a Nov. 1969 N=59	$(1-\beta B)\nabla \nabla_{12} l(Z_t) = (1-\theta B^{12})\alpha t$	$\hat{\beta} = -0.30972$ $\hat{\theta} = 0.79219$	$(-0.5546, -0.06479)$ $(0.6896, 0.8947)$	--- ---	-0.0019286 -0.0884369	0.14629	28.70, 34	$r11(\hat{\sigma}) = 0.367$	$\hat{\sigma}_{19} = -2.10\hat{\sigma}_a$

Tabla 19
Resumen de Resultados de la Estimación de Modelos

Serie $Z_t = \text{Cápt}$		Transformación: $I(Z_t) = Z_t^{0.24}$							
Periodo y Número de Observaciones	Modelo	Parámetros Estimados	Intervalos del 95% de Confianza	Correlación entre Parámetros ± 0.5 o ± 0.5	análisis de residuales				
					$\hat{\alpha}$	$\hat{\sigma}_a$	D. g. l.	$r(\hat{\alpha})=0$	Residuales Grandes
Enero 1965 a Mayo 1971 N=77	$\nabla \nabla_{12} I(Z_t) = (1-\theta)(1-\theta^{12})_a t$	$\hat{\theta} = 0.47555$	(0.2745, 0.6765)	---	-0.0051179	0.12812	37.75, 34	$r(\hat{\alpha}) = -0.257$	
		$\hat{\theta} = 0.82188$	(0.7219, 0.9216)	---	-0.031956			$r(1(\hat{\alpha})) = 0.392$	
Enero 1965 a Mayo 1971 N=77	$(1-\beta)\nabla \nabla_{12} I(Z_t) = (1-\theta)^{12}_a t$	$\hat{\beta} = -0.49369$	(-.6874, -.2998)	---	-0.0051668	0.12784	38.08, 34	$r(\hat{\alpha}) = -0.254$	$\hat{\alpha}_{23} = 2.18 \hat{\sigma}_a$
		$\hat{\theta} = 0.79219$	(0.6927, 0.8916)	---	-0.32079			$r(1(\hat{\alpha})) = 0.398$ $r(2(\hat{\alpha})) = -0.257$	
Enero 1965 a Nov. 1967 N=59	$\nabla \nabla_{12} I(Z_t) = (1-\theta)(1-\theta^{12})_a t$	$\hat{\theta} = 0.32568$	(0.0812, 0.5700)	---	0.0084084	0.11301	29.99, 34	$r(1(\hat{\alpha})) = 0.351$	$\hat{\alpha}_{20} = -2.22 \hat{\sigma}_a$ $\hat{\alpha}_{24} = 2.43 \hat{\sigma}_a$
		$\hat{\theta} = 0.79844$	(0.6965, 0.9022)	---	0.024512				
Enero 1965 a Nov. 1967 N=59	$(1-\beta)\nabla \nabla_{12} I(Z_t) = (1-\theta)^{12}_a t$	$\hat{\beta} = -0.37243$	(-.6127, -.1321)	---	-0.0005415	0.11245	28.28, 34	$r(\hat{\alpha}) = -0.276$	
		$\hat{\theta} = 0.78706$	(0.6854, 0.8926)	---	-0.0323			$r(1(\hat{\alpha})) = 0.353$	

tabla 9 (que se repite para facilitar su lectura). La corrección de sesgo para estos pronósticos, considerando el método de Guerrero, Miller y Taylor se presentan en la tabla 11.

TABLA 9
Pronósticos con las diferentes transformaciones

l	Chatfield & Prothero	M.M.V. Wilson	Máxima Verosim.	Recorrido - Media	Método Guerrero	Obs. Reales
	$\log Z_t$	$Z_{0.34t}$	$Z_{0.24t}$	$Z_{0.25t}$	$Z_{0.254t}$	
78	305	304	284	286	284	260
79	482	423	410	409	407	304
80	673	574	529	511	523	390
81	990	706	743	761	735	614
82	1297	858	935	966	924	783
83	1387	902	988	1091	976	872
M_1	-318.5	-90.6	-111	-133.5	-104.3	
M'_1	-55.72%	-21.87%	-22.24%	-24.66%	-21.14	

TABLA 11

PRONOSTICO EN LA ESCALA ORIGINAL

Serie: $Z_t = C \& P_t$		Transformación $T(Z_t) = \log(Z_t)$		
Obs.	Pronóstico	Guerrero	Miller	Taylor
260	305	311	311	311
304	482	495	495	495
390	673	699	699	699
614	990	1038	1038	1038
783	1297	1374	1374	1374
872	1387	1483	1483	1483
M_1		-362.8	-362.8	-362.8
M'_1		-62.71%	-62.71%	-62.71%

Serie: $Z_t = C \& P_t$		Transformación $T(Z_t) = Z_t^{0.24}$	
Obs.	Pronóstico	Guerrero	Taylor
260	284	287	284
304	410	414	410
390	529	536	529
614	743	754	744
783	935	950	936
872	988	1006	989
	M_1	-120.6	-111.5
	M'_1	-23.91%	-22.31%

Serie: $Z_t = C \& P_t$		Transformación $T(Z_t) = Z_t^{0.25}$		
Obs.	Pronóstico	Guerrero	Miller	Taylor
260	286	288	286	286
304	409	413	409	409
390	511	518	511	511
614	761	771	762	762
783	966	980	967	967
872	1091	1109	1092	1092
	M_1	-142.6	-134	-134
	M'_1	-26.22%	-19.50%	-19.50%

Serie: $Z_t = C \& P_t$		Transformación $T(Z_t) = Z_t^{0.25}$	
Obs.	Pronóstico	Guerrero	Taylor
260	284	286	284
304	407	411	407
390	523	530	523
614	735	745	736
783	924	938	925
872	976	992	977
	M_1	-113.1	-104.8
	M'_1	-22.66%	-21.21%

5.2) Caso "Índice de Precios al Consumidor"

Los datos a considerar son:

GRAFICA 10
INDICE DE PRECIOS AL CONSUMIDOR

Base 1978 = 100

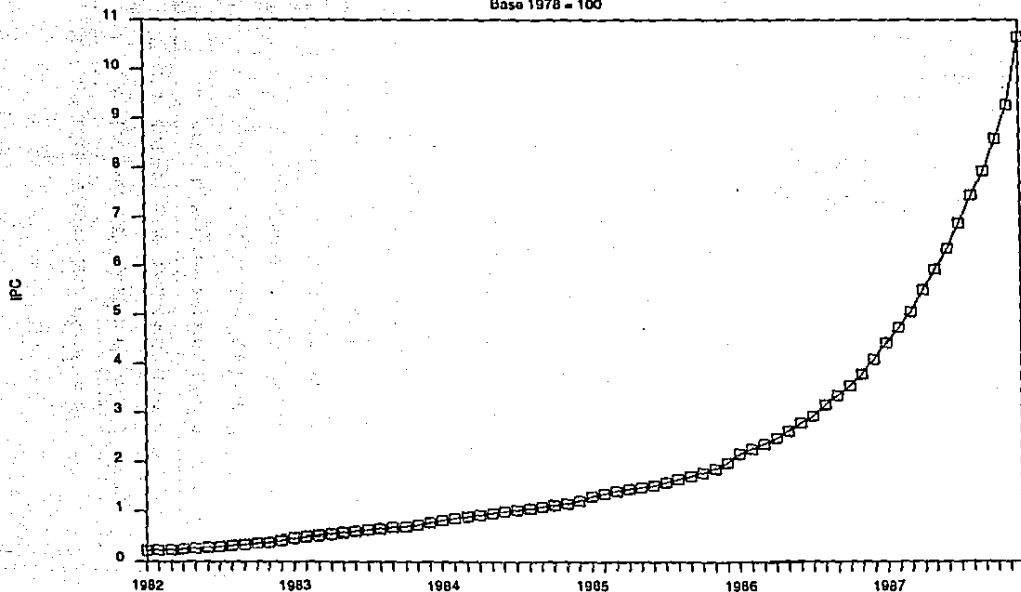


TABLA 20
 INDICE NACIONAL DE PRECIOS AL CONSUMIDOR
 Base 1978 = 100

MES	1982	1983	1984	1985	1986	1987
ENE	223.7	469.9	814.7	1309.8	2173.3	4440.9
FEB	232.5	495.1	857.8	1364.2	2269.9	4761.3
MAR	241.0	519.1	894.5	1417.1	2375.4	5076.0
ABR	254.1	552.0	933.2	1460.7	2499.4	5520.1
MAY	268.4	575.9	964.1	1495.3	2638.3	5936.2
JUN	281.3	597.7	999.0	1532.8	2807.6	6365.7
JUL	295.8	627.3	1031.8	1586.2	2947.7	6881.3
AGO	329.0	651.6	1061.1	1655.5	3182.7	7443.7
SEP	346.5	671.7	1092.7	1721.6	3373.7	7934.1
OCT	364.5	694.0	1130.9	1787.0	3566.5	8595.2
NOV	382.9	734.7	1169.7	1869.5	3807.6	9277.0
DIC	423.8	766.1	1219.4	1996.7	4108.2	10647.2

