

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE INGENIERIA AGRICOLA



**“MODELOS EN SERIES DE TIEMPO
APLICADO A DESCARGAS MEDIAS MENSUALES
DEL RIO PISCO”**

Tesis presentada por:

ESAUL OBREGÓN PARRAGA

**Para optar el título de
INGENIERO AGRICOLA**

LIMA - PERU

1993

CONTENIDO

	Pág.
i. Presentación	
ii. Resumen	i
iii. Lista de símbolos y siglas usadas	ii
iv. Lista de figuras y tablas	v
1.- Introducción.	4
2.- Revisión de Literatura	6
2.1 Conceptos Básicos y Antecedentes	6
2.2 Proceso estocástico	9
2.3 Función de transferencia	9
2.4 Modelos lineales	12
3.- Modelos Auto-regresivos de media móvil (ARMA)	
3.1 Generalidades	15
3.2 Modelos Autorregresivos (AR) y de medias móviles (MA)	
3.2.1 Modelos Auto-regresivos (AR)	17
3.2.2 Modelos de Media Móvil (MA)	21
3.3 Propiedades del modelo ARMA	
3.3.1 Función de autocorrelación y autocovariancia	24
3.3.2 Ecuación de Yule-Walker	26
3.3.3 Relación entre los parámetros ARMA y la función de autocovariancia (FA).	28
3.3.4 Relación de los parámetros AR con predicción lineal.	30
3.3.5 Solución recursiva para la ecuación de Yule- Walker y la función de autocorrelación parcial (FAP)	34

3.3.6	Coeficiente de reflexión.	36
3.3.7	Condiciones de estabilidad y estacionariedad	37
3.4	Identificación del modelo	38
3.5	Estimación de los parámetros	45
3.5.1	Estimación de los parámetros autoregresivos.	45
3.5.2	Estimación de los parámetros del Modelo MA.	48
3.5.3	Estimación preliminar de los parámetros del Modelo ARMA(p,q)	49
3.5.4	Estimación de parámetros por máxima verosimilitud para los modelos ARMA(p,q).	52
3.5.5	Error estandar de los parámetros estimados.	55
3.6	Modelos no estacionarios y estacionales.	
3.6.1	Modelos no estacionarios	56
3.6.2	Modelos estacionales - SARMA(p,q)x(P,Q).	58
3.7	Diagnóstico y revisión del modelo.	71
3.7.1	Obtención de los residuos	71
3.7.2	Uso de los residuos para modificar el modelo	73
3.7.3	Análisis del periodograma	74
3.7.4	Prueba de Box y Pierce	80
4.-	Modelos Auto-regresivos integrados de media móvil (ARIMA)	
4.1	Series no estacionarias	81
4.2	Formas del modelo ARIMA	87
4.3	Término constante del modelo	89
4.4	Modelos estacionales.	90
4.5	Predicción con modelos ARIMA	95
4.5.1	Cálculo de la predicción del error cuadrado medio mínimo.	96

4.5.2 Formas básicas de la predicción y ecuaciones de pronóstico.	98
4.5.3 Actualización de las predicciones.	99
4.5.4 Intervalos de confianza.	100
5.- Materiales y Métodos	
5.1 Materiales y equipos	103
5.2 Métodos y procedimientos	104
6.- Resultados y Discusiones	113
7.- Conclusiones y Recomendaciones	159
7.1 Conclusiones	159
7.2 Recomendaciones	160
8.- Bibliografía	161

ANEXOS

- A. Ejemplos sobre Estacionariedad, Invertibilidad, Autocovariancia y Autocorrelación para algunos Modelos
- B. Algunos comandos y usos del Software FORECAST PLUS y TSP (Time Series Processor)
- C. Predicciones e Intervalos de Confianza (95%) para el ajuste del Modelo y predicciones hasta agosto de 1993.

RESUMEN

Se realizó el estudio de las descargas medias mensuales (m^3/s) del río Pisco para encontrar modelos de predicción utilizando la metodología de los modelos de Box y Jenkins, en base a la información del período de 1974-88. Presento a los modelos ARMA y ARIMA ordinarios y estacionales, de manera explicativa, contando así con una base teórica para continuar trabajando en este campo de Investigación y en Recursos de Agua y Tierra.

Se analizó la serie y se efectuó una transformación no lineal según Box y Cox, encontrándose que la transformación más adecuada es del tipo logarítmica. La serie es diferenciada en relación a un atraso unitario estacional ($D=1, s=12$); de esta manera encontrándose una serie homogénea y estacionaria.

Se determinó los modelos más apropiados para las descargas mensuales, al modelo: $SARIMA(2,0,1) \times (1,1,1)_{12}$ aceptando a los residuos como ruido blanco con 85% de probabilidad según la prueba de Box y Pierce y el modelo $SARIMA(1,0,1) \times (2,1,1)_{12}$ con 73,9 % de aceptar a los residuos como ruido blanco;

Estos modelos estudiados en el presente trabajo y utilizando la misma metodología efectuar para otros ríos y contar con descargas mensuales con cierto período de anticipación para operación de sistemas de riego, planificación y programación de cultivo, principalmente.

LISTA DE SIMBOLOS

- a_t : Serie temporal ruido blanco o "Shock"
- b : Parámetro en T.N.L. de Box y Cox.
- A_t : Vector ruido blanco (S.T.)
- d : Orden de diferenciación de una serie temporal, cuyo modelo integrado: ARIMA(p,d,q).
- D : Orden de diferenciación de la serie temporal; cuyo modelo es estacional de período s .
- e_t : Residuo de la predicción.
- B : Operador atraso, "backward shift operator"
- c_k : Estimado de γ_k ; o autocovariancia.
- k : Número de clases en consideración;
- m : Orden del predictor lineal unidimensional.
- N : Número de observaciones o muestras de un proceso estocástico.
- p : Orden de un modelo AR ó la parte AR de un modelo ARIMA;
- P : Orden de la parte AR estacional de un modelo ARMA estacional;
- q : Orden de un modelo MA, ó parte MA de un modelo ARIMA;
- Q : Orden de la parte MA estacional de un modelo ARMA estacional;
- r_k : Estimador para ρ_k , ó coeficiente de autocorrelación;
- Z_t : Serie temporal, secuencia de valores en el tiempo;
- Φ : Parámetro autorregresivo estacional de un modelo ARMA estacional;
- ϕ : Parámetro autorregresivo de un modelo ARMA univariado;
- Φ : Matriz de coeficientes regresivos en modelos multivariados;

- γ_k : Función de autocovariancia (FA) para un proceso estocástico estacionario;
- Γ : Matriz de covariancia para un proceso multivariado;
- λ : Correlación cruzada entre la entrada y salida de un filtro - modelo ARMA;
- μ : Media de un proceso estocástico estacionario;
- y_t : Muestra de un proceso estocástico Z_t ;
- π : Parámetro de una expansión AR infinita, de un proceso ARMA;
- w_t : Coeficiente de expansión de un modelo ARMA en términos de un modelo MA de orden infinito (Función de transferencia de filtro);
- $\rho_k = \rho_{-k}$: Función coeficiente de autocorrelación (FAC) de proceso estocástico estacionario;
- σ_x : Desviación estándar de la v.a. x
- θ : Parámetro de media móvil de un modelo ARMA univariado;
- δ : Parámetro de media móvil estacional de un modelo ARMA estacional;

Indices superiores

- T : indica transpuesta
+ : indica pseudo inversa

Indices inferiores

- t : tiempo
k : atraso o "lag", intervalo entre 2 instantes de tiempo,
ij: indica el coeficiente bidimensional o de componente de matrices.

LISTA DE SIGLAS Y ABREVIATURAS

- AED : Análisis exploratorio de datos;
- AR : Autorregresivo;
- ARMA: Modelo autorregresivo y media móvil (Autorregressive moving average);
- FA : Función de autocovariancia;
- FAC : Función de autocorrelación;
- FAP : Función de autocorrelación parcial;
- FAR : Función de autocorrelación de residuos;
- GL : Grados de libertad
- GN : Gauss-Newton
- MA : Media móvil;
- SECMP: Suma de Cuadrados de los errores de predicción;
- S.T.: Serie temporal;
- SUM : Suma
- T.N.L.: Transformación no lineal
- v.a.: Variable aleatoria;
- YW : Yule - Walker; (ecuación)
- YW-2D: Yule - Walker Ecuación de 2 dimensiones.

LISTA DE FIGURAS

Nº		Pág.
1.	Filtro lineal, y función de transferencia;	10
2.	Función de autocorrelación para modelos AR(1) y AR(2);	19
3.	Región de estacionariedad para un modelo AR(2);	20
4.	FAC para un modelo MA(1);	22
5.	Región de estacionariedad para un modelo MA(2);	23
6.	FAC y FAP de 6 modelos ARMA(1,1) estacionarios e invertibles;	42
7.	Función de autocorrelación de modelo AR(1) y AR(1) ₄ ;	62
8.	Espectro y espectro acumulado del ruido blanco;	77
9.	Períodograma acumulado normalizado VS. frecuencias para un ruido blanco;	77
10.	Períodograma acumulado de una serie y límites de confianza;	79
11.	Amplitud VS. media de la serie, mostrando algunos valores de b, para una transformación no lineal;	86
12.	Observación de una serie temporal con predicción de origen t y horizonte h;	95
13.	Curvas con la misma Suma cuadrado (SQ) para un modelo ARMA (1,1);	110
14.	Descargas medias mensuales, datos originales, río Pisco;	114
15.	Serie suavizada por: suavizamiento robusto 4253HT;	116
16.	Ploteo de cajas - serie original (caudal);	117
17.	Ploteo de cajas agregadas - serie original;	119
18.	Ploteo nivel VS. dispersión de los datos originales;	121
19.	Ploteo nivel VS. dispersión de los datos transformados - Ln(caudal);	122

20. Ploteo de la serie transformada - $\ln(\text{caudal})=Z_t$;	123
21. Ploteo de caja - serie transformada;	125
22. Ploteo de caja agregada - serie transformada;	126
23. Función de autocorrelación y autocorrelación parcial de la serie original;	129
24. FAC Y FAP de la serie transformada;	130
25. Periodograma y periodograma acumulado (Z_t)	131
26. FAC Y FAP para Z_t y $D=1$ (diferencia de orden 12) estacionaria;	133
27. Función de autocorrelación de Residuos (FAR) para el modelo SARIMA(1,0,1)x(2,1,1) ₁₂ ;	135
28. FAR para el modelo SARIMA (1,0,1)x(1,1,1) ₁₂ ;	138
29. FAR para el modelo SARIMA (1,0,0)x(1,1,1) ₁₂ ;	141
30. FAR para el modelo SARIMA (2,0,1)x(1,1,1) ₁₂ ;	143
31. FAR para el modelo SARIMA (2,0,1)x(2,1,1) ₁₂ ;	145
32. FAR para el modelo SARIMA (1,0,1)x(2,1,1) ₁₂ ; estimado por el Micro TSP.	149
33. Pronósticos para set.87 a ago.88. Modelos SARIMA(2,0,1)x(1,1,1) ₁₂ y SARIMA(1,0,1)x(2,1,1) ₁₂	157
34. Pronósticos de Descargas medias mensuales hasta Ago.93 Modelo: SARIMA(2,0,1)x(1,1,1) ₁₂	158

LISTA DE TABLAS

1. FAC y FAP en modelos AR(p), MA(q) y ARMA(p,q);	41
2. FAC en un modelo AR(1) y un AR(1) _s ;	61
3. FAC y FAP; para modelos AR(p)xAR(P) _s y MA(q)xMA(Q) _s ;	68
4. Coeficientes para los cálculos de límites de confianza para el periodograma acumulado;	78
5. Descargas medias mensuales (m ³ /s) río Pisco, años de set. 74 a Ago. 88	104
6. Comportamiento de la FAC y FAP de un proceso ARIMA(p,d,q);	108
7. Varianzas aproximadas para algunos modelos;	109
8. Varianzas aproximadas para los parámetros de los modelos más comunes;	111
9. Observaciones mensuales para la serie en estudio (set.74 - ago.86);	113
10. Variancias anuales de la serie transformada;	127
11. Estimación de parámetros y estadísticos para el modelo SARIMA(1,0,1)x(2,1,1) ₁₂	137
12. Estimación de parámetros y estadísticos para el modelo SARIMA(1,0,1)x(1,1,1) ₁₂	137
13. Estimación de parámetros y estadísticos para el modelo SARIMA(1,0,0)x(1,1,1) ₁₂	140
14. Estimación de parámetros y estadísticos para el modelo SARIMA(2,0,1)x(1,1,1) ₁₂	140
15. Estimación de parámetros y estadísticos para el modelo SARIMA(2,0,1)x(2,1,1) ₁₂	144
16. Estimación de parámetros mediante el uso del TSP para el modelo SARIMA(1,0,1)x(2,1,1) ₁₂	148
17. Predicción para orígenes 156, 132 y 120;	152
18. Intervalos de confianza al 95% y los pronósticos;	154
19. Errores de la predicción set.87 - ago.88.	155
20. Suma de Cuadrados de los errores de predicción.	155

I INTRODUCCION

Los recursos hídricos no controlados pueden presentar consecuencias desventajosas, daños por inundaciones, sequías, mala planificación en los sistemas de riego y planificación de cultivo; ocurriendo pérdidas económicas y de orden social y ambiental considerables, por lo tanto es preocupante.

Las predicciones hidrológicas permiten un mayor racionamiento de las políticas operacionales de corto y mediano plazo de los recursos hídricos, dando buenos resultados.

Una de las características desventajosas en los modelos predictivos es la posibilidad de una continua corrección de las predicciones a partir de las observaciones previas, llamado recalibración.

En el presente trabajo se describe la metodología para encontrar el modelo adecuado para las descargas medias mensuales del río Pisco para ello en la sección 3 se menciona el desarrollo para un modelo ARMA y en la sección 4 para un modelo ARIMA; como podremos ver que el modelo ARIMA presentado recurre al procedimiento del modelo ARMA cuando la serie tiene ciertas características como estacionariedad e invertibilidad.

La selección de un modelo adecuado no solo es aquel que estadísticamente presenta un buen coeficiente de aceptación, sino aquel que sea el más representativo del comportamiento de

las descargas medias mensuales para cualquier ciclo hidrológico y presentando con un intervalo de confianza de 95% para tener un buen pronóstico. Así mismo se verifica el modelo con los datos del último año (1987-88).

En el capítulo 3 y 4 se describe la base teórica de los modelos de Box y Jenkins para poder ser aplicados en la parte de metodología y obtención de resultados.

OBJETIVOS

Los Objetivos del presente trabajo son:

- 1.- Determinar la serie de las descargas medias mensuales del río Pisco con fines de predicción;
- 2.- Encontrar el o los modelos de series temporales para predecir las descargas medias mensuales (m^3/s) del río Pisco.
- 3.- Analizar los modelos y pronósticos obtenidos.

II REVISION DE LITERATURA

2.1 CONCEPTOS BASICOS Y ANTECEDENTES

Es importante mencionar que el tratamiento de series temporales tiene una larga historia como herramienta de predicción alternativa al análisis estructural característico de los modelos econométricos.

El enfoque tradicional es muy conocido; como la descomposición temporal de la serie en diferentes partes que se consideran relevantes y que básicamente se identifica como:

- 1) Tendencia.- Momento de larga duración que se mantiene durante el periodo de observación.
- 2) Variación Estacional.- Movimiento que se produce, dentro de un periodo anual, inicialmente por motivos climáticos y con base en estos, por motivos económicos y de organización social (recolecciones, vacaciones, etc.)
- 3) Movimientos cíclicos.- Oscilaciones alrededor de la tendencia producido por periodos alternativos de prosperidad y depresión (mayor de un año).
- 4) Movimientos Accidentales ó Irregulares.- Oscilaciones erráticas que aún quedan en la serie original tras eliminar los tres anteriores componentes.

Escribir un modelo de series en base a los datos que represente el comportamiento del proceso que lo originó y

brindar pronósticos aceptables de su comportamiento futuro.

Hidrológicamente se puede decir que una predicción de ocurrencia o no de un suceso, en función del tiempo muy próximo se le conoce como Prognóstico.

El inicio del enfoque remonta los trabajos de Yule y Markov (1910), que plantearon modelos autoregresivos.

Según PULIDO (1980); en los libros de Granger (1964) y Fishman (1969) en la línea de análisis espectral atrajeron la atención de los económetros sobre la posibilidad del análisis de series en el dominio de frecuencias, concepto de una serie para diferentes frecuencias.

Pero la obra de los estadísticos George Box y Guilym Jenkins (1970) quienes han cimentado el interés por el tratamiento de las series cronológicas.

Una clase importante de los modelos estocásticos son los llamados modelos estacionarios, los cuales asumen que el proceso permanece en equilibrio en torno a un cierto nivel medio constante.

La descripción de los modelos de Box & Jenkins, tanto estacionarios, y no estacionarios homogéneos y los estacionales serán descritas en los capítulos 3 y 4; siguiendo una metodología adecuada, para poder hacer uso de los Software como el "Forecast Plus", T.S.P., etc. y

lograr el modelo adecuado que nos representa el proceso que la originó y la serie hidrológica modelada para los caudales medios mensuales; así mismo nos permita pronosticar caudales futuros con cierto intervalo de confianza y que describa el comportamiento futuro de la serie.

Trabajos en modelos de series hidrológicas en nuestro país no se han desarrollado aún a la fecha dentro del campo de recursos hídricos. Sin embargo citamos algunos trabajos relacionados.

TORRES S., (1979), encontró un modelo ajustado para simular la serie de volúmenes de descargas anuales del río Chicama con el modelo Autorregresivo de tercer orden AR(3) y uso los modelos de Fourier y modelos multiplicativos de Box y Jenkins para el caso de volúmenes mensuales del mismo río; no así llegando a resultados satisfactorios con el último modelo por falta de análisis.

SOTOMAYOR, (1989) describe el modelo ARIMA, utilizando datos de precipitación mensual de San Jerónimo - Cusco; mas no concluye cual es mejor modelo para dicha variable y explica parte de la metodología del modelo ARIMA.

VELASQUEZ B., (1982) utilizando análisis de serie de Fourier de los parámetros periódicos, encontró que los caudales medios mensuales se pueden modelar con una componente periódica determinística.

Trabajos en otros países, como en el Brasil, si se

han desarrollado con más amplitud como de Boris A. y Lanna A. (1982) para dimensionamiento y capacidad del reservorio de Passo Real del río Jacui; Genovez A. y Chaudhry F. (1983) encontraron Modelos SARIMA para valores mensuales del río Teite de Cumbica-Brasil.

2.2 PROCESO ESTOCASTICO

Interpretado como una familia de v.a. (Z_t) , definidas sobre el mismo espacio de probabilidad.

La base fundamental del análisis de series de tiempo es considerar la secuencia de observaciones obtenidas como la realización de variables aleatorias conjuntamente distribuidas, esto es la serie Z_1, Z_2, \dots, Z_N ; tomando intervalos de tiempos discretos e igualmente espaciados 1, 2, ... N es considerada extraída de una distribución de probabilidad conjunta $P_1, \dots, P_N(Z_1, Z_2, \dots, Z_N)$.

de donde $p(\cdot)$ es la función de densidad de probabilidad conjunta.

Los subíndices, 1, 2, ..., N, de p indican que la distribución está asociada a estos periodos de tiempo y las variables en cuestión son Z_1, Z_2, \dots, Z_N .

2.3 FUNCION DE TRANSFERENCIA DEL FILTRO.

El modelo de una serie temporal es generado a través de un filtro lineal (o un sistema lineal), cuya entrada es el ruido blanco ó proceso de perturbación Puro (elemento fundamental). Como muestra en la Fig. N° 01.

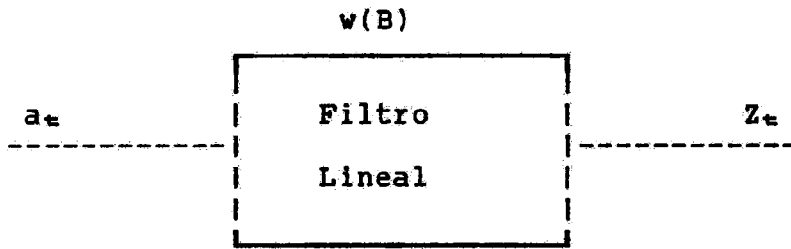


Fig. N^o 01: Filtro lineal, entrada a_t , salida Z_t y $w(B)$ es función de transferencia.

$$Z_t = a_t + w_1 a_{t-1} + \dots +$$

Hay muchos tipos de procesos lineales que dependen de cuantos términos uso para formar Z_t , finalmente tenemos:

$$\begin{aligned} Z_t &= \mu + a_t + w_1 a_{t-1} + w_2 a_{t-2} + \dots \\ &= \mu + w(B)a_t \end{aligned} \tag{2.1}$$

donde:

$$w(B) = 1 + w_1 B + w_2 B^2 + w_3 B^3 + \dots \tag{2.2}$$

denominado función de transferencia de filtro

μ : es un parámetro que define el nivel de la serie.

Z_t : proceso lineal (discreto); donde a_t :

$$\begin{aligned} E(a_t) &= 0 && \forall t \\ \text{Var}(a_t) &= \sigma^2_a && \forall t \\ E(a_t a_s) &= 0 && s \neq t \end{aligned}$$

Si llamamos $\bar{Z}_t = Z_t - \mu$ tenemos que (1)

$$\bar{Z}_t = w(B)a_t \tag{2.3}$$

Siendo la secuencia de pesos $\{w_j ; j \geq 1\}$ finita o infinita pero convergente, el filtro es estable y \bar{Z}_t es estacionario. En este caso μ es la media del proceso; caso contrario Z_t es no estacionario y μ no tiene significado específico, a no ser un punto de referencia para el nivel de la serie.

de (2.1) tomando esperanzas tenemos que:

$$E(Z_t) = \mu + E(a_t + \sum_{j=1}^{\infty} w_j a_{t-j})$$

como $E(a_t) = 0 \forall t$; tenemos $E(Z_t) = \mu$, osea la serie $\sum_{j=1}^{\infty} w_j$ converge; de la misma manera la variancia de \bar{Z}_t y dado por:

$$\text{Var}(Z_t) = \sigma^2 \sum_{j=1}^{\infty} w_j^2 \quad (2.4)$$

con $w_0 = 1$; $\sum_{j=1}^{\infty} w_j < \infty$ y la función de autocovariancia γ_j dada por:

$$\gamma_j = \sigma^2 \sum_{j=1}^{\infty} w_j w_{j+1} \quad (2.5)$$

se observa que (2.4) es obtenido de (2.5) para $j=0$, así $\gamma_0 = \text{Var}(\bar{Z}_t)$. Tenemos que la media y la variancia de \bar{Z}_t son constantes y la covariancia depende del atraso j , luego \bar{Z}_t es estacionario. Podemos escribir \bar{Z}_t en forma alternativa como una suma ponderada de valores pasados Z_{t-1} , Z_{t-2} más un ruido blanco a_t :

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j Z_{t-j} + a_t \quad (2.6)$$

despejando:

$$(1 - \sum_{j=1}^{\infty} \pi_j B^j) Z_t = a_t.$$

$$\pi(B) \bar{Z}_t = a_t \quad (2.7)$$

donde $\pi(B)$ es el operador:

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots \quad (2.8)$$

de (2.7) y (2.3) tenemos:

$\pi(B)w(B)a_e = a_e$; de modo que:

$$\pi(B) = w^{-1}(B) \quad (2.9)$$

esta relación puede ser usada para obtener los pesos π_j en función de los pesos w_j y viceversa.

Estacionariedad e Invertibilidad

Un proceso lineal será estacionario si la serie $w(B)$ converge para $|B| \leq 1$ e invertible si $\pi(B)$ converge para $|B| \leq 1$ (Box & Jenkins, 1976)

2.4 MODELOS LINEALES

Los modelos en estudio son llamados modelo lineal estacionarios, ellos son: modelos autorregresivos (AR), de medias móviles (MA) y mixtos (ARMA) y en particular los modelos no estacionarios (ARIMA).

Proceso Lineal

Estos procesos se caracterizan porque se pueden representar como una combinación lineal de variables aleatorias. Estos pueden ser procesos puramente aleatorios, procesos autorregresivos, procesos de media móvil y los obtenidos de combinaciones de estos dos últimos.

El proceso puramente aleatorio es el más simple de todos, expresados de la siguiente forma:

$$Z_t = a_t$$

cuyas propiedades son las definidas en la sec. 2.2 y que en adelante designaremos a a_t unicamente a una variable aleatoria. Variable aleatoria del tipo a_t no haria falta la teoria de los procesos estocásticos.

Esta afirmación es cierta. Sin embargo el "ruido blanco" es una pieza clave en la construcción de procesos estocásticos más complicados como son los AR y MA, etc.

Un proceso autorregresivo de orden p , utilizando la notación usual, un proceso AR(p) se expresa de la siguiente forma:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (2.10)$$

Algunas características de los modelos AR(1), AR(2) y generalizando para un AR(p) se tiene:

para un modelo AR(1) definido por:

$$Z_t = \phi_1 Z_{t-1} + a_t \quad (2.11)$$

para que sea estacionario las raíces del polinomio $1 - \phi_1 B = 0$ deben caer fuera del circulo unitario es decir $|B| = |1/\phi_1| > 1$ lo que equivale a $|\phi_1| < 1$.

Si el proceso se considera iniciado en $-\infty$, entonces para $|\phi_1| < 1$ y cualquiera que sea el valor inicial se verifica que $E[Z_t] = 0$ si al modelo 2.11 se incluye una constante, se tendrá:

$$Z_t = \varphi + \phi_1 Z_{t-1} + a_t \quad (2.12)$$

Entonces si el proceso se inicia en $-\infty$ y es estacionario, se verifica que la media del proceso será constante para

cualquier valor de t .

$E[Z_t] = E[Z_{t-1}] = \mu$; tomando esperanzas a 2.12 y teniendo en cuenta lo anterior se verifica que:

$$E[Z_t] = \mu = g/(1 - \phi_1).$$

Para un proceso de media móvil de orden q o un proceso MA(q) viene dado por:

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.13)$$

La expresión 2.13 de medias móviles hace referencia a que la variable Z_t , se obtiene como un promedio de variables de "ruido blanco" (en este caso $q+1$), siendo los θ_i coeficientes de ponderación. Como las variables que forman parte de este promedio varía a lo largo del tiempo, reciben el apelativo de móviles.

Estos procesos, así como la combinación de un proceso AR(p) y un proceso MA(q) se obtiene un proceso ARMA(p,q), donde p indica el retardo máximo de la parte regresiva y q señala el correspondiente a la parte de medias móviles. la que es presentada fundamentalmente en la obra de Box y Jenkins (1970) y que se describen en los siguientes capítulos.

III MODELO AUTO-REGRESIVO DE MEDIA MOVIL (ARMA)

3.1 GENERALIDADES

Primeramente vamos introducir algunas nociones de operadores que será usada frecuentemente en el presente trabajo. La familiarización con esta notación facilitará bastante al manipuleo de los modelos en estudio.

Estos Operadores son:

- 1) Operador de traslación para el pasado, denotado por B y definido por:

$$BZ_t = Z_{t-1} ; B^p Z_t = Z_{t-p}$$

- 2) Operador de traslación para el futuro, denotado por F, y definido por:

$$FZ_t = Z_{t+1} ; F^p Z_t = Z_{t+p}$$

- 3) Operador diferencia; denotado por Δ definido por:

$$\Delta Z_t = Z_t - Z_{t-1} = (1-B)Z_t$$

$$\text{entonces } \Delta = 1 - B$$

- 4) Operador suma denotado por S definido por:

$$\begin{aligned} SZ_t &= \sum_{j=-1}^{\infty} Z_{t-j} = Z_t + Z_{t-1} + Z_{t-2} + \dots \\ &= (1 + B + B^2 + B^3 + \dots)Z_t \end{aligned}$$

de allí se reduce a:

$$SZ_t = (1 - B)^{-1}Z_t = \Delta^{-1}Z_t ; \text{ entonces } S = \Delta^{-1}$$

por el desarrollo de $1/(1-B) = 1 + B + B^2 + B^3 + \dots$

Según (Box & Jenkins, 1976) los modelos de serie de tiempo; que usan datos medidos y generados de series independientes de "Shocks" a_t . Estos shocks son variables aleatorias de entera distribución, usualmente asumimos Normal, con Media cero y variancia σ^2_a (constante) y ausencia de correlación y se llama un modelo de filtro lineal. Este modelo supone que la serie temporal es generada a través de un filtro lineal (o un sistema lineal) cuya entrada es el ruido blanco a_t (White noise) proceso que se transforma a la salida Z_t por medio del filtro lineal $w(B)$ como se mostró en la Fig. N° 01 y en las Ecuaciones. (2.1, 2.2, 2.3 y 2.4).

La forma general del comportamiento del filtro puede ser escrito por un modelo ARMA (Auto-regressive móvil average) de la forma:

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \dots + \phi_p \bar{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.1)$$

$$\text{ó } \phi(B)\bar{Z}_t = \theta(B)a_t$$

donde:

$$\bar{Z}_t = Z_t - \mu ; \quad \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad \text{y} \quad \theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

La función $\phi(B)$ del proceso autoregresivo y $\theta(B)$ de media móvil, de orden p y q , respectivamente y otra forma de describir la ecuación 3.1 será:

$$\bar{Z}_t = \frac{\theta(B)}{\phi(B)} a_t = \frac{1 - \theta_1 B - \dots - \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} a_t \quad (3.2)$$

Tomando la transformada z del proceso que sigue la ec. 3.1 obtenemos: $(1-\phi_1z^{-1}-\dots-\phi_pz^{-p})Z(z)=(1-\theta_1z^{-1}-\dots-\theta_qz^{-q})Z(z)$ comparandose con 3.2 se tiene:

$$H(z) = Z(z) / A(z) = \theta(z^{-1}) / \phi(z^{-1}) \quad (3.2b)$$

Modelo que buscamos debe ser de tal manera que tenga el menor número de parámetros estimados posibles (condición de parsimonia)

3.2 MODELOS AUTORREGRESIVOS (AR) Y DE MEDIA MOVIL (MA)

3.2.1 MODELOS AUTO-REGRESIVOS

Si en la ecuación 2.6, $\pi_j = 0$, $j > p$, obtenemos un modelo auto-regresivo de orden p, que denotaremos por AR(p): (mencionado anteriormente 2.10).

$$\bar{Z}_t = \phi_1\bar{Z}_{t-1} + \phi_2\bar{Z}_{t-2} + \dots + \phi_p\bar{Z}_{t-p} + a_t \quad (3.3)$$

recordando los pesos de π_j para ϕ_j de acuerdo con la notación usual.

Fue definido el operador auto-regresivo estacionario de orden p.

$$\phi(B) = 1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p$$

de esto se escribió la ecuación : $\phi(B)\bar{Z}_t = a_t$.

Algunos ejemplos se mencionan en la parte de apéndice y así por ejemplo el caso más simple del modelo auto-regresivo de orden $p=1$, AR(1) de manera que Z_t depende apenas de Z_{t-1} , y del ruido blanco en el instante t. (Z_t es un proceso de Markov).

Como $\pi(B) = \phi(B) = 1 - \phi B$, el proceso es siempre invertible.

Sustituyendose, sucesivamente, \bar{Z}_{t-1} , \bar{Z}_{t-2} , etc. en 2.6 obtenemos:

$$Z_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots = \sum_{j=0}^{\infty} \phi^j a_{t-j}$$

o sea;

$$\bar{Z}_t = w(B)a_t = (1 + \phi B + \phi^2 B^2 + \dots)a_t$$

vemos entonces que

$$w(B) = \sum \phi^j B^j = [\phi(B)]^{-1} = (1 - \phi B)^{-1}$$

y de acuerdo con la proposición (2.3), el proceso será estacionario si $w(B)$ converge para $|B| \leq 1$. Si debemos tener $|\phi| < 1$. Como las raíces de la ecuación $\phi(B) = 1 - \phi B = 0$ y $B = \phi^{-1}$, esta condición es equivalente a decir que la raíz de $\phi(B)=0$ deben caer fuera del círculo unitario.

Función de Autocorrelación

En general la función de autocorrelación de un proceso autorregresivo consiste de una mezcla de amortiguamientos exponenciales y amortiguamientos de oscilación sinusoidal.