Estos datos se muestran en la gráfica 10. Es notorio que esta serie no es estacionaria, por lo que aplicaré una transformación estabilizadora de varianza utilizando el metodo de Guerrero para seleccionar la λ adecuada. Los datos obtenidos en este caso se muestran en la tabla 21.

TABLA 21
 DETERMINACION DE λ PARA LA SERIE DE IPC

Grupo	Potencia						
	1	0	-0.1	-0.103	-0.104	-0.105	-1
1	65.042	0.2142	0.1209	0.1189	0.1182	0.1175	0.00071
2	94.618	0.1543	0.0812	0.0797	0.0791	0.0786	0.00025
3	126.64	0.1248	0.0625	0.0612	0.0608	0.0603	0.00012
4	211.31	0.1320	0.0631	0.0617	0.0613	0.0608	0.00008
5	632.23	0.2122	0.0953	0.0931	0.0923	0.0916	0.00007
6	1930.5	0.2795	0.1154	0.1124	0.1114	0.1104	0.00004
$M(\lambda)$	510.06	0.1862	0.0897	0.0878	0.0872	0.0866	0.00021
$DE(\lambda)$	726.57	0.0597	0.0252	0.0247	0.0245	0.0243	0.00025
$CV(\lambda)$	1.4244	0.3210	0.281253	0.2812055	0.2812054	0.28121	1.18889

En la tabla anterior se observa que el mínimo coeficiente de variación se logra con $\lambda = -0.104$, el cual se encuentra muy cerca de cero y me llevaría a considerar la posibilidad de utilizar logaritmos; sin embargo analizaré la serie $T(Z_t) = Z^{-0.104} z_t$. En la gráfica 11 se muestra la serie transformada.

Para determinar el grado de diferenciación requerido para volver estacionaria la serie, consideraremos la FAC muestral para $\{T(Z_t)\}$, $\{\nabla T(Z_t)\}$, $\{\nabla^2 T(Z_t)\}$, $\{\nabla^3 T(Z_t)\}$ comparando sus correspondientes desviaciones estándar muestrales (véase Guerrero, 1983, cap.4).

Se tiene que

$$\begin{array}{ll} S(0) = 0.05067 & S(1) = 0.0009893 \\ S(2) = 0.0009765 & S(3) = 0.001531 \end{array}$$

El mínimo se encuentra dado por $S(2)$, aunque $S(1)$ se encuentra muy cerca de $S(2)$ por lo que habría que considerar una primera diferencia o quizá una diferencia de orden 2. En la gráfica 12 se muestran las FAC muestral para la serie transformada con una y dos diferencias.

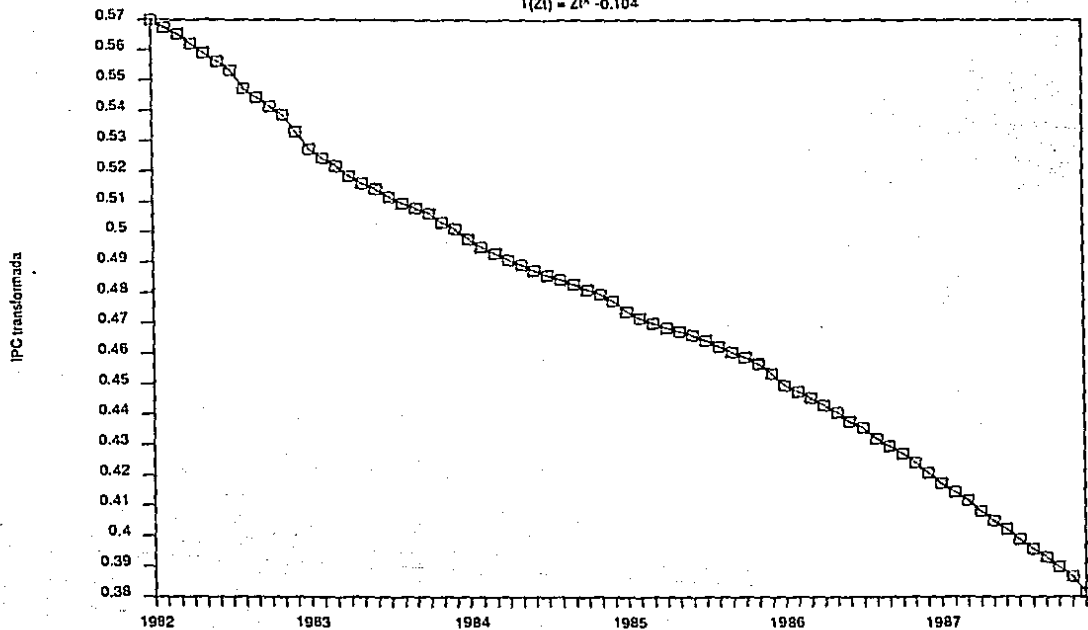
Analizaré primeramente la FAC y FACP de $\{\nabla^2 T(Z_t)\}$ para poder asociar un modelo ARIMA a la serie. Para poder decir si las autocorrelaciones son cero a partir de un retraso q deben de compararse los valores r_k con sus desviaciones estándar. En la tabla 22 se muestran los valores de las autocorrelaciones.

Si $\rho_k = 0$ para $k > 0$ la desviación estándar estimada de r_k , para $k=1,2,\dots$, es

$$\sqrt{\hat{\text{var}}(r_k)} = \sqrt{1/70 (1+2(0)^2)} = 0.119522$$

GRAFICA 11
INDICE DE PRECIOS AL CONSUMIDOR

$$T(Z_1) = Z_1^{\alpha} - 0.104$$



ahora para $|r_1|$ es dos veces este valor por lo que:

$$\sqrt{\hat{\text{var}}(r_1)} = \sqrt{1/70 (1+2(-0.255)^2)} = 0.12705$$

TABLA 22
FAC y FACP de $\{\nabla Z_t(Z_t)\}$

Retraso	1	2	3	4	5	6	7
FAC	-0.2550	-0.2125	-0.0861	0.2138	0.0322	-0.1757	0.0160
FACP	-0.2550	-0.2968	-0.2713	0.0334	0.0561	-0.1043	-0.0123
Retraso	8	9	10	11	12	13	14
FAC	0.1231	-0.1297	-0.1006	0.1143	0.1465	-0.0630	-0.0581
FACP	0.0603	-0.1494	-0.1532	0.0016	0.0871	0.0715	0.1112
Retraso	15	16	17	18	19	20	21
FAC	0.0579	-0.0777	0.0209	0.0265	0.0131	-0.0504	0.0353
FACP	0.0925	-0.1436	-0.0403	0.0124	-0.0397	-0.0063	0.1414

Sólo la autocorrelación r_1 satisface que:

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

es decir: $|r_1| > 0.25411$, entonces se supone que la única autocorrelación distinta de cero es la primera por lo que se podría considerar un IMA(2,1).