Algunos ejemplos de la función de autocorrelación se presentan en las figuras N^o 02 para algunos procesos autorregresivos:

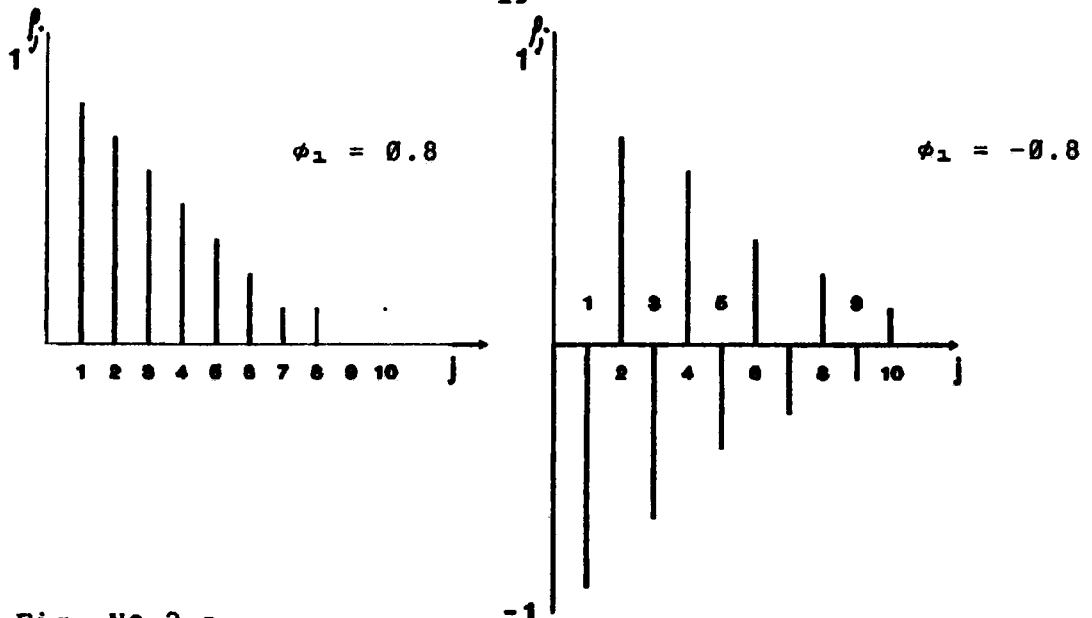
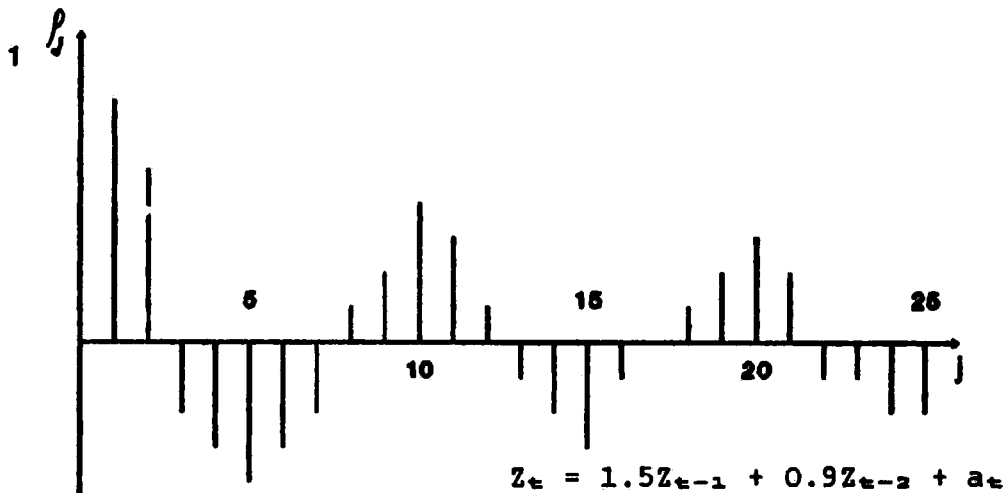


Fig. No 2.a Función de autocorrelación (FAC) para un proceso AR(1).



2.b Función de autocorrelación para un proceso AR(2)

Estacionariedad e Invertibilidad

Sean G_i^{-1} , $i=1, \dots, p$, las raíces de la ecuación característica $\phi(B) = 0$; entonces se pueden escribir:

$$\phi(B) = (1-G_1)(1-G_2)\dots(1-G_p)$$

y expandiendo en fracciones parciales;

$$w(B) = \phi^{-1}(B) = \sum_{i=1}^p \frac{A_i}{1-G_i B} \tag{3.4}$$

Si $w(B)$ debe converger para $|B| \leq 1$ debe tener $|G_i| < 1$, $i=1, \dots, p$. Esta condición es equivalente a la de la ecuación característica $\phi(B)=0$ tiene raíces fuera del círculo unitario. Esta es la condición de estacionariedad.

Si consideramos un modelo AR(2):

$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$; escrito de la forma $\phi(B)Z_t = a_t$, con $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$. se puede mostrar que Z_t es estacionario si (las raíces de $\phi(B)=0$ están fuera del círculo unitario).

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 &< \phi_2 < 1 \end{aligned} \quad (3.5)$$

En la Fig. N° 03 se presenta la región de estacionariedad para un modelo AR(2), donde se puede apreciar la región de raíces reales y las de complejas.

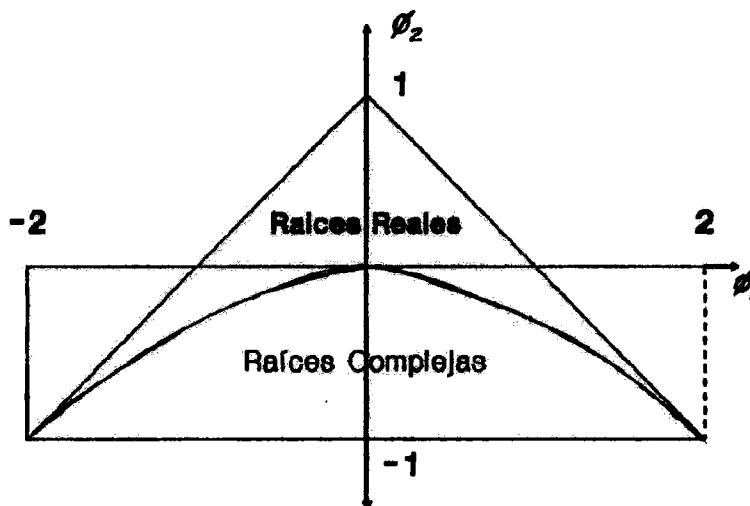


Fig. N° 03 Región de estacionariedad para un modelo AR(2)

3.3.2 MODELOS DE MEDIA MOVIL

Considerase el proceso lineal (2.3) y suponga que $w_j=0, j>q$; obtenemos el proceso de medias móviles de orden q , que denotaremos por $MA(q)$ (de "moving average"). Y ahora usaremos la notación:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.6)$$

y haciendo $\bar{Z}_t = Z_t - \mu$, tenemos

$$\bar{Z}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t \quad (3.7)$$

donde

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

que se ha definido anteriormente como el operador de media móvil de orden q .

La media móvil más simple es $MA(1)$,

$$\bar{Z}_t = a_t - \theta a_{t-1}$$

ó

$$\bar{Z}_t = (1 - \theta B) a_t,$$

de modo que $\theta(B)=1-\theta B$. Como $w(B)=1-\theta B$ es finito, el proceso es siempre estacionario de acuerdo con 2.3 como

$$a_t = [\theta(B)]^{-1} \bar{Z}_t = \frac{1}{1-\theta B} \bar{Z}_t = (1+\theta B+\theta^2 B^2+\dots) \bar{Z}_t$$

obtenemos la forma invertida:

$$\bar{Z}_t = a_t - \theta \bar{Z}_{t-1} - \theta^2 \bar{Z}_{t-2} - \dots$$

Si $|\theta| < 1$, o sea, la serie anterior $\pi(B)=\theta^{-1}(B)$ converge para $|B| \leq 1$. Esto equivale a decir que los ceros de $\theta(B)=1-\theta B=0$ están fuera del círculo unitario.

Función de Autocorrelación (FAC)

De la ecuación 3.7, tomando esperanzas y resolviendo se obtiene la FAC del proceso MA:

$$\rho_j = \frac{-\theta_j + \theta_1\theta_{j+1} + \theta_2\theta_{j+2} + \dots + \theta_{q-j}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} ; j=1,2,\dots,q$$

$$= 0 ; j>q \tag{3.8}$$

observandose; entonces que la FA de un proceso MA(q) es igual a cero para atrasos mayores que q, diferente a lo que acontece en un proceso AR, donde la función se extiende indefinidamente. En la Figura N^o 04 se presenta la FAC para procesos MA.

Estacionariedad e Invertibilidad

Dado que $w(B) = 1 - \theta_1B - \dots - \theta_qB^q$, no hay restricciones sobre los parámetros θ_j para que el proceso sea estacionario.

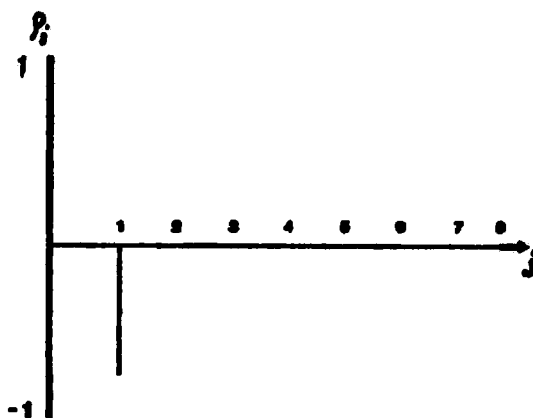


Fig. N^o 04. FAC para el modelo MA(1), $\theta=0,8$

Para un modelo MA(2) la condición de invertibilidad, implica que los parámetros del modelo deben caer en la región: como se muestra en la Fig. N° 05. equivalente a la estacionariedad en el modelo AR(2).

$$\theta_1 + \theta_2 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$-1 < \theta_2 < 1$$

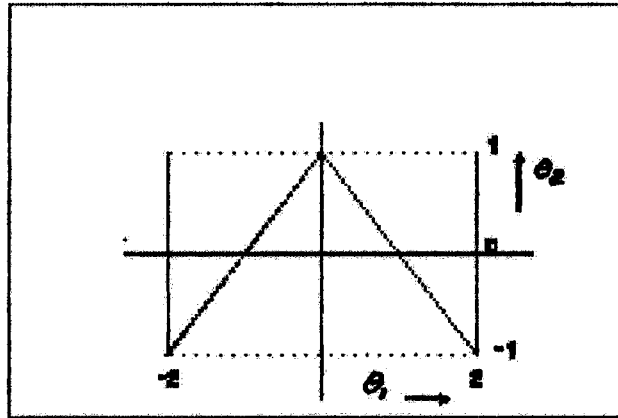


Fig. N° 05 Región de invertibilidad para un modelo MA(2).

Usando un argumento completamente similar al que fue visto para un modelo auto-regresivo AR(p), en caso de estacionariedad, podemos verificar que las condiciones de invertibilidad para un modelo MA(q) es que las raíces de la ecuación característica $\theta(B)=0$ están fuera del círculo unitario. En estas condiciones, un modelo MA(q) es equivalente a un modelo AR de orden infinito.

3.3 PROPIEDADES DEL MODELO ARMA

3.3.1 FUNCION DE AUTOCORRELACION Y AUTOCOVARIANCIA

Sea Z_t un proceso estocástico y la covariancia entre Z_t y sus muestras Z_{t+k} separada por k intervalos de tiempo y llamada autocovariancia y definida por:

$$\gamma_{t,t+k} = \text{cov}[Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)] \quad (3.9a)$$

En esta definición o valor de covariancia depende del tiempo t y del intervalo k . Un proceso estrictamente estacionario de distribución conjunta asociada con las observaciones $Z_{t1}, Z_{t2}, \dots, Z_{tm}$ es la misma que es asociada a $Z_{t1+k}, \dots, Z_{tm+k}$ para cualquier t . Una condición menos estricta de la condición de estacionariedad es cuando se somete los momentos ante cierto orden son considerados solo la función de retardo k . Así consideramos estacionariedad de segundo orden; de la Ec. 3.9a puede ser escrita:

$$\gamma_k = \text{cov}[Z_t, Z_{t+k}] \quad (3.9b)$$

La autocorrelación relativa a un retardo k , ρ_k (también llamado coeficiente de autocorrelación) puede ser definida como:

$$\rho_k = \gamma_k / \gamma_0 \quad (3.10)$$

γ_k y ρ_k considerados como función de k , son llamados función de autocovariancia (FA) y función de

autocorrelación (FAC) respectivamente. Para la estimación de la función de autocorrelación cuando se tiene un conjunto de N observaciones, Box & Jenkins (1970) presenta las siguientes expresiones:

$$r_k = C_k/C_0 \quad (3.11)$$

donde:

$$C_k = \frac{1}{N} \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})$$

$$C_{-k} = C_k,$$

de donde:

$$\bar{Z} = \frac{1}{N} \sum_{t=1}^{N-k} Z_t$$

Esta estimación es polarizada y es preferible la estimación no polarizada donde la suma de los productos es dividida por $(N-k-1)$ porque en general la variancia del estimador es menor Marple (1987, p.148), es conducir a la matriz de autocorrelación siempre definido-positivo. La estimación polarizada cuando usan las ecuaciones de Yule-Walker corresponden al método de autocorrelación para resolución de problemas de predicción lineal (Sec. 3.3.4) y es equivalente a extender el límite superior de la sumatoria de la expresión 3.11 para N , es considerar los datos que faltan como ceros. Por eso el método es también llamado abierto ("JANELADO"). La estimación no polarizada corresponde al método de covariancia o no "JANELADO". Esas

denominaciones para los métodos son utilizadas por razones históricas y no tienen relación con significado estadístico.

3.3.2 ECUACION DE YULE-WALKER

Considerandose el modelo AR(p), (p es el orden del modelo) obtenido a partir de la Ec. (2.4)

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \dots + \phi_p \bar{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.1)$$

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \dots + \phi_p \bar{Z}_{t-p} + a_t$$

multiplicando ambos miembros por Z_{t+k} y tomando el valor esperado deducimos la ecuación:

$$Y_k = \phi_1 Y_{k-1} + \dots + \phi_p Y_{k-p} \quad k > 0$$

dividiendo todo por Y_0 :

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (3.12)$$

si sustituimos $k = 1, 2, \dots, p$ en la ecuación 3.12, se obtiene un conjunto de ecuaciones lineales para ϕ_1, \dots, ϕ_p en terminos de ρ_1, \dots, ρ_p :

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 \dots + \phi_p \rho_{p-2} \\ &\vdots \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \dots + \phi_p \end{aligned}$$

o en forma matricial:

$$\begin{aligned}
 \underline{\phi} &= \begin{vmatrix} \phi_1 \\ \cdot \\ \cdot \\ \cdot \\ \phi_p \end{vmatrix} & \underline{r}_p &= \begin{vmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_p \end{vmatrix} \\
 P_p &= \begin{vmatrix} 1 & r_1 & r_2 & \dots & r_{p-1} \\ r_1 & 1 & r_1 & \dots & r_{p-2} \\ r_2 & r_1 & 1 & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ r_{p-1} & r_{p-2} & \dots & \dots & 1 \end{vmatrix} \\
 \underline{r}_p &= P_p \underline{\phi} & & (3.13)
 \end{aligned}$$

La ecuación 3.13 es llamada de Yule-Walker, es una manera aproximada de estimar los parámetros de una serie AR(p), que puede ser obtenida sustituyendo los coeficientes de correlación teóricos por los estimados y calculados de la Ec. 3.14

$$\bar{\phi} = \bar{P}^{-1} \bar{r}_p \quad (3.14)$$

3.3.3 RELACION ENTRE LOS PARAMETROS ARMA Y LA FUNCION DE AUTOCOVARIANCIA (FA)

Sea la Ecuación 3.1 - Modelo ARMA completo - que multiplicada por \bar{Z}_{t-k} (de aqui en adelante dispensaremos el $\bar{\cdot}$, o sea, consideraremos que la media es siempre sustraída) y tomando el valor esperado podemos deducir:

$$Y_k = \sum_{i=1}^p \phi_i Y_{k-i} - \sum_{i=0}^q \theta_i \lambda_{k-i} \quad (3.15)$$

donde:

$$\theta_0 = -1 \quad \text{y} \quad \lambda_k = E(a_t, Z_{t+k})$$

La correlación cruzada λ_k entre la entrada y la salida puede ser expresada en términos de la expansión MA(ω) (Ecuación 2.1) (Marple, 1987, p.181):

$$\lambda_k = \begin{cases} \sigma^2_m & k = 0 \\ \sigma^2_m w & k < 0 \\ 0 & k > 0 \end{cases} \quad (3.16)$$

La relación final entre los parámetros ARMA y la función de autocovariancia (FA) del proceso usando la ecuación 3.16:

$$Y_k = \begin{cases} \sum_{i=1}^p \phi_i Y_{k-i} - \sigma^2_m \sum_{i=0}^q \theta_i w_{i-k} & 0 \leq k \leq q & (3.17a) \\ \sum_{i=1}^p \phi_i Y_{k-i} & k > q & (3.17b) \\ Y_{-k} & k < 0 & (3.17c) \end{cases}$$

Observamos las ecuaciones 3.17 verificándose los parámetros auto-regresivos del modelo son relacionados

por un conjunto de ecuaciones lineales con la función de autocovariancia. La ecuación 3.17 puede ser válida para los p índices de retardo $q < k \leq q+p$, por ejemplo en las ecuaciones agrupadas en la matriz:

$$\begin{pmatrix} Y_q & Y_{q-1} & \dots & Y_{q-p+1} \\ Y_{q+1} & Y_q & \dots & Y_{q-p+2} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ Y_{q+p-1} & \dots & & Y_q \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \phi_q \end{pmatrix} = \begin{pmatrix} Y_{q+1} \\ Y_{q+2} \\ \cdot \\ \cdot \\ Y_{q+p} \end{pmatrix} \quad (3.18)$$

Por lo tanto la FA para los atrasos $q-p+1$ al $q+p-1$ los parámetros AR pueden ser resueltos encontrados separadamente para los parámetros MA. esa es la llamada ecuación de Yule-Walker modificada (Banon, 1971).

Infelizmente los parámetros MA de los modelos no pueden ser resueltos simplemente como solución de ecuaciones lineales, pues están involucrados con los coeficientes de la expansión MA infinita, resultando en una relación no lineal con la FA.

Haciendo $q=0$ obtenemos las relaciones para un proceso AR puro:

$$Y_k = \begin{cases} \sum_{i=1}^p \phi_i Y_{k-i} & k > 0 \\ \sum_{i=1}^p \phi_i Y_{k-i} + \sigma^2_m & k = 0 \end{cases} \quad (3.19)$$

que puede ser agrupada la expresión:

$$\begin{pmatrix} Y_0 & Y_1 & \dots & Y_p \\ Y_1 & Y_0 & \dots & Y_{p-1} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ Y_p & Y_{p-1} & \dots & Y_0 \end{pmatrix} \begin{pmatrix} 1 \\ -\phi_1 \\ \cdot \\ \cdot \\ \cdot \\ -\phi_p \end{pmatrix} = \begin{pmatrix} \sigma^2_m \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad (3.20a)$$

ó

$$\Gamma_p \begin{pmatrix} 1 \\ -\underline{\phi} \end{pmatrix} = \begin{pmatrix} \sigma^2_m \\ 0 \end{pmatrix} \quad (3.20b)$$

Esta es una forma más completa para la ecuación 3.13 donde la variancia del ruido aparece explícitamente.

La relación entre FA y el modelo MA puede ser obtenida haciendo $p=0$ notandose que $w_k = -\theta_k$ para $1 \leq k \leq q$:

$$Y_k = \begin{cases} 0 & k > q \\ \sigma^2_m \sum_{i=k}^q \theta_i \theta_{i-k} & 0 \leq k \leq q \end{cases} \quad (3.21)$$

Existe por lo tanto una relación no lineal entre la FA y los parámetros MA.

3.3.4 RELACION DE LOS PROCESOS AR CON PREDICCION LINEAL

Los problemas de predicción lineal consiste en estimar una muestra futura a partir de m muestras del pasado mas reciente, por mas de una relación lineal:

$$\bar{Z}_t = \sum_{i=1}^m \alpha_i Z_{t-i} \quad (3.21)$$

Los errores o residuo de predicción es definido por:

$$e_t = Z_t - \bar{Z}_t \quad (3.22)$$

Este residuo es estacionario y su variancia esta dada por:

$$\sigma^2_m = E(e^2_t) \quad (3.23)$$

Procurándose que los coeficientes α de 3.21 minimizan la Ecuación 3.23 y estos son encontrados como solución de ecuaciones normales, estructura por tanto idéntica a la expresión de estimaciones del modelo AR.

Dos distinciones deben ser hechas:

- 1.- Ninguna suposición fue hecha sobre la caracterización estadística de los errores de predicción, mas los coeficientes del filtro predictor obedecen a la relación:

$$\Gamma_m \begin{vmatrix} 1 \\ -\alpha \end{vmatrix} = \begin{vmatrix} \sigma^2_m \\ 0 \end{vmatrix} \quad (3.24)$$

- 2.- Los errores no correlacionados con los datos, y por tanto con predicción lineal (principio de ortogonalidad), mas en general no es un ruido blanco. Esto acontece cuando $m = p$ o sea de orden del filtro de predicción de igual o mayor que el orden de un proceso AR(p) generando Z_t . En estas condiciones el filtro es llamado "blanqueador".

Los coeficientes del filtro, en la práctica, pueden ser obtenidos sustituyéndose de las correlaciones teóricas

por muestras en ecuaciones de Yule-Walker. Utilizando los datos directamente en forma de un sistema de ecuaciones sobredeterminado la solución clásica de pseudo-inversa puede ser obtenida.

Sea $y = [y_1, \dots, y_N]^T$ una secuencia de muestras conocidas del proceso Z_t , esta para cada y_j , $m \leq j \leq N$, de 3.22:

$$y_j = \underline{y}_j + e_j \quad (3.25a)$$

Usando 3.21

$$y_j = \sum_{i=1}^m \alpha_i y_{j-i} + e_j \quad (3.25b)$$

En forma matricial:

$$\underline{y}_j = \Omega \underline{\alpha} + \underline{e}_m \quad (3.25c)$$

donde:

$$\underline{y}_m^T = [y_{m+1}, \dots, y_N] \quad (3.25d)$$

$$\underline{e}_m^T = [e_{m+1}, \dots, e_N], \quad (3.25e)$$

$$\Omega = \begin{bmatrix} y_m & y_{m-1} & \dots & y_1 \\ y_{m+1} & y_m & \dots & y_2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ y_{N-1} & \dots & \dots & y_{N-m} \end{bmatrix} \quad (3.25f)$$

$$\underline{\alpha}^T = [\alpha_1, \dots, \alpha_m] \quad (3.25g)$$

Quierase hacer a los valores de $\underline{\alpha}$ tal que $\underline{e}_m^T \underline{e}_m$ sea mínimo; entonces la solución de mínimos cuadrados será:

$$\underline{\hat{\alpha}} = (\Omega^T \Omega)^{-1} \Omega^T \underline{y}_m = \Omega^+ \underline{y}_m \quad (3.26)$$

donde:

Ω^+ es la llamada pseudo-inversa de Ω .

Si comparamos la ecuación 3.26 con la ecuación 3.14 vease que la matriz $C_{\Omega P}$ corresponde al $\Omega^T \Omega$ y $C_{\Omega r}$ al $\Omega^T \bar{y}_m$, con la distinción de que es estimada de la FA usadas en la Ecuación 3.14 no son las mismas que las calculadas implícitamente con la ecuación 3.26. La diagonal principal de $\Omega^T \Omega$ no tienen los valores exactamente iguales porque existen diferencias en los límites de los cálculos en cada posición de la diagonal principal de $\Omega^T \Omega$.

Por tanto el de las estimaciones polarizadas y no polarizadas para las autocorrelaciones, existe también la estimación calculada implícitamente por la ecuación 3.26. Como los podemos observar entre tanto en Brockwell y Davis (1987), $n^{-1}(\Omega^T \Omega) \xrightarrow{P} \gamma_{\Omega P}$ y $n^{-1} \Omega^T y_n \xrightarrow{P} \gamma_{\Omega P}$, cuando $n \rightarrow \infty$ (\xrightarrow{P} significa convergencia en probabilidad, o sea x_n de una secuencia x_t converge con probabilidad para $x(x_n \xrightarrow{P} x) \Leftrightarrow P(|x_n - x| > \epsilon) \rightarrow 0$ cuando $n \rightarrow \infty$). Esto significa, como asegura Brockwell y Davis (1987, p.257), que las estimaciones de pseudo-inversa y las estimaciones de Yule-Walker convergen en probabilidad para el mismo valor.

En caso general de predicción lineal cuando los residuos son correlacionados entre si, la estimación para $\underline{\alpha}$ en la ecuación 3.26 puede ser polarizada (Bennet, 1979) cuando se conoce la matriz de covariancia de los residuos (Σ_{ϵ}), el mejor estimador es el estimador de mínimos cuadrados generalizados (GLS). Para el modelo 3.25c (sobredeterminado) el valor de $\underline{\alpha}$ es dado por:

$$\underline{\hat{\alpha}} = (\Omega^T \Sigma_{\epsilon}^{-1} \Omega)^{-1} \Omega^T \Sigma_{\epsilon}^{-1} y_m \quad (3.27)$$

Este estimador y BLUE ("best linear unbiased estimator"), reduciendo por tanto la variancia de los residuos. La matriz Σ_{ϵ} normalmente no es conocida, por tanto puede ser estimada por procesos iterativos, tomando como inicio la estimación de mínimos cuadrados simples.

Los usos de GLS no es frecuente en predicción lineal debido a las dificultades que se tiene con la matriz de covariancia del residuo.

3.3.5 SOLUCION RECURSIVA PARA LA ECUACION DE YULE-WALKER Y FUNCION DE AUTOCORRELACION PARCIAL (FAP)

La matriz de la expresión 3.20 es simétrica puesto que cada entrada a_{ij} corresponde a entrada a_{ji} de igual valor. Para solución de ecuaciones de este tipo existe una solución recursiva simple, llamada algoritmo de Levinson. Con este algoritmo puede determinarse secuencialmente los coeficientes del modelo de predicción lineal en orden

creciente. Eventualmente el m-ésimo coeficiente del modelo toma cero ó aproximadamente cero, si el proceso de origen son generados por un modelo AR(m-1). En consecuencia los últimos coeficientes del modelo de predicción lineal en función del orden del modelo da la función de autocorrelación parcial (FAP). y se constituirá herramienta de auxilio é identificación del modelo como se vera posteriormente.

Algoritmo de Levinson:

$$\bar{\phi}_1 = \bar{r}_1 \quad 3.28a$$

$$\bar{\phi}_{p+1,j} = \bar{\phi}_{p,j} - \bar{\phi}_{p+1,p+1} \bar{\phi}_{p,p-j+1} \quad j = 1, \dots, p \quad (3.28b)$$

$$\bar{\phi}_{p+1,p+1} = \frac{r_{p+1} - \sum_{j=1}^p \bar{\phi}_{p,j} r_{p+1-j}}{1 - \sum_{j=1}^p \bar{\phi}_{p,j} r_j} \quad (\text{FAP}) \quad (3.28c)$$

Ejemplo de coeficientes de tercer orden:

$$\bar{\phi}_{31} = \bar{\phi}_{21} - \bar{\phi}_{33} \bar{\phi}_{22}$$

$$\bar{\phi}_{32} = \bar{\phi}_{22} - \bar{\phi}_{33} \bar{\phi}_{21}$$

$$\bar{\phi}_{33} = \frac{r_3 - \bar{\phi}_{21} r_2 - \bar{\phi}_{22} r_1}{1 - \bar{\phi}_{21} r_1 - \bar{\phi}_{22} r_2} \quad (3.29)$$

3.3.6 COEFICIENTES DE REFLEXION

Los coeficientes de reflexión $\bar{\phi}_{1,1}$ (definidos por la Ec. 3.28 son los llamados coeficientes de autocorrelación parcial (FAP(i)) en la literatura estadística, porque representan una correlación normalizada entre Z_t y Z_{t+1} con las correlaciones de Z_{t+1} ante Z_{t+1-1} removidas. Los negativos de estos coeficientes, $k_1 = -\phi_{1,1}$ son frecuentemente llamados coeficientes de reflexión debido a su interpretación como parámetros físicos de un modelo en un tubo acústico o en modelos sísmicos (Robinson y Treitel, 1980).

Puede definirse un error de predicción inversa ("backwards"), donde los coeficientes para cada orden del predictor son los mismos del predictor directo, mas donde se usan las m muestras (por tanto anti casual) y derivar una relación (Marple, 1987) envolviendo los errores de predicción directa e inverso de un proceso estocástico complejo.

La predicción inversa esta dada por:

$$\bar{Z}^1_t = \sum_{j=1}^m \alpha^1_j Z_{t+j} \quad (3.30)$$

Los errores de predicción inversa es:

$$e_t^1 = Z_{t-m} - \bar{Z}_{t-m}^1 \quad (3.31)$$

definido de esta forma para que e_t y e_t^1 sean función de un mismo conjunto de datos (Z_{t-m}, \dots, Z_t) .

La relación entre e_t y e^*_t teniéndose en consideración el orden m del predictor lineal y los coeficientes de reflexión dado por (Marple, 1987):

$$e_{(m)t} = e_{(m-1)t} + k_m e^*_{(m-1)(t-1)} \quad (3.32a)$$

$$e^*_{(m)t} = e^*_{(m-1)(t-1)} + k^*_m e_{(m-1)t} \quad (3.32b)$$

donde $*$ = conjugado.

La ecuación 3.32 provee una interpretación alternativa de filtro de predicción como. Los residuos de un filtro predictor de mayor orden son funciones de los residuos directos e inversos de los filtros de orden inmediatamente inferior. Este tipo de filtro tienen ventajas de implementación pues minimizan la propagación de errores de arredondamiento y tienen menor sensibilidad a los valores de los coeficientes.

3.2.7 CONDICIONES DE ESTABILIDAD Y ESTACIONARIEDAD

Anteriormente hemos introducido cierta hipótesis de estacionariedad para la cual un proceso en particular permanece estable en torno de una media. Esto también significa (Box y Jenkins, 1970) que las raíces de la ecuación característica $\phi(B) = 0$, de la parte autorregresiva del modelo ARMA generado esta fuera del círculo unitario o plano complejo.

Adicionalmente las condiciones denominadas de invertibilidad debe ser observada por el modelo ARMA (Box

y Jenkins, 1970). Esta condición puede ser ilustrada por la idea de que, para un proceso MA(q) puro puede ser representado por un AR infinito de manera convergente, o inverso de $\theta(B)$ debe ser una serie convergente o sea las raíces de $\theta(B) = 0$ deben caer fuera del círculo unitario.

Las dos condiciones se resumen en llamadas condiciones de fase mínima (Oppenheim y Schaffer, 1975; Dudgeon y Mersereau, 1984), que implica que la señal y su transformada Z presenta las siguientes propiedades: es aleatorio los polos (opuestos) y ceros de $H(Z)$ están dentro del círculo unitario y tiene la mayor parte de energía parte del origen (en espectro de potencia) y además de eso H e H^{-1} son absolutamente aditivos.

3.4 IDENTIFICACION DEL MODELO

En los desenvolvimientos efectuados hasta ahora fue suponiendo que se conocían los ordenes del modelo ARMA o (SARIMA si fuera el caso). La determinación del orden entre tanto no es trivial y se constituye en realidad una fase mas difícil del uso de la metodología de series temporales, denominada fase de identificación del sistema.