Si consideramos la FAC y FACP de $\{\nabla T(Z_t)\}$ se tiene:

$$\sqrt{\hat{\text{var}}(r_1)} = 0.141544$$

debido a que $|r_4| > 0.28308$, desechamos el modelo IMA(1,1) (ver tabla 23).

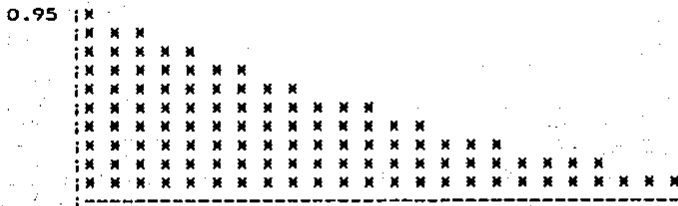
Para verificar la no significación estadística de $\hat{\beta}_{11}$ para $|i| > 1$ se tiene:

$$|\hat{\beta}_{11}| > 2/\sqrt{N-d} = 0.23735 \text{ para } |i| > 1$$

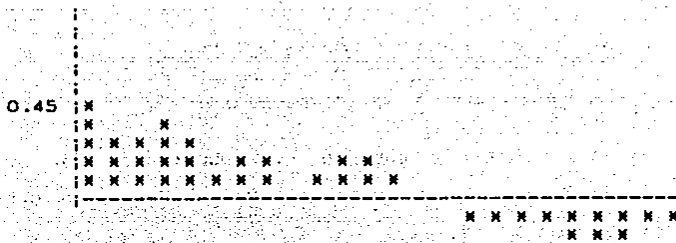
debido a que la autocorrelación parcial estimada para el retraso 4 no satisface esto se rechaza un ARI(1,1), pero quizá sea razonable pensar en un modelo ARIMA(1,1,1). Entonces

GRAFICA 12
 FUNCION DE AUTOCORRELACION MUESTRAL

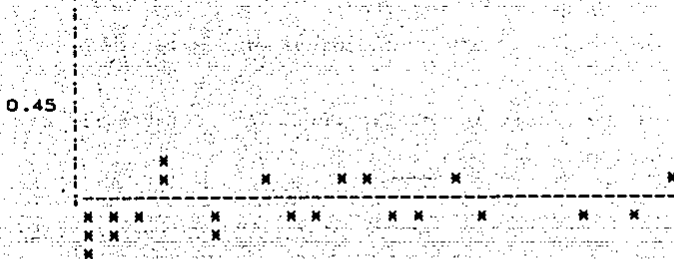
$\{T(z_t)\}$



$\{\nabla T(z_t)\}$



$\{\nabla^2 T(z_t)\}$



tentativamente se han identificado un IMA(2,1) y un ARIMA(1,1,1).

TABLA 23
FAC y FACP de $\{\nabla T(z_t)\}$

Retraso	1	2	3	4	5	6	7
FAC	0.4593	0.2611	0.2649	0.3782	0.2574	0.1063	0.1558
FACP	0.4593	0.0635	-0.1438	0.2436	-0.0158	-0.0906	0.0809
Retraso	8	9	10	11	12	13	14
FAC	0.1670	-0.0382	0.0670	0.1987	0.2063	0.0597	-0.0105
FACP	-0.0161	-0.1271	0.1079	0.1648	0.0394	-0.0570	-0.0834
Retraso	15	16	17	18	19	20	21
FAC	-0.0036	-0.0640	-0.0877	-0.0774	-0.1224	-0.1668	-0.1617
FACP	-0.1437	-0.1601	0.0071	0.0198	-0.0773	-0.0049	-0.0092

En la tabla 24 se presenta un resumen de resultados de la estimación de los dos modelos considerados para la serie IPC, usando la transformación potencia con $\lambda = -0.104$. (para detalles de la tabla remitase a Guerrero, 1983).

La gráfica de los residuales $\{a_t\}$ de ambos modelos se presenta en la gráfica 13, así como las autocorrelaciones muestrales en la tabla 25.

TABLA 25
FAC DE LOS RESIDUALES DE
MODELOS ARIMA(1,1,1), IMA(2,1)

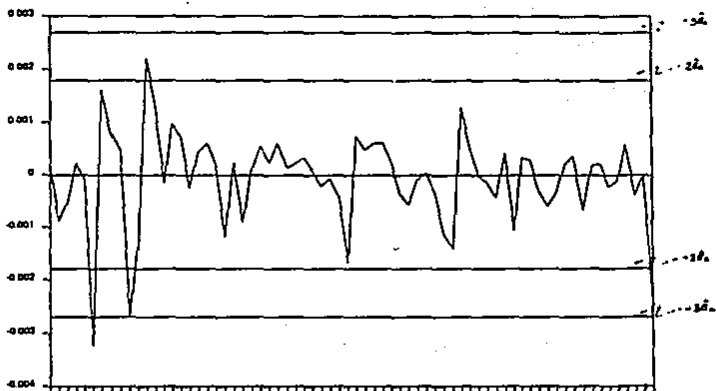
Retraso	1	2	3	4	5	6	7
ARIMA	-0.0045	-0.2541	-0.1201	0.1873	0.0335	-0.1702	-0.0242
IMA	0.1233	-0.1575	-0.0815	0.1633	0.0163	-0.1590	-0.0509
Retraso	8	9	10	11	12	13	14
ARIMA	0.0596	-0.1582	-0.0928	0.1642	0.2004	-0.0178	-0.0651
IMA	0.0142	-0.1661	-0.0726	0.1847	0.2186	0.0165	-0.0457
Retraso	15	16	17	18	19	20	21
ARIMA	0.0107	-0.0628	-0.0114	0.0306	-0.0069	-0.0680	-0.0314
IMA	0.0035	-0.0638	-0.0323	0.0070	-0.0351	-0.0902	-0.0508

Con estos datos podré llevar a cabo la verificación de

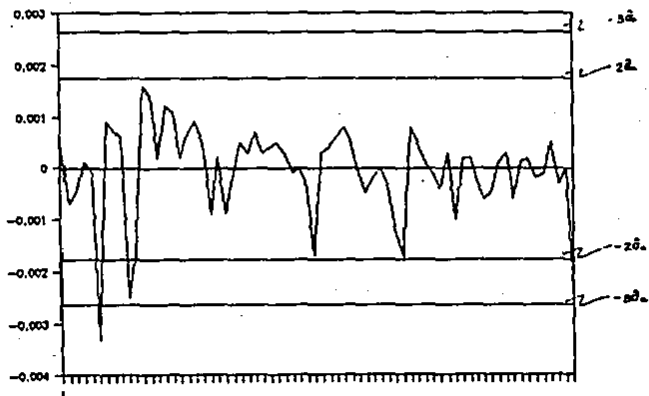
Tabla 24
Resumen de Resultados de la Estimación de Modelos

Serie $Z_t = IPCT$		Transformación: $Y(Z_t) = Z_t$							
Periodo y Número de Observaciones	Modelo	Parámetros Estimados	Intervalos del 95% de Confianza	Correlación entre Parámetros >0.5 o <-0.5	análisis de residuales				
					$\hat{\mu}(\hat{\alpha})$	$\hat{\sigma}_a$	Q*,d.l	$r_2(\hat{\alpha})=0$	Residuales Grandes
Enero 1982 a Dic. 1987 N=72 Modelo A	$(1-\hat{\beta})\nabla I(Z_t)=$	$\hat{\beta}=0.997539$	(0.9812, 1.0138)	---	-0.0000661	0.0008961	34.45,22	$r_2(\hat{\alpha})=-0.2541$	$\hat{\sigma}_a = -3.6\hat{\sigma}_a$ $\hat{\sigma}_a = -2.9\hat{\sigma}_a$ $\hat{\sigma}_a = 2.4\hat{\sigma}_a$ $\hat{\sigma}_a = -2.7\hat{\sigma}_a$
	$(1-\hat{\theta})\Delta t$	$\hat{\theta}=0.4706487$	(-.3011, -.6401)	---	-0.617808				
Enero 1982 a Dic. 1987 N=72 Modelo B	$\nabla^2 I(Z_t)=$	$\hat{\theta}=0.68630$	(0.5105, 0.8620)	---	-0.0000571	0.0008807	30.72,23	---	$\hat{\sigma}_a = -3.7\hat{\sigma}_a$ $\hat{\sigma}_a = -2.8\hat{\sigma}_a$
	$(1-\hat{\theta})\Delta t$			---	0.543111				

GRAFICA 13 A
Residuales del modelo ARMA(1,1,1)



GRAFICA 13 B
RESIDUALES DE IMA(2,1)



los supuestos de los modelos.