La primera herramienta de análisis de la serie temporal es de formar la propia serie temporal (usualmente mostrada en un gráfico). Una tendencia

polinomial o mismo exponencial (como es el caso de la inflación peruana entre los años de 1980 a 1990), puede ser observada y esa tendencia retirada. Transformaciones del tipo Log o raíz cuadrada pueden ser utilizadas, normalmente sugeridas por un conocimiento a priori sobre el proceso. El objetivo es producir una serie estacionaria. Como alternativa para el caso de tendencia polinomial se puede optar por diferencias sucesivas de la serie, hasta que se alcance un aspecto de media constante. Aún ese caso, al analizarse la FAC, puede ser que se nota que la misma cae lentamente, sugiriendo que el polinomio del modelo auto-regresivo del proceso tenga un cero muy próximo del círculo unitario. Una muestra finita nunca tendrá una FAC constante, misma que la serie ha sido generada a partir de un modelo integrado de orden $d = 1$. En ese caso será necesaria hacer más de una diferenciación en "lags" unitarios. Una vez eliminada la parte correlación unitaria y estacional se podrá manifestar. Repartir la FAC en atrasos periódicos sugiere un modelo estacional; si en esos atrasos la FAC cae lentamente puede ser necesaria una diferenciación estacional.

En la FAC los coeficientes ρ_1, \dots, ρ_k no tienen un patrón fijo de comportamiento, ya que dependen de los valores de los coeficientes de medias móviles del modelo. Para $k > q$ los coeficientes ρ_k se determinan según una ecuación en

diferencias y decrecen según una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas. La FAP también presenta p valores iniciales sin un patrón fijo, seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas.

En la figura N^o 06 se ha representado la FAC y FAP del modelo ARMA(1,1) estacionario e invertible. La zona conjunta de estacionariedad e invertibilidad está definida por todas las combinaciones situadas en el interior del cuadrado de la figura.

Como puede verse que la diagonal trazada (en la Fig. 6) corresponde a aquellas combinaciones en que $\theta_1 = \phi_1$. Cuando se da esta circunstancia el modelo ARMA queda reducido a un modelo de ruido blanco. En efecto $(1 - \phi_1 B)Z_t = (1 - \theta_1 B)a_t$ se convierte en: $Z_t = a_t$ para $\phi_1 = \theta_1$.

También las combinaciones cercanas a esa diagonal presentan dificultades de identificación, puesto que no serán fácilmente distinguibles de un ruido blanco.

En la tabla N^o 01 se presenta un resumen del comportamiento de la FA y FAP en los modelos AR(p), MA(q) y ARMA(p,q).

Como se puede notar la fase de identificación necesita de una cierta experiencia. En el caso de duda el analista retiene algunas alternativas para coger en el

futuro al analizar el comportamiento del residuo, asunto que será discutido en la sección sobre diagnóstico.

Tabla Nº 01 La FAC y FAP en modelo AR(p), MA(q) y ARMA(p,q)

	FAC	FAP
AR(p)	Decrecimiento rápido de tipo geométrico puro y geométrico con alteración de signos, sinusoidal o mezcla de varios tipos.	Se anula para retardos superiores a p.
MA(q)	Se anula para retardos superiores a q.	Decrecimiento rápido de tipo exponencial y/o sinusoidal.
ARMA(p,q)	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas.	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales y/o exponenciales amortiguadas.

Después de obtenida la estacionariedad de la serie falta establecer las ordenes p, q, P y Q las diferencias $W_t = (1 - B)^q(1 - B^P)^P$. De la teoría (Sec. 3.3) se sabe que la FAC se anula a partir del "lag" q para los modelos MA y la FAP también se anula a partir del atraso p para modelos AR; para el modelo ARMA completo en la FAC y en la FAP se anulan. Se observa por tanto la FAC y la FAP se

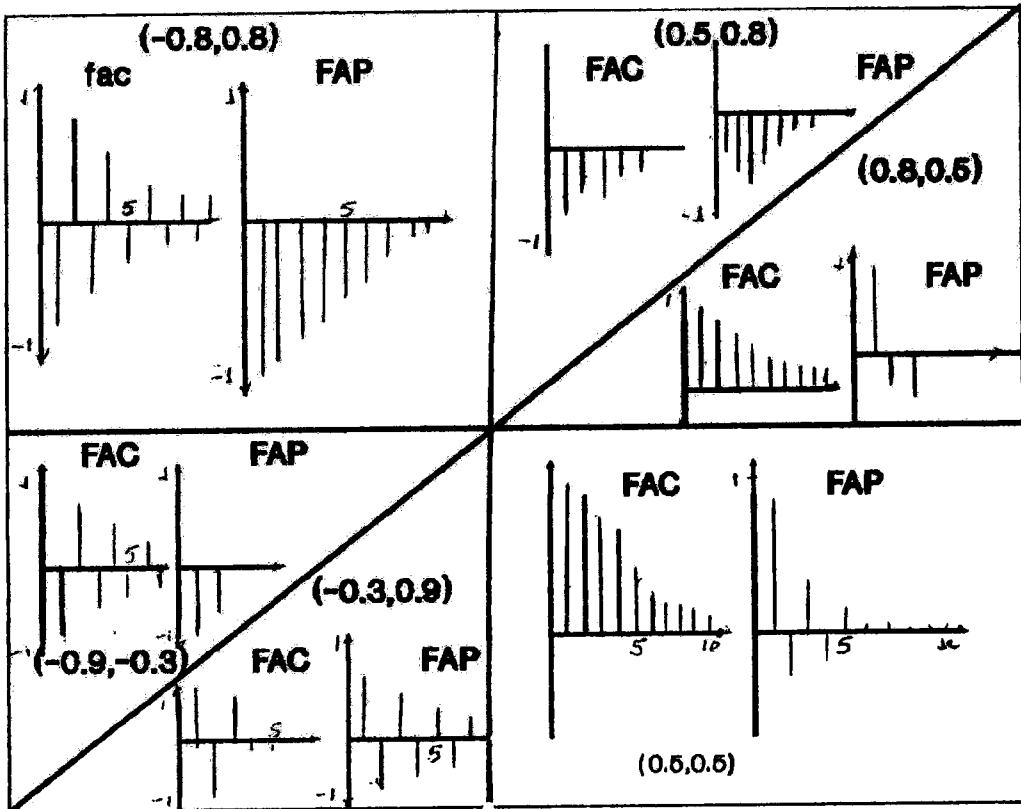


Fig. Nº 06 FAC y FAP de 6 modelos ARMA (1,1) estacionarios e invertibles.

puede tener una idea del modelo subyacente a un proceso. En la práctica podrá ser preciso establecer un criterio para saber cuando las estimaciones no son nulas, debido a las variancias de los estimadores de la autocorrelaciones. Segundo Granger y Newbold (1977) que citan a Anderson (1942) los estimadores de FAC es FAP son aproximadamente gaussianos para todos los "lags", cuando la señal es un ruido blanco, con variancia n^{-1} (donde n es el número de muestras), asi se puede usar el limite de $\pm 2n^{-1/2}$ (equivalente al doble de desviación estandar de la distribución del estimador) como un indicativo mas o menos grosero para saber si las funciones no son cero en ciertos

atrasos, afectandose asimismo la hipótesis del ruido blanco. De esta manera se obtiene inmediatamente de los valores provisionarios para p y q (\bar{p} y \bar{q}), correspondiendo a los valores de "lag" para los cuales la FAP y la FAC caen dentro del intervalo $\pm 2n^{-1/2}$, por tanto consideradas nulas. Para obtener el valor de P y Q cuando fuera el caso, se analiza las $FAC(s_1)$ y $FAP(s_1)$ donde s es el período de $l = 0, 1, 2, \dots$

Para terminar el problema de identificación bien definido, fueron introducidos diversos índices, entre los cuales, los más utilizados es el llamado índice de AKAIKE (AIC) (Priestley, 1981), que es puesto de la siguiente forma:

$$AIC(p,q) = n \ln \bar{\sigma}_{p,q}^2 + 2(p+q) \quad (3.33)$$

donde $\bar{\sigma}_{p,q}^2$ es igual a la variancia estimada para el ruido para un modelo ARMA(p,q).

La idea general para el uso de este criterio de cualidad para el ajustamiento es que en general al aumentar el número de parámetros del modelo la variancia del residuo disminuye y el residuo es mas "blanco". En tanto aumentar demasiadamente el número de parámetros es desventaja, al disminuir la precisión en la estimación de los parámetros, es parte contraria al principio de parsimonia, par la cual se debe procurar modelos en su forma equivalente al más compacto posible.

De tener los valores \bar{p} y \bar{q} iniciales, escojemos algunas ordenes $p < \bar{p}$ y $q < \bar{q}$ para estimar los parámetros de los modelos ARMA(p,q). Los modelos escogidos es aquel que minimiza el índice de AIC.

Aún así se verifica la necesidad de tener que estimar los parámetros de un número grande de modelos, principalmente si hubiera opciones con relación al grado de diferenciación y transformación utilizadas en la serie original. Para disminuir esas dificultades se puede utilizar la transformación sobre las desviaciones estandar de los estimadores de los parámetros para eliminar coeficientes poco significativos. La literatura presenta también otras funciones auxiliares para la identificación del modelo tales como las autocorrelaciones inversas, autocorrelación parcial inversa, arreglos R^* y S^* y el método Corner; todos esos revisados en Pereira (1984). Esos métodos también facilitan la identificación del Proceso SARIMA pues muchas de esas funciones auxiliares indican que los coeficientes intermediarios deben ser eliminados. En caso de identificación de modelos estacionales el índice de AIC debe ser modificado para

$$AIC_s = 2 n \ln \sigma_s + 2m \quad (3.34)$$

donde m es el número de coeficientes estimados.

3.5 ESTIMACION DE LOS PARAMETROS

Vease en esta sección como, una vez fijada el orden del modelo, obtener los valores de los coeficientes del modelo ARMA.

3.5.1 ESTIMACION DE LOS PARAMETROS AUTO-REGRESIVOS

Sea $y^T = [y_1, \dots, y_m]$ una secuencia de muestras conocidas del proceso Z_t ; deseandose que a partir de estos datos, estimar los parámetros de un modelo tipo AR(p), p dado, ajustado a esos datos. Existen diversos métodos de estimación de parámetros entre los cuales se pueden citar (Melsa y Cohn, 1978): mínimos cuadrados (LS), máxima verosimilitud (ML) máximo a posteriori (MAP), que difieren entre sí la cantidad de conocimiento estadístico que se puede tener con respecto a los parámetros, siendo estimado y de comportamiento estadístico de ruido. Los métodos MAP no serán discutidos en el presente.

Para el métodos de los mínimos cuadrados, conforme se ha visto en la sección 3.3.4 la solución es obtenida por la sustitución en las ecuaciones de Yule-Walker, de las correlaciones teóricas por las estimadas, además de eso, la solución también puede ser obtenida por la pseudo-inversa de la matriz de datos. Lo que cambian básicamente en esas soluciones son las fórmulas usadas para los cálculos de las estimaciones de las autocorrelaciones.

El método ML, da una estimación de mejor cualidad que el LS, pues permite incorporar en la solución el conocimiento a priori con respecto al ruido que afecta al modelo (ecuación 3.25c). Al ajustar el modelo a los datos contenidos en y , obteniéndose el siguiente sistema en forma matricial:

$$y_m = \Omega\phi + e_m \quad (3.35)$$

como la ecuación 3.25 $e^T_m = [e_{m+1}, \dots, e_N]$ y ϕ ocupa el lugar de los parámetros α de posición al orden del predictor en ese caso es igual al orden del modelo AR(p).

La estimación ML para ϕ , cuando el residuo e_m tiene una distribución gaussiana, y dada por (Melsa y Cohn, 1978):

$$\bar{\phi} = (\Omega^T \Sigma^{-1} \Omega)^{-1} \Omega^T \Sigma^{-1} y_m \quad (3.36)$$

Observa que para ese caso la solución es idéntica al GLS (ecuación 3.27). esta estimación ML por no ser "exacta" pues la construcción de los sistemas de ecuaciones, con parte de los datos disponibles para utilizar como condición inicial (implicito la formación del vector y_m). Esa es la estimación de máxima verosimilitud condicional aquellas condiciones iniciales.

A partir de una mejor aproximación de verosimilitud "verdadera" de los datos disponibles Box y Jenkins (1970) presentan la siguiente expresión para la

estimación ML de ϕ :

$$\bar{\phi} = D^{-1}_p d \quad (3.37)$$

donde:

$$d = \begin{bmatrix} D_{12} \\ \vdots \\ D_{1,p+1} \end{bmatrix} \quad D_p = \begin{bmatrix} D_{22} & D_{23} & \dots & D_{2,p+1} \\ D_{33} & D_{33} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ D_{p+1,p+1} & \dots & \dots & D_{p+1,p+1} \end{bmatrix}$$

$$D_{12} = D_{21} = \bar{y}_1 \bar{y}_2 + \bar{y}_{1+1} \bar{y}_{2+1} + \dots + \bar{y}_{n+1-2} \bar{y}_{n+1-1} \quad (3.38)$$

observese: \bar{y}_t forma centralizada de y_t .

Descomponiendo 3.38 para algunos términos en la diagonal principal de D_p tenemos:

$$D_{22} = y^2_{n-1} + \dots + y^2_3 + y^2_2$$

$$D_{33} = y^2_{n-2} + \dots + y^2_3$$

$D_{44} = \dots$ disminuyendo el número de términos en la sumatoria.

Como se puede observar las estimaciones ML aproximadas conducen a soluciones semejantes al de las ecuaciones de Yule-Walker con las correlaciones estimadas, con variación observada al cálculo de las mismas autocorrelaciones. Los métodos vistos hasta ahora, el último presentado y el que presenta mayor aproximación con la llamada verosimilitud exacta (Priestley, 1981).

3.5.2 ESTIMACION DE PARAMETROS DEL MODELO MA

No existe una manera directa para estimar los parámetros de los modelos MA, del tipo de ecuación de YW, así como se puede ajustar modelos AR de ordenes crecientes a los datos $y^T = [y_1, \dots, y_n]$ aplicando el algoritmo de Levinson a las autocovariancias muestrales, se puede ajustar a modelos de media móvil de ordenes crecientes

$$\bar{z}_t = a_t - \theta_{m1}a_{t-1} - \dots - \theta_{mm}a_{t-m} \quad (3.39)$$

$q = 1, 2, \dots$, por medio del algoritmo de innovaciones (Brockwell y Davis, 1987), donde θ_{mn} significa el n -ésimo coeficiente de un modelo de orden m y $v_m = \sigma^2_{a_m}$ es la variancia de a_t :

Algoritmo de innovaciones

$$\bar{v}_{m0} = c_0 \text{ (variancia estimada de } Z_t \text{)}$$

$$\bar{\theta}_{m, m-k} = -\bar{v}^{-1}_{m, k} \left[c_{m-k} - \sum_{j=0}^{k-1} \bar{\theta}_{m, m-j} \bar{\theta}_{k, k-j} \bar{v}_{m, j} \right] \quad (3.40a)$$

$k = 0, \dots, m-1$; c_k : es la autocovariancia estimada para el atraso k .

$$\bar{v}_{m, m} = c_0 - \sum_{j=1}^{m-1} \bar{\theta}^2_{m, m-j} \bar{v}_{m, j} \quad (3.40b)$$

Marple (1987) presenta un método alternativo para estimación de parámetros MA, que usa la representación de un proceso MA para un modelo AR de orden elevada. Los

parámetros AR(M), donde $M \gg q$ son posibles calcular usando por ejemplo el algoritmo de Levinson. Como los parámetros MA tiene una relación convolucional con los parámetros AR visto que $TZ(MA(q))/TZ(AR(\infty)) = 1$, (donde TZ significa la transformada Z) construyéndose un sistema de ecuaciones que permite calcular los parámetros MA ya que conocemos los parámetros AR. No sería posible usar la ecuación 3.21 porque la variable incógnita aparece como producto $\theta_1 \theta_{1-k}$ dificultando la resolución del sistema.

3.5.3 ESTIMACION PRELIMINAR DE LOS PARAMETROS DEL MODELO ARMA

Sea el modelo completo, de media cero, estable.

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.4)$$

Las condiciones de estabilidad asegura que la expresión MA(∞) converge.

$$Z_t = \sum_{j=0}^{\infty} w_j a_{t-j} = w(B)a_t \quad (3.41)$$

Si observamos e interpretamos del punto de vista de transformada Z (ecuación 3.2b), se deduce que

$$w(Z^{-1}) = \theta(Z^{-1}) / \phi(Z^{-1}) \quad \delta:$$

$$\theta(Z^{-1}) = w(Z^{-1})\phi(Z^{-1}) \quad (3.42)$$

Tomando la anti-transformada de 3.42 obsérvese que se obtiene la convolución de los términos w y ϕ

(Brockwell y Davis, 1987):

$$w_0 = 1$$

$$\theta_j = w_j + \sum_{i=1}^{\min(j,p)} \phi_i w_{j-i} \quad j=1,2,3,\dots \quad (3.43)$$

donde por convención $\theta_j = 0$ para $j > q$ y $\phi_j = 0$ para $j > p$. observándose a partir de la ecuación 3.43 es posible, utilizándose de esas relaciones para $j = q + 1$ hasta $q+p$, extraer los p valores de ϕ_i ya que en ese intervalo $\theta_j = 0$ desde que se conoce w_1 . Para estimar los necesarios w_1, \dots, w_{q+p} se puede usar las estimaciones generadas por el algoritmo de las innovaciones. Sustituyendo w_j por $-\bar{\theta}_m$ en 3.43 y resolviendo las ecuaciones resultantes, para m suficientemente grande:

$$\bar{\theta}_m = \theta_j + \sum_{i=0}^{\min(j,p)} \phi_i \bar{\theta}_m, j-1, \quad j=1,2,\dots,q+p \quad (3.44)$$

La ecuación 3.43 con $j = q+1, \dots, q+p$, observándose que $\bar{\phi}$ debe satisfacer la ecuación:

$$\bar{\theta}_m = \bar{\phi} \bar{\phi} \quad (3.45)$$

donde:

$$\bar{\theta}_m = \begin{vmatrix} \bar{\theta}_m, q+1 \\ \bar{\theta}_m, q+2 \\ \vdots \\ \bar{\theta}_m, q+p \end{vmatrix} \quad \bar{\phi} = \begin{vmatrix} \bar{\phi}_1 \\ \bar{\phi}_2 \\ \vdots \\ \bar{\phi}_p \end{vmatrix}$$

$$\bar{\phi} = \begin{vmatrix} \bar{\theta}_{m,q} & \bar{\theta}_{m,q-1} & \dots & \bar{\theta}_{m,q+1-p} \\ \bar{\theta}_{m,q+1} & \bar{\theta}_{m,q} & \dots & \bar{\theta}_{m,q+2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\theta}_{m,q+p-1} & \dots & \dots & \bar{\theta}_{m,q} \end{vmatrix}$$

siendo resuelto 3.45 para $\bar{\phi}$ entre $\bar{\theta}$ y haciendo uso de 3.44, tenemos:

$$\bar{\theta}_j = \bar{\theta}_{m,j} - \sum_{i=1}^{\min(j,p)} \bar{\phi}_i \bar{\theta}_{m,j-i}, \quad j = 1, 2, \dots, q \quad (3.46)$$

La variancia del ruido blanco es calculado a partir de 3.40b:

$$\sigma^2_m = \bar{v}_{mm} \quad (3.47)$$

Un criterio para escoger m suficientemente grande para considerar una buena aproximación para un modelo $MA(\infty)$ es escoger m de forma que sean ajustados al modelo $MA(m)$ de orden creciente, de tal forma que se minimice el índice del criterio AIC, que corresponde a la sección de identificación del modelo.

MARPLE (1987) presenta otro método para la estimación preliminar de los parámetros $\bar{\theta}$ y $\bar{\phi}$ en los siguientes pasos; primero son estimados los parámetros AR utilizandose de una variante de la ecuación de YW modificada, llamada la ecuación de YW modificada de

mínimos cuadrados. La serie es filtrada para así obtener residuos del tipo MA. Un modelo MA(q) es ajustado a estos residuos usándose el método anteriormente descrito.

Vale observar que la ecuación 3.45 es una ecuación del tipo YW donde los coeficientes de expansión MA(∞) sustituyen las correlaciones muestrales.

3.5.4 ESTIMACION DE PARAMETROS POR MAXIMA VEROSIMILITUD PARA LOS MODELOS ARMA(p,q)

Sea $y^T = [y_1, y_2, \dots, y_n]$ un vector de observaciones de un proceso estocástico Z_t de media cero. Conociéndose el orden del modelo ARMA(p,q) que generó el proceso, mas no se conoce los parámetros $\phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q$ y σ^2_ϵ (la media se supone cero), o sea las observaciones y_t fueron extraídas de una distribución conjunta $p(y|\phi, \theta, \sigma^2_\epsilon)$. Contando con las observaciones deseamos saber cuales son los valores para los parámetros más probables de haber dado origen a esos datos. Las estimaciones de máxima verosimilitud de los parámetros son considerados determinísticos, y deseandose esta maximizar una función $L(\phi, \theta, \sigma^2_\epsilon | y)$ que es de la misma forma que $p(y|\phi, \theta, \sigma^2_\epsilon)$.

Asúmese que las observaciones vienen de una distribución normal.

Para inicio del análisis, considérase, por lo tanto una distribución de ruido de entrada o "shocks", que suponen gaussianos e independientes entre si. siendo la

distribución de 1 "shock" simple dado por:

$$p(a_t | \sigma^2_{\epsilon}) = \frac{(2\pi)^{-1/2}}{\sigma_{\epsilon}} \exp\left[-\frac{a_t^2}{2\sigma^2_{\epsilon}}\right] \quad (3.48)$$

La distribución conjunta de los "shocks" de entrada será:

$$p(a_1, \dots, a_n) = (2\pi)^{-n/2} \sigma_{\epsilon}^{-n} \exp\left\{-\frac{1}{2\sigma^2_{\epsilon}} \sum_{t=1}^n a_t^2\right\} \quad (3.49)$$

más de 3.1

$$a_t = y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (3.50)$$

sustituyendo 3.50 en 3.49 se obtiene la densidad de probabilidad conjunta de y dado que el Jacobiano de la transformación es unitario (Nelson, 1973; Priestley, 1981):

$$p(y | \phi, \theta, \sigma^2_{\epsilon}) = (2\pi)^{-n/2} \sigma_{\epsilon}^{-n} \exp\left[-\frac{1}{2\sigma^2_{\epsilon}} \sum_{t=1}^n (y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} + \dots + \theta_q a_{t-q})^2\right] \quad (3.51)$$

La función de verosimilitud de los datos es por lo tanto:

$$L(\phi, \theta, \sigma^2_{\epsilon} | y) = (2\pi)^{-n/2} \sigma_{\epsilon}^{-n} \exp\left[-\frac{1}{2\sigma^2_{\epsilon}} \sum_{t=1}^n a_t(\phi, \theta)^2\right] \quad (3.52)$$

Puede observarse que la expresión 3.50 no siempre puede ser calculada para todos los índices t , pues, conforme el orden del modelo no tiene los valores de a_t o y_t para $t \leq 0$. Para resolver el problema o se ajustan los límites de la sumatoria en 3.52 o se asumen los valores no

existentes como cero. Box y Jenkins (1976) sugieren también un método denominado "back-forecasting" para estimar los valores iniciales necesarios para el cálculo de $L(\mathbf{z})$. Así siendo la función de verosimilitud $L(\mathbf{z})$ debería ser exacta, pues no usa todo los datos disponibles o esto es necesario establecer hipótesis para las condiciones iniciales (c.i.). De cualquier forma para un número grande de datos de tratamiento las c.i. no influenciaran mucho el resultado. Refiriéndose a (Priestley, 1981) para mayores detalles sobre la elaboración de verosimilitud exacta.

Lo que se procura son los valores de ϕ , θ y $\sigma^2_{\mathbf{a}}$ que maximizan la ecuación 3.52. Para facilitar los cálculos e eliminándose los valores constantes basta maximizar la expresión:

$$\text{Log } L(\mathbf{z}) \propto l(\phi, \theta, \sigma^2_{\mathbf{a}}) = -n \ln \sigma_{\mathbf{a}} - \frac{S(\phi, \theta)}{2\sigma_{\mathbf{a}}} \quad (3.53)$$

donde:

$$S(\phi, \theta) = \sum_{t=1}^n a_t(\phi, \theta)^2 \quad (3.54)$$

Para maximizar 3.53 basta inicialmente minimizar $S(\mathbf{z})$, y una vez calculado ese valor y los parámetros $\bar{\phi}$ y $\bar{\theta}$, el valor $\sigma^2_{\mathbf{a}}$ es dado por (derivando e igualando a cero):

$$\sigma_{\epsilon}^2 = \frac{S(\bar{\phi}, \bar{\theta})}{n} \quad (3.55)$$

En el cálculo de los parámetros del modelo ARMA, en general la expresión $S(\mathbf{a})$ será altamente no lineal en relación a los parámetros; la solución directa es por lo tanto muy difícil en la mayoría de los casos. Para la solución aplíquese los métodos de estimación no lineal, por ejemplo basándose en aproximación lineal del modelo, el llamado método de Gauss-Newton (GN) (Nelson, 1973), que es un método iterativo y que utiliza inicialmente las estimaciones preliminares generadas por los esquemas descritos en las secciones anteriores. Las diversas estimaciones para los parámetros son generados hasta que $S(\bar{\phi}, \bar{\theta})$ no cambie más dentro de ciertos límites y los valores alcanzados para $\bar{\phi}$ y $\bar{\theta}$ en la última iteración son las estimaciones de máxima verosimilitud procurada.

En principio la función $S(\phi, \theta)$ puede tener varios mínimos locales, mas usandose las estimaciones preliminares como partida, normalmente se alcanza al menor valor entre los mínimos (Nelson, 1973).

3.4.5 ERROR ESTANDAR DE LOS PARAMETROS ESTIMADOS

Las estimaciones por máxima verosimilitud de los parámetros son v.a. desde que son función de los datos. De la teoría estadística (Nelson, 1973) se sabe que en

transformaciones en los datos hasta tomar la estacionariedad. Una transformación muy utilizada es la función logarítmica o la raíz cuadrada por ejemplo. Ciertos cuidados deben ser tomados cuando se quiere hacer la previsión de los datos originales, pues no siempre la previsión en campo original es dada por la inversa de la transformación utilizada (Pereira, 1984, Sec. 9.5).

El tipo de no-estacionariedad más comunes es la llamada no-estacionariedad homogénea cuando sucesivas diferencias de la serie en estudio son estacionarias, o sea Z_t es esta una serie no-estacionaria homogénea de la serie:

$$W_t = \nabla^d Z_t = (1 - B)^d Z_t \quad (3.57)$$

son estacionaria; d es el orden de diferenciación y $(1-B)Z_t = Z_t - Z_{t-1}$

Una serie temporal con no estacionariedad homogénea es llamada un proceso ARIMA(p,d,q) (Autorregresive Integrated Moving Average), donde d es el grado de diferenciación y $W_t = \nabla^d Z_t$ es un proceso ARMA(p,q).

Como ejemplo, podemos observar una serie cuya media sea función lineal ascendente en el tiempo (también llamada de no-estacionariedad en tendencia). Basta diferenciar una vez y fácilmente se observa que la media

tendra media constante. Muchas veces, podra ser necesario conjugar una transformación y una diferenciación, para tener la serie estacionaria.

En adelante se verá como identificar una posible no-estacionariedad a partir del análisis de FAC y de FAP.

3.6.2 MODELOS ESTACIONALES

Estacionalidad es un fenómeno que ocurre con mucha frecuencia en diversos procesos naturales, como datos hidrometeorológicas, particularmente en los procesos económicos, donde prácticamente todo se refiere a períodos anuales. Existen 2 métodos principales que procuran modelar la estacionalidad: un determinístico, donde el proceso estacional es considerado una suma de un proceso determinístico periódico con un proceso estocástico; otro estadístico que permite cierta aleatoriedad en la componente estacional. Aquí el proceso estocástico es considerado en una multiplicación de componentes, una no estacional para modelar los aspectos intraperiodos y otras componentes cuantas fueron los aspectos periodicos determinados.

El primer método denominado de "descomposición clásica" o proceso Z_t es considerado como:

$$Z_t = m_t + s_t + W_t \quad (3.58)$$

donde m_t es una función que varía lentamente, llamada de

tendencia, s_t es la componente estacional y W_t es el proceso ARMA.

Obsérvese que en ese caso W_t ya es considerado estacionario, pues el término tendencia es retirado de la serie, por ejemplo, ajustando una curva poligonal a los datos observados de la serie y retirando a esos valores determinísticos en función de t , de la serie observada. La diferenciación secuencial discutida en la sección anterior también podría eventualmente ser aplicada aquí para así alcanzar la estacionariedad. El término s_t -estacional-, aquí también es una componente determinística sería calculada, por ejemplo, por considerar la media de valores correspondientes dentro de cada estación (por ejemplo: la media del número de accidentes de carro en Lima en un periodo de 10 años, después de retirar la leve tendencia del aumento de número de carros vendidos de año por año). Esas medias de los valores en las posiciones temporales dentro de la estación, definen un periodo de la función que se repite periódicamente durante todo el tiempo al que se refieren las observaciones. De las observaciones y_t , instancias del proceso Z_t , se extrae por tanto la serie W_t que es la serie a ser modelada habiendo sido retirada la tendencia y la estacionalidad. Otros métodos pueden ser vistos en (Brockwell y Davis, 1987; Granger y Newbold, 1977).

Función de Autocorrelación.

Las FAC y FAP son calculadas con las ecuaciones para series no estacionales, pero esta vez con período estacional. Examinaremos en particular un modelo AR(1) estacional ó AR(1)_s y viene definido por:

$$Z_t = \phi_1 Z_{t-s} + a_t \quad (3.59)$$

entonces bajo el supuesto de estacionariedad, que exige $|\phi| < 1$ la variancia y las covarianzas se pueden calcular multiplicando ambos miembros de 3.59 por Z_{t-s} y tomando esperanzas; así tenemos:

$$k=0$$

$$\gamma_0 = E[Z_t^2] = \phi_1^2 E(Z_{t-s})^2 + E(a_t)^2 + 2E[a_t Z_{t-s}] = \phi_1^2 \gamma_0 + \sigma_a^2$$

$$\text{entonces } \gamma_0 = \sigma_a^2 / (1 - \phi_1^2)$$

$$k=1$$

$$\gamma_1 = E(Z_t Z_{t-s}) = \phi_1 E(Z_{t-s} Z_{t-s}) + E(a_t Z_{t-s});$$

las dos esperanzas del segundo miembro son iguales a cero (0), lo que Z_{t-s} y Z_{t-1} no existe ninguna relación, entonces $\gamma_1 = 0$.

Se puede resumir las distintas autocovarianzas (FA) y se pueden expresar:

$$\gamma_k = \begin{cases} \sigma_a^2 / (1 - \phi_1^2) & \text{para } k=0 \\ \phi_1^k \gamma_0 & \text{para } k=s, 2s, 3s, \dots \\ 0 & \text{en los demás casos.} \end{cases} \quad (3.60)$$

dividiendo ambos miembros de 3.60 entre γ_0 , se obtiene la FAC:

$$\rho_k = \begin{cases} 1 & \text{para } k=0 \\ \phi_1 \rho_{k-s} & \text{para } k=s, 2s, 3s, \dots \\ 0 & \text{en los demás casos.} \end{cases} \quad (3.61)$$

Como puede verse en el modelo AR(1)_s son aplicable los resultados obtenidos para un AR(1) ordinario, con la única salvedad de que solamente hay que considerar los coeficientes correspondientes a los retardos $s, 2s, 3s, \dots$. Como ejemplo de la función de autocorrelación para un modelo AR(1) y un AR(1)_s con $\phi_1 = \phi_1$ y $s = 4$; se muestra en la tabla N^o 02 de igual manera se muestra los graficos en la Fig. N^o 07.

TABLA N^o 02 Función de autocorrelación en un modelo AR(1) y un modelo AR(1)_s.