1.- Para el modelo ARIMA se tiene

$$|\sqrt{N-d-p} \ m(\hat{a}) / \hat{\sigma}_a| = 0.617808$$

para el modelo IMA se tiene

$$|\sqrt{N-d-p} \ m(\hat{a}) / \hat{\sigma}_a| = 0.54311$$

por lo que no podemos decir que la media de los residuales es distinta de cero para ambos modelos.

2.- La varianza constante se verifica con la gráfica 13.

3.- Para el modelo A, la χ^2 con 22 grados de libertad son

$$\chi^2_{0.05}(22) = 33.9 \quad \chi^2_{0.10}(22) = 30.8 \quad \chi^2_{0.25}(22) = 26.0$$

para el modelo B se tiene

$$\chi^2_{0.05}(23) = 35.2 \quad \chi^2_{0.10}(23) = 32.0 \quad \chi^2_{0.25}(23) = 27.1$$

por lo que los valores respectivos de 34.45 y 30.726 no conducen al rechazo de la hipótesis de que los residuales son ruido blanco al nivel de significación del 5% en ninguno de los modelos.

4.- En la gráfica de los residuales de ambos modelos, se observa que solamente cuatro residuales se encuentran fuera de la banda $(-2 \hat{\sigma}_a, 2 \hat{\sigma}_a)$ para el modelo ARIMA y 2 para el IMA. Esto indica que no hay violación de importancia al supuesto de distribución Normal.

5.- Para ambos modelos se tiene una observación sospechosa, la cual debe ser investigada para saber si corresponde a algún evento exógeno que pudiera explicarla.

6.- Para el modelo A, los intervalos de confianza para $\hat{\theta}$ y $\hat{\beta}$ no cubren el valor de cero, podríamos decir entonces que ambos parámetros son requeridos. Para el modelo B, el intervalo de confianza para el parámetro $\hat{\theta}$ tampoco cubre el valor cero. Se

diría entonces, que se requiere de dicho parámetro para explicar el comportamiento de la serie.

7.- Para verificar la admisibilidad del modelo revisaremos si los intervalos de confianza para los parámetros se encuentran totalmente incluidos dentro de la región admisible de los modelos MA(1) y ARMA(1,1) respectivamente. Para el MA(1) es (-1,1) y se cumple, pero para $\hat{\beta}$ su intervalo cubre el valor 1 lo cual conduciría a tener una diferencia más.

8.- No existen correlaciones altas positivas en ninguno de los modelos.

Debido a que el modelo A no cumple con el supuesto de admisibilidad debemos considerar al IMA (2,1) como el modelo adecuado para la serie en estudio, siendo este

$$\nabla^2 T(\text{IPC}_t) = (1 - 0.68630 B)a_t \quad \text{para } t=3, \dots, 77$$

Los pronósticos obtenidos por este modelo, así como los pronósticos corregidos usando el método de Guerrero, se muestran en la tabla 26

TABLA 26
Pronósticos para la serie IPC

Adelanto	1	2	3	4	5	6
Pronóstico	11727.7	12959.6	14298.6	15792.0	17506.0	19375.5
Prón. Corregido	11731.3	12969.2	14320.9	15836.2	17581.1	19501.3

Debido a que en 1988 cambiarán las condiciones económicas en México por el "Pacto de Solidaridad" las observaciones de este año no pueden ser comparadas con el pronóstico obtenido por el modelo IMA. Para efectos de comparación conviene más estimar el modelo con dato hasta diciembre de 1986 y pronosticar todos los meses de 1987 y comparar los pronósticos con los valores

actuales.

TABLA 27
Pronóstico para la serie IPC
(Pronósticos corregidos)

Obs.	Pronóstico	M. Guerrero	M. Taylor
61	4440.9	4390.7	4390.7
62	4761.3	4694.7	4694.7
63	5074.0	5022.1	5022.2
64	5520.1	5387.5	5387.6
65	6936.2	5768.8	5769.1
66	6365.7	6180.2	6180.5
67	6881.3	6624.1	6624.7
68	7443.7	7103.6	7104.4
69	7934.1	7621.6	7622.6
70	8595.2	8201.8	8203.4
71	9277.0	8809.3	8811.5
72	10647.2	9466.8	9469.8
	M ₁	221.84	299.78
	M ¹ ₁	2.80%	3.71%

CONCLUSIONES

Como se mencionó anteriormente, existen varios métodos en el área estadística para poder llevar acabo el análisis de Series de Tiempo. Makridakis (1976) hace una recopilación y análisis de estos métodos para tratar de guiar a la comunidad estadística en la selección de la metodología. El señala que la exactitud del pronóstico se ve influenciada por el método utilizado y por el número de observaciones con que se cuente, así como los programas y computadora que se utilice para efectuar dichos cálculos.

Makridakis hace referencia a los resultados que Reid (1971), Granger y Newbold (1972) y Groff (1973) obtuvieron al comparar un número limitado de metodologías en lo que a exactitud se refiere. Ellos reportan que el método de Box y Jenkins es superior a los otros métodos comparados, cuando la serie contienen más de 50 observaciones y que se ve afectado en su exactitud cuando la transformación aplicada a la serie no se selecciona de manera adecuada.

Tomando en cuenta el problema de la selección del parámetro λ , en esta tesis se han presentado los métodos de Guerrero, siendo el objetivo principal, el de mostrar que dichos métodos cumplen con los requisitos necesarios para llevar a cabo un análisis efectivo de series de tiempo y así poder dar al estadístico otra herramienta de trabajo que facilite su análisis.

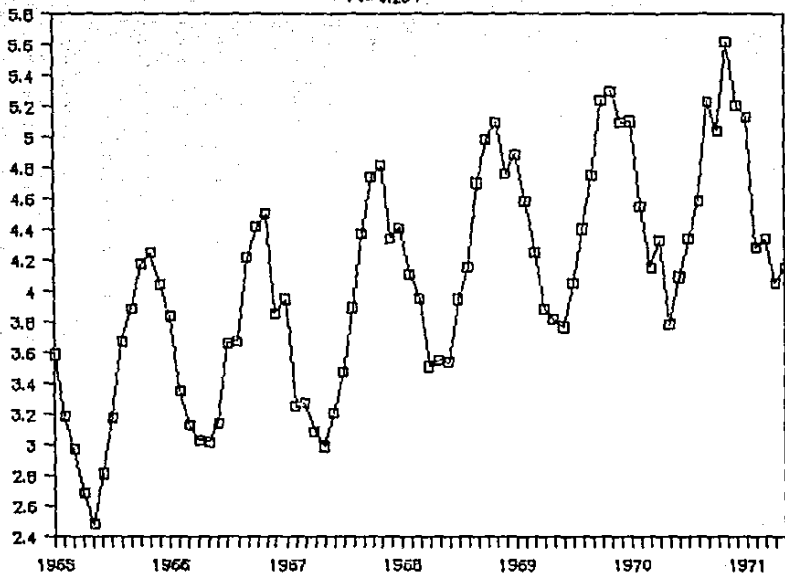
La "complejidad" es un factor que determina si un

estadístico usa tal o cual método para desarrollar su análisis y es importante hacer notar que los dos métodos de Guerrero son relativamente fáciles de calcular, es decir, pueden calcularse con los resultados de paquetes estadísticos que tengan implementada la metodología de Box y Jenkins, y no requieren de un experto en probabilidad y estadística para desarrollarlos e interpretar sus resultados una vez que han sido aplicados, por lo tanto el factor complejidad no es un factor que dificulte su uso en la práctica.