($\phi_1 = \phi_1 = 0,8$)

Retardo	AR(1)	AR(1) _s
1	0,8000	0
2	0,6400	0
3	0,5120	0
4	0,4096	0,8000
5	0,3277	0
6	0,2621	0
7	0,2097	0
8	0,1678	0,6400
9	0,1342	0
10	0,1074	0
11	0,0859	0
12	0,0687	0,5120

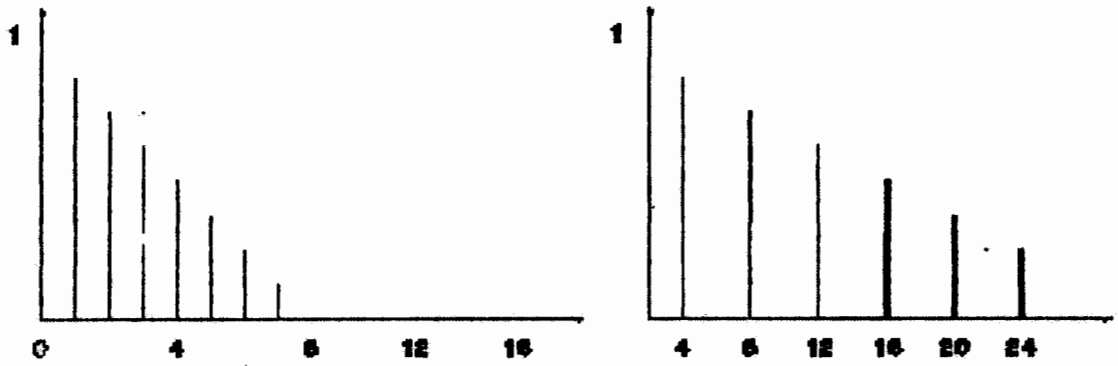


Fig. No 07 Función de autocorrelación de modelos AR(1) y AR(1)_s.

Otros modelos estacionales puros más usados son:

$$AR(2)_s : Z_t = \phi_1 Z_{t-s} + \phi_2 Z_{t-2s} + a_t \quad (3.62)$$

$$\rho_k = \begin{cases} 1 & k=0 \\ \phi_1 \rho_{k-s} + \phi_2 \rho_{k-2s} & k=s, 2s, 3s, \dots \\ 0 & \text{en los demás casos} \end{cases} \quad (3.63)$$

$$MA(1)_s : Z_t = a_t - \delta_1 a_{t-s}$$

$$\rho_k = \begin{cases} 1 & k=0 \\ -\delta_1 / (1 + \delta_1^2) & k=s, 2s, 3s, \dots \\ 0 & \text{en los demás casos} \end{cases} \quad (3.64)$$

$$MA(2)_s : Z_t = a_t - \delta_1 Z_{t-s} - \delta_2 a_{t-2s}$$

$$\rho_k = \begin{cases} 1 & k=0 \\ \frac{-\delta_1 + \delta_1 \delta_2}{1 + \delta_1^2 + \delta_2^2} & k=s \\ \frac{-\delta_1}{1 + \delta_1^2 + \delta_2^2} & k=2s, 3s, \dots \\ 0 & \text{en los demás casos} \end{cases} \quad (3.65)$$

$$\text{ARMA}(1,1)_s : Z_t = \phi_1 Z_{t-s} + a_t - \delta_1 a_{t-s}$$

$$\rho_k = \begin{cases} 1 & k=0 \\ \frac{(1 - \phi_1 \delta_1)(\phi_1 - \delta_1)}{1 + \delta_1^2 - 2\phi_1 \delta_1} & k=s \\ \phi_1 \rho_{t-s} & k=2s, 3s, \dots \\ 0 & \text{en los demás casos} \end{cases} \quad (3.66)$$

En general, un modelo ARMA(P,Q) estacional puro tendrá la siguiente formulación:

$$Z_t - \phi_1 Z_{t-s} - \dots - \phi_p Z_{t-sp} = a_t - \delta_1 a_{t-s} - \dots - \delta_q a_{t-sq}$$

utilizando el operador polinomial de retardos:

$$\phi_s(B)Z_t = \delta_s(B)a_t \quad (3.67)$$

para $s=12$

$$\phi_{12}(B)Z_t = \delta_{12}(B)a_t \quad (3.68)$$

cuyos operadores son definidos por:

$$\phi(B) = 1 - \phi_1 B^{12} - \dots - \phi_p B^{12p} ; \text{ para el proceso AR}(P)_s$$

$$\delta(B) = 1 - \delta_1 B^{12} - \dots - \delta_q B^{12q} ; \text{ para el proceso MA}(Q)_s$$

Modelos Estacionales Multiplicativos Estacionarios

Anteriormente se han examinado modelos ARMA sin ningún calificativo y que denominaremos ordinarios, después los modelos ARMA estacionales y ahora combinando estos dos modelos se pueden formular el siguiente modelo general mixto multiplicativo ARMA(p,q)xARMA(P,Q)_s, también llamada SARMA(p,q)x(P,Q)_s y representada :

$$\phi(B)\Phi(B)Z_t = \theta(B)\delta(B)a_t \quad (3.69)$$

En 3.69, tanto la parte autorregresiva como en la parte de medias móviles aparecen el producto de dos operadores polinomiales: uno ordinario y otro estacional y como la serie de caudales tiene un periodo estacional con $s=12$, como ejemplo de un modelo multiplicativo simple tenemos:

MA(1) x MA(1)₁₂, expresado:

$$Z_t = (1 - \theta_1 B)(1 - \delta_1 B^{12})a_t \quad (3.70)$$

desarrollando 3.70 se tiene:

$$Z_t = a_t - \theta_1 a_{t-1} - \delta_1 a_{t-12} + \theta_1 \delta_1 a_{t-13} \quad (3.71)$$

Evidentemente la formulación equivalente a 3.71 utilizando solamente parámetros ordinarios sería la siguiente:

$$Z_t = a_t - \theta_{12} a_{t-1} - \theta_{13} a_{t-12} + \theta_{13} a_{t-13} \quad (3.72)$$

Como puede observarse, el modelo 3.71 contiene dos parámetros, uno ordinario y otro estacional, reflejando el término a a_{t-13} la interacción entre ambos y es utilizado por el "Forecast Plus". En cambio en 3.72 se ha introducido un nuevo parámetro para recoger la influencia de este último término llevando a un ordinario θ_{13} y que es utilizado por el "TSP".

Naturalmente si existe la interacción entre la parte ordinaria y la parte estacional es preferible utilizar el modelo 3.71 ya que es más escueto.

Función de Autocorrelación

El cálculo de la FAC correspondiente a los modelos

SARMA pueden llegar a ser laboriosos.

En Uriel (1985), menciona a Peña (1979) que ha deducido una fórmula general que permite aproximar los coeficientes de autocorrelación de un proceso ARMA(p,q) x ARMA(P,Q)_s. Esta aproximación viene dada por:

$$\rho_k^x = \rho_k + \sum_{i=1}^{\infty} \rho_{k+i}^{\bar{}} (\rho_{k+i}^x + \rho_{k+i-x}^x) \quad (3.73)$$

donde:

ρ_k : coeficiente de autocorrelación de orden k en un proceso ARMA(p,q).

$\rho_k^{\bar{}}$ coeficiente de autocorrelación de orden k en un proceso estacional puro ARMA(P,Q)_s

ρ_k^x coeficiente de autocorrelación de orden k en un proceso multiplicativo ARMA(p,q) x ARMA(P,Q)_s .

La ecuación 3.73 es exacta en caso de que la parte ordinaria sea un proceso de medias móviles de orden inferior al período estacional. Si la parte ordinaria tiene términos autorregresivos, entonces es claro que $\rho_k^{\bar{}}$ no es igual a cero (0); sin embargo, si el orden de la parte autorregresiva es bajo y si el período estacional es relativamente elevado.

En el caso de datos mensuales, se verifica que $\rho_k^{\bar{}} \approx 0$ para $k \geq s$; adoptando esta aproximación, la aplicación de la fórmula 3.73 permitirá obtener los siguientes resultados:

- i) En los retardos bajos $1, 2, \dots, s/2$ la única estructura que aparece es la correspondiente a la parte ordinaria.

$$\rho_{-}^{x} \approx \rho_{x} ; \quad \text{para } k = 0, 1, 2, \dots, s/2 \quad (3.74)$$

ya que $\rho_{k+s-1} \approx 0$ y $\rho_{s-1-k} \approx 0$ para $i=1, 2, \dots$

- ii) En los retardos $k=s, 2s, 3s, \dots$ el único efecto importante es el correspondiente a la parte estacional, en efecto:

$$\rho_{-}^{x} = \rho_{-} + \bar{\rho}_{-}(\rho_{2s} + \rho_{0}) + \bar{\rho}_{-}(\rho_{3s} + \rho_{-s}) \approx \bar{\rho}_{-} \rho_{0} = \bar{\rho}_{-} \quad (3.75)$$

- iii) En los retardos contiguos a múltiplos de s , se obtiene simétricamente la reproducción de la parte ordinaria tomando como referencia los valores de los retardos estacionales.

Así para $k=s+g$, siendo $g < s/2$, se obtiene el siguiente resultado:

$$\rho_{-+g}^{x} = \rho_{-+g} + \bar{\rho}_{-}(\rho_{2s+g} + \rho_{-g}) + \dots \approx \bar{\rho}_{-} \rho_{-g} \approx \bar{\rho}_{-} \rho_{g} \quad (3.76)$$

análogamente:

$$\rho_{--g}^{x} = \rho_{--g} + \bar{\rho}_{-}(\rho_{2s-g} + \rho_{g}) + \dots \approx \bar{\rho}_{-} \rho_{g}$$

de acuerdo a lo anterior;

$$\rho_{-+g}^{x} \approx \rho_{--g}^{x}$$

en general se verifica que:

$$\rho_{-+n+g}^{x} \approx \rho_{--n-g}^{x} \approx \bar{\rho}_{-}^{n} * \rho_{g} \quad (3.77)$$

A los coeficiente de autocorrelación contiguos y multiples de s se los denomina a veces coeficientes satélites, caracterizandose porque recogen la interacción entre la parte ordinaria y la parte estacional, y por la propiedad de simetría. Esta propiedad se puede perder cuando el periodo estacional es corto ($s=4$) y la parte ordinaria es autorregresiva con raíces cercanas al círculo unitario. Podemos citar como ejemplo un modelo SARMA(0,2) x (0,1), cuya expresión es: $Z_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \delta_1 B^s) a_t$ y de acuerdo a 3.73, 3.65 y 3.8 y las FAC toman los siguientes valores:

$$\begin{aligned}
 & 1 & k=0 \\
 p_1 &= \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2} & k=1 \\
 p_2 &= \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} & k=2 \\
 p_k &= p_2 p_1^{s-k} & k=s-2 \\
 & p_1 p_2 & k=s-1 \\
 p_s &= \frac{-\delta_1}{1 + \delta_1^2} & k=s \\
 p_1 p_s &= p_{s-1}^k & k=s+1 \\
 p_2 p_s &= p_{s-2}^k & k=s+2 \\
 0 & & \text{en los demás casos}
 \end{aligned}
 \tag{3.78}$$

Aunque existen estudios referente al comportamiento de la FAP en un modelo estacional multiplicativo, los patrones son más difíciles de establecer. En general, puede decirse que la parte de medias móviles ordinaria y/o estacional, introduce un decrecimiento exponencial con posibles oscilaciones sinusoidales. Por el contrario la FAP de un proceso AR estacional presenta un corte brusco. Uriel (1985) menciona lo cual dificultoso es calcular, sin embargo dice se puede complementar con Hamilton y Watts (1978).

Como resumen de modelos multiplicativos estacionales puede construirse la Tabla N^o 03

Tabla N^o 03 FUNCION DE AUTOCORRELACION Y
AUTOCORRELACION PARCIAL

Modelos	FAC	FAP
AR(p)xAR(P) _s	Decaimiento exponencial y/u oscilaciones sinusoidales atenuadas.	Se anula para $k > p+sP$
MA(q)xMA(Q) _s	Se anula para $k > q+sQ$	Decaimiento exponencial y/u oscilaciones sinusoidales atenuadas.

En cuanto a la estimación de estos modelos, SARMA, son válidos algunos métodos estudiados para modelos ordinarios. Sin embargo es necesario hacer algunas consideraciones.

Así un proceso mixto, con componente estacional y no estacional, no se puede estimar por métodos lineales, a diferencia que ocurría con los procesos autorregresivos no estacionales.

Por otra parte, en la estimación condicionada de los proceso autorregresivos estacionales (SAR) hay que tener presente que se perderán D^* s observaciones iniciales. La estimación no condicionada requiere del cálculo hacia atrás de muchos más términos que en los procesos no estacionales; por este motivo, en los modelos estacionales las diferencias entre una estimación condicionada y una estimación no condicionada puede llegar a ser considerable, siendo preferible, como es lógico, utilizar este último procedimiento.

Para la validación de los modelos estacionales se aplican las mismas técnicas para los modelos ordinarios. Si el modelo seleccionado subsiste en sus residuos algún componente de carácter estacional. Cuando la FAR presenta valores significativos en algún retardo múltiplo del período estacional será indicativo que el modelo seleccionado no explica completamente el componente estacional. Una forma alternativa para detectar si aún subsisten periodicidades sin explicar es mediante el periodograma acumulativo normalizado tal como propone Box y Jenkins (1976).

El modelo SARIMA(P,D,Q) (Seasonal Autorregressive Integrated Moving Average) es una generalización del proceso ARIMA, que lleva en consideración la dependencia de observaciones afectadas por periodos de tamaño s. Así se tiene el modelo:

$$Z_t = \phi_1 Z_{t-s} + \dots + \phi_p Z_{t-sp} + e_t - \delta_1 e_{t-s} - \dots - \delta_q e_{t-sq} \quad (3.79)$$

donde e_t es proceso ruido blanco y Z_t tiene media cero y es estacionario.

Denotando la diferencia estacional por $W_t = Z_t - Z_{t-s}$, o $W_t = (1 - B^s)Z_t$ para permitir series estacionarias, llegando al modelo:

$$(1 - \phi_1 B^s - \dots - \phi_p B^{sp})(1 - B^s)^D Z_t = (1 - \delta_1 B^s - \dots - \delta_q B^{sq})e_t \quad (3.80)$$

donde D permite diferencia estacional de mayor orden.

El modelo de encima contiene defectos por no contemplar relación entre estaciones, o sea las autocorrelaciones intermedias de las estaciones serian cero, el que no es muy razonable en la mayoría de los casos.

Para modelar esas interacciones entre atrasos unitarios se supone que el ruido e_t siga un modelo ARMA.

Tengase finalmente el modelo general multiplicativo:

$$(1 - \phi_1 B^s - \dots - \phi_p B^{sp})(1 - \theta_1 B - \dots - \theta_q B^q)(1 - B)^d(1 - B^s)^D Z_t = (1 - \delta_1 B^s - \dots - \delta_p B^{sq})(1 - \theta_1 B - \dots - \theta_q B^q)a_t \quad (3.81a)$$

en una notación más compacta:

$$\phi_p(B^s)\phi_p(B)\nabla^d\nabla^D Z_t = \delta_q(B^s)\theta_q(B)a_t + \mu \quad (3.81b)$$

El modelo es multiplicativo en sentido que la serie final pasa por dos filtros ARMA, uno simple y otro estacional.

La estimación de parámetros de un modelo SARIMA no es muy diferente, de un proceso ARIMA. La diferencia $W_t = (1 - B)^d(1 - B^s)^D Z_t$ constituye un proceso ARMA($p+sP, q+sQ$) donde algunos de los coeficientes son cero. Expandiéndose las relaciones multiplicativas de los parámetros del modelo para un dato conjunto p, P, q, Q, s se relacionan con los valores calculados para el modelo ARMA, hallándose los valores para $\phi(\square), \phi(\square), \delta(\square)$ y $\theta(\square)$.

3.7 DIAGNOSTICO Y REVISION DEL MODELO

Después de estimar un modelo tentativo para los datos existentes es necesario verificar la adecuabilidad del modelo a esos mismos datos. Este procedimiento es denominado de diagnóstico y sirve también para orientar la modificación del modelo cuando es necesario. El diagnóstico se basa en los llamados residuos.

3.7.1 OBTENCION DE LOS RESIDUOS

Supongase el modelo ARMA(p, q) obtenido a partir de 3.81, donde $W_t = \nabla^d\nabla^D Z_t$

$$\phi(B)W_t = \theta(B)a_t \quad (3.82)$$

fuera ajustado a las observaciones $y_n = [y_1, \dots, y_n]$ con estimadores $\bar{\phi}$ y $\bar{\theta}$.