Uno de los puntos que observé al desarrollar este trabajo, es que para algunos métodos no hay una documentación adecuada o completa y no hay una descripción accesible de como utilizar el método y como interpretar los resultados, o en otros casos, la descripción de los métodos es muy matemática y poco explicativa para los no-introducidos en la materia, pero que desean hacer un análisis de series de tiempo. Al respecto Chatfield y Prothero (1973) comentan que "básicamente el análisis de series de tiempo es un arte que requiere conocimientos especializados, que muy pocos estadísticos tienen y el cual desalienta a los principiantes". Este comentario puede aplicarse todavía en 1989 a algunos métodos mostrados en este trabajo, como el de Box y Cox, el de Granger y Newbold o el de Neyman y Scott.

Al analizar los resultados obtenidos al aplicar los métodos de Guerrero a la serie de Chatfield y Prothero nos damos cuenta que los métodos han cumplido con los objetivos esperados, aunque no es claro que el método para corregir por sesgo sea tan

GRAFICA 14
SERIE TRANSFORMADA
 $\lambda=0.254$



bueno como los otros métodos considerados.

En primera instancia, al transformar la serie de Chatfield y Prothero, utilizando el parámetro $\lambda=0.254$ se logra obtener una serie estacionaria en función de la serie original. Esto es, definitivamente se logró estabilizar la varianza (esto se puede constatar en la gráfica 14 donde se presenta la serie transformada) más aún, esta transformación mejora mucho a la Logarítmica (véase la gráfica 2).

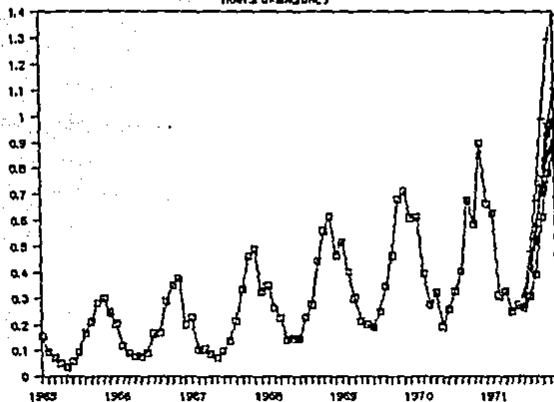
El método de Guerrero para seleccionar el parámetro λ es mucho más accesible que el de Box y Cox y proporciona resultados muy cercanos al de Máxima Verosimilitud, por lo que se puede confiar en la exactitud relativa del método, al menos en lo que se pudo observar del análisis aquí realizado con la serie de Chatfield y Prothero.

Por último los pronósticos obtenidos con $\lambda=0.254$ son mucho más cercanos a las observaciones reales que los pronósticos que se obtienen con los otros valores de λ (esto se confirma con la media de los errores de pronóstico expresada en porcentaje). Obsérvense las gráficas 15 y 16.

Estos resultados nos hacen pensar que el método de Guerrero para seleccionar λ , cumple con los objetivos y el analista podrá tener la seguridad de que los resultados ayudarán a obtener un buen análisis de series de tiempo.

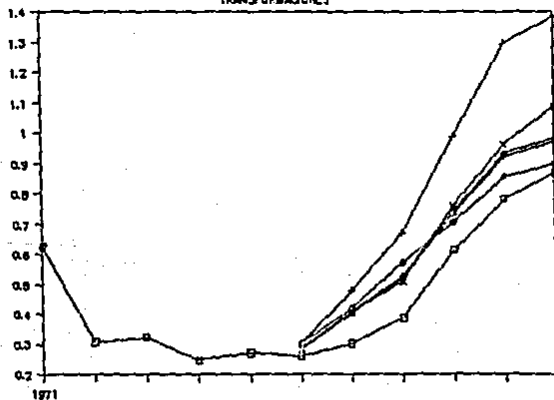
En lo que se refiere a la corrección del sesgo introducido por la transformación inversa, el método de Guerrero también da buenos resultados. Las fórmulas (4.5.11) y (4.5.12) no son tan complicadas y pueden ser aplicadas sin mucho problema;

GRAFICA 15
 PRONOSTICOS CON LAS DIFERENTES
 TRANSFORMACIONES



□ 085 + L00 ◻ 0.24 Δ 0.24 × 0.25 ▽ 0.284

GRAFICA 16
 PRONOSTICOS CON LAS DIFERENTES
 TRANSFORMACIONES



□ 085 + L00 ◻ 0.24 Δ 0.24 × 0.25 ▽ 0.284

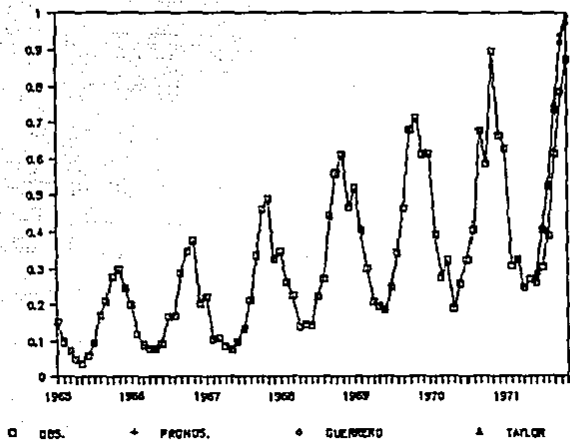
sin embargo habría que demostrar con otros análisis que el método es al menos tan bueno como los competidores.

Las gráficas 17 y 18 muestran las observaciones reales y los pronósticos corregidos con los métodos de Guerrero y Taylor . Obsérvese que estos dan resultados muy similares para $\lambda=0.254$ (observa la media del error de pronóstico).

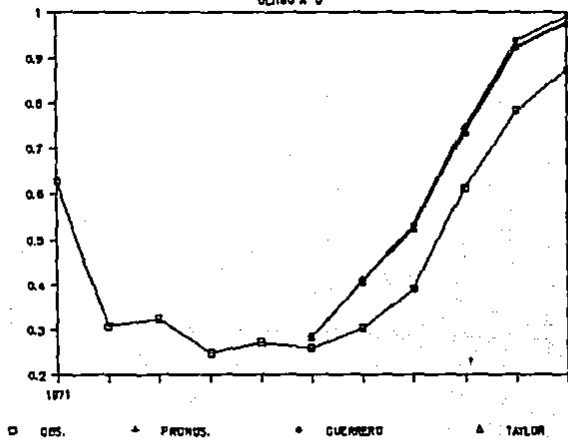
Aunque Guerrero hace notar que estos métodos son aproximaciones, puedo atreverme a decir que los métodos de Guerrero ofrecen al analista buenos resultados, tanto como los que se podrían obtener con métodos más exactos. Esto se puede observar en los resultados del análisis de la serie "Índice de precios al consumidor"

Los métodos aquí presentados han requerido de muchos años de estudio y de esfuerzo personal y de grupo, por lo que el trabajo a realizar ahora es el de la divulgación para que la comunidad estadística tenga conocimiento de su existencia y más aún, hacer del conocimiento de los no-especializados la existencia del análisis de series de tiempo así como del beneficio que ofrece su uso.

GRAFICA 17
PRONOSTICOS CORREGIDOS



GRAFICA 18
PRONOSTICOS CORREGIDOS
ULTIMO AÑO



APENDICE

A.1) Propiedades de los Polinómios Hermitianos.

El Sistema de Polinómios Hermitianos $H_n(x)$ se define en terminos de la Distribución Normal Estándar como:

$$H_n(x) = \exp(x^2/2) (-d/dx)^n \exp(-x^2/2) \text{ ó } H_n(x) = (-1)^n \rho^{(n)}(x) / \rho(x)$$

donde ρ es la función de densidad de la distribución Normal Estándar. Explicitamente, se puede escribir como:

$$H_n(x) = n! \sum_{m=0}^{[n/2]} (-1)^m \{2^m (n-2m)!\}^{-1} x^{n-2m}$$

donde $[N]$ es el entero mas grande menor o igual a N . Entonces se tiene

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1,$$

$$H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3, \quad H_5(x) = x^5 - 10x^3 + 15x$$

y así sucesivamente.

Se define la Esperanza de los operadores E_0 y E como:

$$E_0\{\psi(x)\} = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x) \exp(-x^2/2) dx,$$

$$E\{\psi(x)\} = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} \psi(x) \exp(-(x-\mu)^2/2\sigma^2) dx$$

Los Polinómios Hermitianos constituyen un sistema ortogonal con respecto a la función de densidad de la distribución Normal Estándar por lo que

$$E_0\{H_n(x)H_k(x)\} = \begin{cases} 0, & n \neq k, \\ n!, & n = k. \end{cases} \quad (A.1)$$

y desde que $H_0(x) = 1$, se sigue que

$$E_0\{H_n(x)\} = 0, \quad n > 0. \quad (A.2)$$

Los Polinómios obedecen la fórmula recursiva

$H_{n+1}(x) - xH_n(x) + nH_{n-1}(x) = 0$ y tiene como función generadora

$$\exp(tx - t^2/2) = \sum_{n=0}^{\infty} H_n(x) t^n / n! \quad (\text{A.3})$$

Un resultado particular importante, es que la función de densidad Normal Bivariada para variables con la misma varianza puede ser escrita como

$$f(x_1, x_2) = \frac{(1 - \rho^2)^{-1/2}}{2\pi\rho^2} \exp\left\{-1/2 \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{\sigma^2(1 - \rho^2)}\right\}$$

$$= \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(x_1^2 + x_2^2)\right\} \sum_{n=0}^{\infty} \frac{\rho^n H_n(x_1/\sigma) H_n(x_2/\sigma)}{n!}$$

entonces se sigue que si X y Y se distribuyen como una función Normal Bivariada con media cero, varianza uno y coeficiente de correlación

$$E\{H_n(X); Y=y\} = \rho^n H_n(y)$$

y

$$E\{H_n(X)H_k(Y)\} = \begin{cases} 0, & n \neq k, \\ \rho^n n!, & n = k \end{cases} \quad (\text{A.4})$$

Un resultado muy útil para el Método de Granger y Newbold es

$$H_n(Ax+Bx) = \sum_{k=0}^n n C_k A^k B^{n-k} H_k(x) H_{n-k}(y), \quad \text{para } A^2+B^2=1. \quad (\text{A.5})$$

Un método general para obtener la expansión de polinomios Hermitianos para una función $T(x)$ se sigue de que si

$$\sum_{j=0}^{\infty} \alpha_j H_j(x) = T(x)$$

entonces $\alpha_n = E_0\{(d/dx)^n T(x)\}/n!$. Si $T(x)$ es una función tal que $E_0\{T(x)\} < \infty$, entonces siempre existirá una expansión asintótica de la forma $T(x) = \lim_{N \rightarrow \infty} S_N(x)$ cuando $N \rightarrow \infty$ donde

$$S_N(x) = \sum_{j=0}^N \alpha_j H_j(x)$$

y tal que $E_0\{[T(x) - S_N(x)]^2\} = 0$ cuando $N \rightarrow \infty$. Algunas expansiones particulares son

$$|x|^\lambda = \alpha_0 \sum_{m=1}^{\infty} \left\{ \prod_{v=0}^{m-1} (\lambda - 2v) \right\} H_{2m}(x) / (2m)!, \quad \lambda + 1 > 0,$$

donde

$$\alpha_0 = 2^{\lambda/2} \Gamma(1/2(\lambda+1)) / \Gamma(1/2),$$

y

$$|x|^{\lambda+1} = (\lambda+1) \alpha_0 \left\{ H_1(x) + \sum_{m=1}^{\infty} \prod_{v=0}^{m-1} (\lambda - 2v) H_{2m+1}(x) / (2m+1)! \right\}, \quad \lambda + 2 > 0$$

La expansión para $\exp(tx)$ se obtiene de la función generadora (A.3).

A.2) Programa que calcula el parametro Lambda dado un conjunto de observaciones de acuerdo al Metodo de Guerrero.

{ Programa: Maria del Carmen Minor Molinã Noviembre, 1987 }

type

Grupo1 = array [1..12] of Real;
Reg1 = array [1..30] of Real;

Const

Enc01 = ' DATOS DE LA SERIE.';
Enc02 = ' ENE FEB MAR ABR MAY JUN JUL AGO SEP OCT NOV DIC';
Enc03 = ' CALCULOS PARA SELECCIONAR LA POTENCIA
(LAMBDA)';
Enc04 = ' DE LA TRANSFORMACION ESTABILIZADORA DE
VARIANZA';
Enc05 = ' LAMBDA';

var

C : Integer ;
Corte : Real ;
CV : array [1..30] of Real ;
DS : array [1..10] of Real; { Desv. Estandard }
DS1 : array [1..30] of Real; { Desv. Estandard }
Elementos : Integer;
Grupos : Integer;
H : Integer; {Numero de grupos}
I : Integer;
J : Integer;
L : Real ;
Lamda : array [1..30] of Real ;
Low1 : Integer ;
Lower : Real ;
Max : Real ;
Min : Real ;
Minimo : Real ;
N : Integer;
R : Integer; {Numero de elementos por grupo}
Res : Integer;
Sum : Real ;
Tabla : array [1..10] of Grupo1; { Observaciones Iniciales }
Tablam : array [1..10] of Reg1; { Tabla de Constantes }
Z : array [1..10] of Real; { Medias por Grupo }
Z1 : array [1..10] of Real; { Medias de Lambda por Corte }

{ Esta rutina calcula la media por Grupo I dada la Tabla [I,J] de valores }

Procedure Media;

Begin

For I := 1 to H do
Z[I] := 0.00 ;

```

for I := 1 to H do
  Begin
    for J := 1 to R do
      Z[I] := Z[I] + (Tabla[I,J] / R);
    [ Writeln ('Media Grupo ',I,' = ',Z[I]:8:6); ]
    End;
End;

{ Esta rutina calcula la Desviacion Estandar por Grupo I dada la Tabla }
{ de valores Tabla [I,J] }
Procedure DevStd;
Begin
  for I := 1 to H do
    Begin
      Sum := 0.00 ;
      for J := 1 to R do
        Begin
          Sum := Sum + ((Exp ( 2 * Ln(Abs(Tabla[I,J] - Z[I] ))));
        End;
      Sum := Sum / ( R - 1 ) ;
      DS[I] := Sqrt (Sum) ;
    [ Writeln ('Desviacion Grupo ',I,' = ',DS[I]:8:6); ]
    End;
End;

{ Esta rutina calcula la media dada la Tablam [I,J] de valores }
Procedure MediaI;
Begin
  for I := 1 to C do
    ZI[I] := 0.00 ;
    for J := 1 to C do
      Begin
        for I := 1 to H do
          Begin
            ZI[J] := ZI[J] + ( Tablam[I,J] / H );
          End;
        [ Writeln ('Media: ',J,' = ',ZI[J]); ]
      End;
    End;
End;

{ Esta rutina calcula la Desviacion Estandar de la Tabla }
{ de valores Tablam [I,J] }
Procedure DevStdI;
Begin
  for J := 1 to C do
    Begin
      Sum := 0.00 ;
      for I := 1 to H do
        Begin
          Sum := Sum + ((Exp ( 2 * Ln(Abs(Tablam[I,J] - ZI[J] ))));
        End;
      Sum := Sum / (H - 1);
      DS1[J] := Sqrt (Sum) ;
    [ Writeln ('Desviacion: ',DS1[J]); ]
    End;
  End;
End;

```

```
End;  
End;
```

```
{ Esta rutina calcula el Coeficiente de Variacion }
```

```
Procedure CalCV;
```

```
Begin
```

```
  for I := 1 to C do
```

```
    Begin
```

```
      CV[I] := DS1[I] / Z1[I];
```

```
      { Writeln ('CV ',I,'=',CV[I]); }
```

```
    End;
```

```
  for I := 1 to C-1 do
```

```
    If CV[I] < CV[I+1] then
```

```
      Begin
```

```
        CV[I+1] := CV[I];
```

```
        Lower := CV[I];
```

```
        LowI := I;
```

```
      End
```

```
    else
```

```
      Begin
```

```
        Lower := CV[I+1];
```

```
        LowI := I+1;
```

```
      End
```

```
End;
```

```
{ Esta rutina calcula los valores de la tabla TabLam [I,J] dada una }
```

```
{ Lambda = L }
```

```
Procedure Lambda;
```

```
Begin
```

```
  for I := 1 to H do
```

```
    Begin
```

```
      L := MinI;
```

```
      For J := 1 to C do
```

```
        Begin
```

```
          Lamda[J] := L;
```

```
          TabLam[I,J] := DS[I] / Exp ((1 - L) * Ln (Z1[I]));
```

```
          L := L + Corte;
```

```
        End;
```

```
      End;
```

```
End;
```

```
{ Esta rutina separa los resultados en grupos de 8 columnas e imprime }
```

```
Procedure Imprime;
```

```
Begin
```

```
  If C > 8 then
```

```
    begin
```

```
      Writeln (Lst);
```

```
      Writeln (Lst,Enc03);
```

```
      Writeln (Lst,Enc04);
```

```
      Writeln (Lst);
```

```
      Write (Lst,'Lambda');
```

```
      for I := 1 to 8 do
```



```

    write (Lst,Lamda[1]:9:5);
writeln (Lst);
Writeln (Lst,'Grupo ');
for I := 1 to H do
  begin
    write (Lst,1:6);
    for J := 1 to 8 do
      Write (Lst,' ',TabLam[1,J]:8:4);
    Writeln (Lst);
  end;

writeln (Lst);
write (Lst,'Media ');
for J := 1 to 8 do
  write (Lst,Z1[J]:9:5);
writeln (Lst);

write (Lst,'DevStd');
for J := 1 to 8 do
  write (Lst,DS1[J]:9:5);
writeln (Lst);

write (Lst,'Cf Var');
for J := 1 to 8 do
  write (Lst,CV[J]:9:5);
writeln (Lst);

writeln (Lst);
Writeln (Lst,EncO3);
Writeln (Lst,EncO4);
Writeln (Lst);
Write (Lst,'Lambda');
for I := 9 to C do
  write (Lst,Lamda[1]:9:5);
writeln (Lst);
writeln (Lst,'Grupo ');
for I := 1 to H do
  begin
    write (Lst,1:6);
    for J := 9 to C do
      Write (Lst,' ',TabLam[1,J]:8:6);
    Writeln (Lst);
  end;

writeln (Lst);
write (Lst,'Media ');
for J := 9 to C do
  write (Lst,Z1[J]:9:5);
writeln (Lst);

write (Lst,'DevStd');
for J := 9 to C do
  write (Lst,DS1[J]:9:5);
writeln (Lst);

```

```

write (Lst,'Cf Var');
for J := 9 to C do
  write (Lst,CV[J]:9:5);
writeln (Lst);
end (if)
else
begin
  writeln (Lst);
  writeln (Lst,EncO3);
  writeln (Lst,EncO4);
  writeln (Lst);
  write (Lst,'Lambda');
  for I := 1 to C do
    write (Lst,Lamda[I]:9:5);
  writeln (Lst);
  writeln (Lst,'Grupo ');
  for I := 1 to H do
    begin
      write (Lst,I:6);
      for J := 1 to C do
        write (Lst,' ',TabLam[I,J]:8:6);
      writeln (Lst);
    end;
  writeln (Lst);
  write (Lst,'Media ');
  for J := 1 to C do
    write (Lst,Z1[J]:9:5);
  writeln (Lst);
  write (Lst,'DsvStd');
  for J := 1 to C do
    write (Lst,DS1[J]:9:5);
  writeln (Lst);
  write (Lst,'Cf Var');
  for J := 1 to C do
    write (Lst,CV[J]:9:5);
  writeln (Lst);
end;
End;
{ Esta rutina carga datos iniciales en la Tabla[I,J] }
Procedure LeeDatos;
Begin
  for I := 1 to H do
    for J := 1 to R do

```

```

Begin
  Write ('Elemento ',I,',',J,' ');
  Readln (Tabla[I,J]);
End;

Writeln (Lst,Enc01);
Writeln (Lst,Enc02);

for I := 1 to H do
  Begin
    for J := 1 to R do
      Begin
        write (Lst,Tabla[I,J]:5:0);
      End;
      writeln (Lst);
    End;
  End;
End;

( ***** )
( ***** Comienza programa principal ***** )
( ***** )
Begin
  Minimo := -0.0000000001 ;
  Write ('Digite numero de Observaciones:');
  Readln (N);
  Write ('Numero de Observaciones por grupo:');
  Readln (R);
  H := N Div R;
  Res := N mod R;
  Writeln (H,' Grupos de ',R,' elementos, sobrando ',Res,' observaciones
');
  LeeDatos;
  Media;
  DevStd;
  Writeln ('Digite Intervalo de Lambdas:');
  Write ('Minimo: ');
  Readln (Min);
  Write ('Maximo: ');
  Readln (Max);
  Write ('Cortes de las Lambdas:');
  Readln (Corte);
  C := Trunc (( Max - Min ) / Corte ) + 1 ;
  Writeln ('Se generaran ',C,' cortes');
  Lambda;
  MediaI;
  DevStdI;
  CalCV;
  Imprime;
  Writeln (Lst);
  Writeln (Lst,'CV Minimo es: ',Lower:8:6,' Para Lambda
: ',Lamda[LOWI]:8:6);
End.

```

A.3) Programa que calcula el factor de correccion de sesgo introducido por la Transformacion Inversa de acuerdo con el Metodo de Guerrero.

```
Enc01 =
'
          Pronosticos en la Escala Original':
Enc02 =
' h T(Zt)(h) Var[et(h)] T [T(Zt)(h)]    Ct,\(h)    Zt(h) Int.de 90%':
Enc03 = ' de Conf. ':
Linea =
'-----';
```

var

```
Const1 : array [1..12] of Real;
H       : Integer ;
I       : Integer ;
Inv1   : array [1..12] of Real;
LimInf : array [1..12] of Real;
LimSup : array [1..12] of Real;
Lambda : Real ;
Parcial1 : Real;
Parcial2 : Real;
Parcial3 : Real;
Serie   : string[8] ;
t       : integer;
Texto  : string[2];
Trans1 : array [1..12] of Real;
Var1   : array [1..12] of Real;
Z      : array [1..12] of Real;
```

Procedure Constante;

Begin

```
if ( Lambda <= 1 ) and ( Lambda <> 0.00 ) then
  for I := 1 to H do
    Begin
      Parcial1 := 2 * Lambda * ( Lambda - 1.00 );
      { Writeln ( 'Parcial1 ',Parcial1); }
      Parcial2 := Exp((-2.00) * Ln ( 1.00+Lambda*Trans1[I] ));
      { Writeln ( 'Parcial2 ',Parcial2); }
      Parcial3 := Sqrt ( 1 - ( Parcial1 * Parcial2 * Var1[I] ) ) / 2 ;
      { Writeln ( 'Parcial3 ',Parcial3); }
      Const1[I] := Exp ( ( 1/Lambda ) * Ln( 1/2+Parcial3 ));
      { Writeln ( 'Const ',I,' ',Const1[I]); }
    End
  else
    for I := 1 to H do
      Begin
        Const1[I] := Exp (- Var1[I] / 2);
        { Writeln ( 'Const ',I,' = ',Const1[I]:8:4); }
      End
```

End;

{ Esta rutina calcula los Limites Superior e Inferior del Intervalo }
{ de Confianza al 90% del pronostico }

Procedure Intervalo;

Begin

for I := 1 to H do

Begin

Parcial1 := Trans1[I] + (1.630) * Sqrt (Var1[I]);

LimSup[I] := Const1[I] * Exp(-Ln(Parcial1) / Lambda);

{ Writeln ('Lim Sup I.C.: ',I,' ',LimSup[I]); }

Parcial1 := Trans1[I] - (1.630) * Sqrt (Var1[I]);

LimInf[I] := Const1[I] * Exp(-Ln(Parcial1) / Lambda);

{ Writeln ('Lim Inf I.C.: ',I,' ',LimInf[I]); }

End;

End;

Procedure Optimo;

Begin

for I := 1 to H do

Begin

Invl[I] := Exp (Ln (Trans1[I]) / Lambda);

Z[I] := Const1[I] * Invl[I];

{ Writeln ('Media: ',I,' ',Z[I]); }

End;

End;

Procedure Imprime;

Begin

Writeln (Lst,Enc01);

Writeln (Lst,' Serie: ',Serie,',', Origen: t = ',t;',

Lambda = ',Lambda);

Writeln (Lst,Linea);

Writeln (Lst,Enc02);

Writeln (Lst,Enc03);

Writeln (Lst,Linea);

for I := 1 to H do

Begin

Write (Lst,I:2,' ',Trans1[I]:8:6,' ',Var1[I]:10,'
,',Invl[I]:9:5,' ');

Write (Lst,Const1[I]:17,' ',Z[I]:8:4,' ');

Write (Lst,'(',LimSup[I]:7:4,' ',LimInf[I]:8:4,')');

Writeln (Lst);

End

End;

Procedure LeeDatos;

Begin

for I := 1 to H do

Begin

```

        Write ('T(zt)('I,'): ');
        Readln (Transi[I]);
    End;

    for I := 1 to H do
        Begin
            Write ('Var t(et)('I,'): ');
            Readln (VarI[I]);
        End;
    End;
End: [ Termina Procedure Lee Datos ... ]

```

```

{XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX}
{XXXXXXXXXXXXXXXXXXXX Comienza Programa Principal XXXXXXXXXXXXXXXXXXXXXXXXXXXX}
{XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX}

```

```

Begin
    Write ('Digite la Lambda a utilizar: ');
    Readln (Lambda);
    Write ('Digite numero de Periodos por Pronosticar:');
    Readln (H);
    Write ('Digite la Serie Zt : ');
    Readln (Serie);
    Write ('Digite el Origen t : ');
    Readln (t);
    LeeDatos;
    Constante;
    Optimo;
    Intervalo;
    Repeat
        Imprime;
        Write ('Imprimir Otra Vez?');
        Readln (Texto);
    Until (Texto <> 'SI') and (Texto <> 'si') and (Texto <> 'Si');

```

End.

BIBLIOGRAFIA

- Andrews, D.F. (1971)
 "A Note on the Selection of Data Transformations"
Biometrika, 58, pag. 249-254
- Ansley, C.F., Spivey, W.A. y Wroblewski W.J. (1977)
 "A Class of Transformations for Box-Jenkins Seasonal Models"
Appl. Statist. 26, No.2, pag. 173
- Bartlett, M.S. (1947)
 "The Use of Transformations"
Biometrics 3, pag. 37-52
- Bowerman, B.L. y D'Connell, R.T. (1987)
"Time Series Forecasting, Unified Concepts and Computer Implementation",
 Second Edition, Duxbury Press.
- Box, G.E.P. y Cox, D. R. (1964)
 "An Analysis of Transformations"
Jour. Royal Statist. Society B26, pag. 211-252
- Box, G.E.P. y Jenkins, G.M. (1970)
"Time Series Analysis, Forecasting and Control"
 Holden Day, San Francisco.
- Box, G.E.P. y Jenkins, G.M. (1973)
 "Some Comments on a Paper by Chatfield and Prothero and on a Review by Kendall"
J. R. Statist. Soc. A., 135, Part 3, pag. 337
- Carroll, R.J. y Rupert, D. (1981)
 "On Prediction and the Power Transformation Family"
Biometrika, Vol 68, No. 3, pag. 609-615
- Chatfield, C. (1975)
"The Analysis of Time Series: Theory and Practice"
 Chapman and Hall, London.
- Chatfield, C. y Prothero, D.L. (1973)
 "Box-Jenkins Seasonal Forecasting: Problems in a Case-Study"
J. R. Statist. Soc. A., 136, Part 3, pag. 295
- Dolby, James L. (1963)
 "A Quick Method for Choosing a Transformation"
Technometrics, Vol. 5, No.3
- Draper, N.R. y Hunter, W.G. (1969)
 "Transformations: Some Examples Revisited"
Technometrics Vol. 11, No. 1

- Granger, C.W.J. y Morris, M.J. (1976)
 "Time Series Modelling and Interpretation"
Journal of the Royal Statistical Society, a-139, 246
- Granger, C.W.J. y Newbold, P. (1976)
 "Forecasting Transformed Series"
Jour. Royal Statistical Society, Ser. B, 38, pag 189-203.
- Guerrero, Victor M. (1983)
"Análisis Estadístico de Series de Tiempo Económicas"
 Libro no-publicado, disponible en fotocopias.
 Mexico.
- Guerrero, Victor M. (1983)
 "Desestacionalización de Series de Tiempo Económicas: Parte
 1, Una Introducción a la Metodología."
Serie Documentos de Investigación, Banco de México No.54.
- Harvey, A.C. (1981)
"Time Series Models"
 Philip Allan Published Limited.
 London.
- Hopwood, W.S., Mckeown, J.C. y Newbold, P. (1981)
 "Power Transformations in Time Series Models of Quarterly
 Earnings per Share"
The Accounting Review, Vol. LV1, No.4, pag. 927
- Hopwood, W.S. . Mckeown, J.C. y Newbold, P. (1984)
 "Time Series Forecasting Models Involving Power
 Transformations"
Journal of Forecasting vol.3 pag. 57-61
- Ljung, G.M. y Box, G.E.P. (1978)
 "On a measure of lack of fit in time series models"
Biometrika, 65, 249
- Makridakis, S. (1976)
 "A Survey of Time Series"
Int. Stat. Rev., Vol.44, No. 1, pag. 29-70
- Makridakis, S. (1978)
 "Time Series Analysis: An Update and Evaluation"
Int. Stat. Rev., Vol.46, pag. 255-278
- Makridakis, S. y Wheelwright (1978)
"Forecasting Methods and Applications"
 Wiley/Hamilton, New York
- McCleary, R. y Hay, R.A. Jr. (1980)
 "Applied Time Series Analysis for the Social Science"
SAGE Publications

- Miller, D.M. (1984)
"Reducing Transformation Bias in Curve Fitting"
The American Statistician Vol.38, No.2, pag. 124
- Nelson, Ch.R. (1973)
"Applied Time Series Analysis for Managerial Forecasting"
Holden Day, San Francisco.
- Nelson, H.L. Jr. y Granger, C.W.J. (1979)
"Experience with Using the Box-Cox Transformation when
Forecasting Economic Time Series"
Journal of Econometrics Vol. 10, pag. 57-69
- Neyman, J. y Scott, E. (1960)
"Correction for Bias Introduced by a Transformation of
Variables"
Annals of Mathematical Statistics, 31, pag.643-655
- Pankratz, A. y Underwood, D. (1987)
"Forecasts of Power Transformed Series"
Journal of Forecasting 6, 239-248
- Pierce, D.A. (1979)
"R² Measure for Time Series"
Journal of the American Statistical Association 74, 901
- Schmetterer, L. (1960)
"On a Problem of J.Neyman and E. Scott"
Annals of Mathematical Statistics, 31, pag.656-661
- Taylor, M.G. (1986)
"The Retransformed Mean After a Fitted Power Transformaton"
Journal American Statist. Assoc., Vol.81, No.393, pag.114