La cantidad \bar{a}_t en 3.83 es conocida como residuo:

$$a_t = \bar{\theta}^{-1}(B)\bar{\phi}(B)y_t \quad (3.83)$$

Si este modelo estuviera correcto, entonces los residuos \bar{a}_t constituyen un proceso de ruido blanco y como visto antes, las autocorrelaciones muestrales de esos residuos deben ser aproximadamente cero para "lags" mayores que cero. En la práctica, se podrán, como los coeficientes de 3.83 son estimados entonces la secuencia \bar{a}_t no es sucesión independiente e idénticamente distribuida Brockwell y Davis (1987, p. 299), y la distribución de las correlaciones muestrales de $\bar{a}_t(\bar{\phi}, \bar{\theta})$ no es la misma de $a_t(\phi, \theta)$ ahora se podrá esperar que las correlaciones muestrales de \bar{a}_t podrán dar información valiosa respecto a la adecuabilidad del modelo. Lo que sucede es que, como aportan Granger y Newbold (1977); Brockwell y Davis (1987) y otros, la desviación estándar de $r_{\kappa}(\bar{a}(\bar{\phi}, \bar{\theta}))$, puede ser considerablemente menor que la desviación estándar del estimador con el valor real de los parámetros si esos fueran conocidos, ahora el límite de $\pm 2n^{-1/2}$ para así considerar una autocorrelación como cero aún continúa válido para grandes atrasos. Para atrasos pequeños (menor que 6 para modelos no muy complicados

(Granger y Newbold, 1977), el intervalo de confianza puede caer hasta la frontera (meta) con relación al intervalo que sería si se conoce los parámetros reales (Box y Jenkins, 1970). Como la variación del intervalo de confianza es función complicada del modelo, Granger y Newbold (1977) sugieren que se use los intervalos de confianza estándar observándose, por lo menos para $k < 6$, los valores de autocorrelación se pasan los límites del intervalo de confianza, fijándose más próximo a cero.

En el cálculo de los residuos, aparece la ecuación de las condiciones iniciales, ya mencionada en la sección de estimación de máxima verosimilitud. Normalmente se utiliza los mismos criterios ya mencionados, mismo porque al término de la última iteración en el cálculo de estimado por máxima verosimilitud los residuos ya están establecidos.

3.7.2 USO DE LOS RESIDUOS PARA MODIFICAR EL MODELO

Supone que el análisis de los residuos indican la no adecuabilidad del modelo:

$$\phi(B)Z_t = \theta(B)a_t \quad (3.84)$$

y que el modelo identificado y estimado para los residuos fue:

$$\phi'(B)a_t = \theta'(B)a'_t \quad (3.85)$$

sustituyéndose el modelo 3.85 en 3.84 tenemos el nuevo

modelo tentativo es:

$$\phi'(B)\phi(B)Z_{\epsilon} = \theta'(B)\theta(B)a'\epsilon \quad (3.86)$$

El nuevo modelo 3.86 será aquel modelo modificado por los residuos y debe ser el más representativo.

3.7.3 ANÁLISIS DEL PERIODOGRAMA

El periodograma pretende confrontar las posibles frecuencias en la serie para identificar cuáles pueden ser las frecuencias importantes para explicar el comportamiento de la serie. Tal es así se usa en la parte de identificación de modelos como complemento de la sec. 3.4.

Schuster en 1898 definió el periodograma de la siguiente manera:

$$\text{para: } w = 2\pi p/N \quad p = 1, \dots, (N/2 - 1)$$

se calculan:

$$a(w) = (2/N) \sum_{i=1}^n y_{\epsilon} \cos tw \quad \text{y} \quad (3.87)$$

$$b(w) = (2/N) \sum_{i=1}^n y_{\epsilon} \sin tw$$

en algún momento encontramos expresiones del tipo $a(w)$ y $b(w)$ durante el análisis armónico; método para correlacionar una serie $\{y_{\epsilon}\}$ con armónicas de frecuencia angular w .

En base a ellos calculamos la "Intensidad" de cada

frecuencia w como:

$$I(w) = [a(w)]^2 + [b(w)]^2 \quad (3.88)$$

y el gráfico $I(w)$ Vs. w se le denomina periodograma.

en estos periodogramas podemos observar algunos, saltos, picos, etc. con sus correspondientes frecuencias.

Para las observaciones de un proceso estocástico de ruido blanco a_t , $t=1,2,\dots,n$; con frecuencias f_1 ($f_1=w/2\pi$) las expresiones 3.87 y 3.88 podemos escribir, un estimador del espectro del proceso así:

$$a(f_1) = (2/n) \sum_{t=1}^n a_t \cos (2\pi i/n)t$$

y

$$b(f_1) = (2/n) \sum_{t=1}^n a_t \sen (2\pi i/n)t$$

la "intensidad" de cada frecuencia será:

$$I_{\Delta}(f_1) = (2/n) \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right)^2 + \left(\sum_{t=1}^n a_t \sen \frac{2\pi i}{n} t \right)^2 \right] \quad (3.89)$$

$0 < f_1 < 1/2$, llamado periodograma para analizar los residuos del proceso. Este estimador fue uno de los primeros a ser propuestos en los estudios iniciales al Análisis Espectral de series temporales con la finalidad de detectar periodicidades en los datos. Un pico en la frecuencia $f_1=i/n$ indica una periodicidad de periodo $1/f_1$.

Para y_t ; $t=1,2,\dots$, como variable continua, se puede tomar valores en un intervalo $[-\pi,\pi]$ ó equivalentemente a la frecuencia $f=2\pi w$, como una variable continua que puede

tomar valores en $[-1/2, 1/2]$, cuyas modificaciones y otros para el paso del análisis armónico al análisis espectral.

Puede probarse, cuando la diferenciación del espectro a_e es ruido blanco, entonces el espectro es constante e igual a $2\sigma^2_{a_e}$ en el intervalo $[0, 1/2]$.

$$P_{a_e}(f) = 2\sigma^2_{a_e}, \quad 0 \leq f \leq 1/2$$

en consecuencia:

$$P_{a_e}(f) = \begin{cases} 0 & , \quad f < 0 \\ \int_0^f P_{a_e}(g) dg = 2\sigma^2_{a_e} f & 0 \leq f < 1/2 \\ \sigma^2_{a_e} & f \geq 1/2 \end{cases} \quad (3.90)$$

como se muestra en la figura N° 08 el "espectro acumulado" (o función de distribución espectral).

Como la intensidad $I_{a_e}(f)$ es un estimador de $P_{a_e}(f)$ vemos que la estimación de $P_{a_e}(f_j)$ es $(1/n) \sum_{i=1}^j I_{a_e}(f_i)$ y

$$C(f_j) = \frac{\sum_{i=1}^j I_{a_e}(f_i)}{n \sigma^2_{a_e}} \quad (3.91)$$

$C(f_j)$ de 3.91 es una estimación de $P_{a_e}(f_j)/\sigma^2_{a_e}$; y es el periodograma acumulado (normalizado), como se muestra en la Fig. N° 8.b para un ruido blanco.

El gráfico $C(f_j) \times f_j$ de la Fig. N° 09 estaría representado al rededor de una recta que pasa por los puntos $(0,0)$ y $(0.5,1)$.

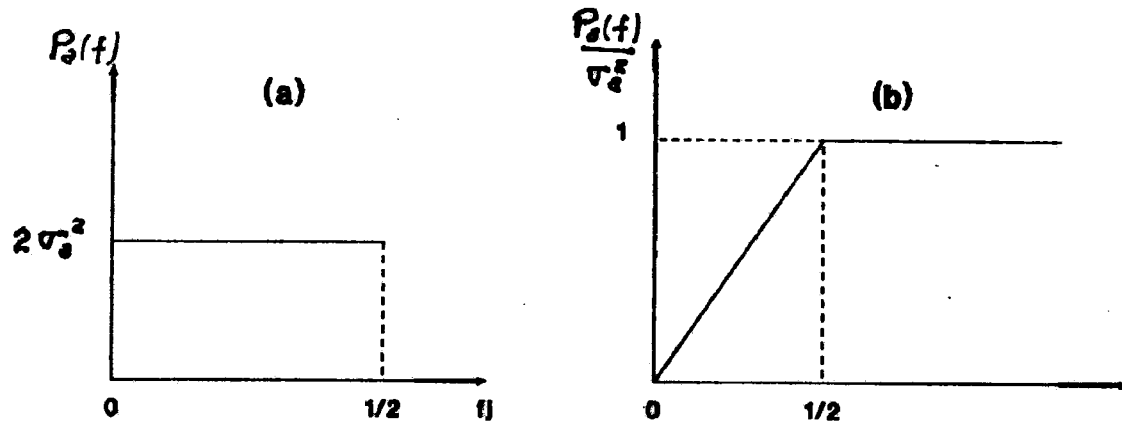


Fig. N^o 08 Espectro y Espectro acumulado del ruido blanco

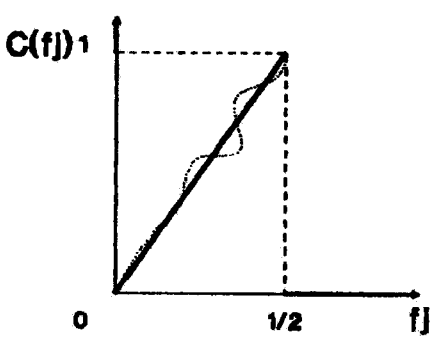


Fig.N^o 09 Periodograma acumulado Normalizado VS. Frecuencia para un ruido blanco (Valores de $C(f_i)$ VS. f_i).

Si el modelo no fuera adecuado, habran desviaciones sistematicas de esta recta. Podemos usar un test de tipo Kolmogorov-Smirnov para evaluar si las desviaciones

observadas son compatibles o no con lo que esperaríamos si a_t fuese ruido blanco.

Podemos obtener límites de confianza alrededor de la recta teórica, hallados a una distancia $\pm K(\alpha)/\sqrt{q}$ de esta recta; valores críticos de $K(\alpha)$ para algunos valores de α son dados en la tabla N^o 04; tomado de Box y Jenkins (1970).

Tabla N^o 04:

Coeficientes para los cálculos del límite de confianza para los test del periodograma acumulado.

α	0,01	0,05	0,10	0,25
$K(\alpha)$	1,63	1,36	1,22	1,02

aquí, $q = [(n-1)/2]$ si n es impar; igual a $(n-2)/2$ para n par. Si el gráfico $C(f_j)$ cruzan los límites paralelos en una proporción mayor que $100\alpha\%$ de veces, los residuos no serán aleatorios.

En la figura N^o 10 se representa el comportamiento de un periodograma acumulado de una serie Z_t y para una variable a_t ruido blanco, con su límite de confianza; para mayor información revisar Box y Jenkins (1970) p. 294.

Por ejemplo, el intervalo que se muestra en la Fig. 10 es para $n=224$; y un nivel de confianza para α 5% y 25%: de la tabla 4 $K(\alpha)$ tenemos $K(0.05)=1,36$ y $K(0.25)=1,02$; y el valor de $q=(224-2)/2 = 111$; entonces el intervalo será:

$$\begin{aligned} \pm 1,36/\sqrt{111} &= \pm 0,13 & \alpha &= 5\% \\ \pm 1,02/\sqrt{111} &= \pm 0,10 & \alpha &= 25\% \end{aligned}$$

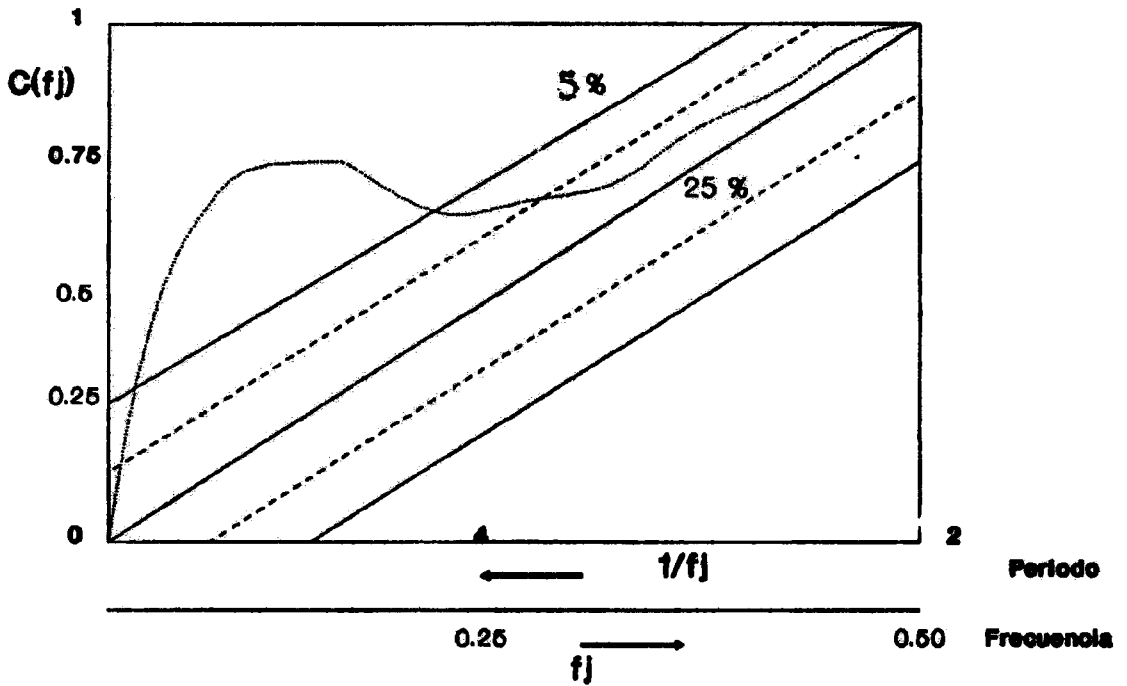


Fig. N° 10 Periodograma acumulado de la serie y límites de confianza.

3.7.4 PRUEBA DE BOX Y PIERCE.

Box y Pierce (1970) sugirieron una prueba para las autocorrelaciones en los residuos estimados; a pesar de no presentar diferencias ó discrepancias específicas del comportamiento del ruido blanco, y puede indicar si estos valores se presentan muy altos, con relación a las autocorrelaciones residuales (r_k).

Si el modelo fuese adecuado, el estadístico: $QB = n \sum_{k=1}^k r_k^2$ tiene una distribución Chi-cuadrado (X^2) con $k-p-q-P-Q$ grados de libertad (G.L.) para el modelo SARMA(p,q)x(P,Q). La hipótesis del ruido blanco es rechazada para los valores grandes de QB. En general basta tomar las primeras 20 ó 25 r_k .

IV. MODELOS AUTOREGRESIVO INTEGRADOS CON MEDIA MOVIL
(ARIMA)

4.1 SERIE NO ESTACIONARIA

Los modelos vistos hasta ahora son apropiados para describir series estacionarias. Esto es que las series se desenvuelven en el tiempo alrededor de una media constante. Probablemente las series encontradas en la práctica nunca son estacionarias, con excepción tal vez para diversas formas de "ruidos", subyacentes a ciertos fenómenos físicos.

Las series económicas, por ejemplo, son no estacionarias, y cuando se diferencian recién tornan estacionarias, las series como de precios, de PNB, etc. así por ejemplo Z_t es no estacionaria, más

$$W_t = Z_t - Z_{t-1} = (1-B)Z_t = \hat{Z}_t \quad (4.1)$$

es estacionaria.

Una serie puede presentar de varias formas de no estacionariedad. Considerando un modelo AR(1)

$$(1 - \phi B)\bar{Z}_t = a_t \quad (4.2)$$

Vimos que las condiciones de estacionariedad es $|\phi| < 1$. Si $\phi=1$ obtenemos un proceso no estacionario $\bar{Z}_t = \bar{Z}_{t-1} + a_t$ y es fácil verificar que si $|\phi| > 1$ el proceso 4.2 "explote", a medida que t aumenta.

Los modelos a estudiar, son apropiados para presentar series cuyo comportamiento es no explosivo, en particular series que presentan alguna homogeneidad en su "comportamiento estacionario". En el caso anterior si $\phi=1$, \bar{Z}_t es no estacionario, pero $\hat{Z}_t = a_t$ es estacionaria.

Las series Z_t tales que se toman un número finito de diferencias, d se tornan estacionarias, son llamadas no estacionarias homogeneas.

Si

$$W_t = \nabla^d Z_t$$

es estacionaria, podemos representar W_t por un modelo ARMA(p,q), o sea,

$$\phi(B)W_t = \theta(B)a_t \quad (4.3)$$

Si W_t es una diferencia de Z_t , entonces Z_t es una integral de W_t , de ahí decimos que Z_t sigue un modelo auto-regresivo-integrado-medias móviles, o modelo ARIMA,

$$\phi(B)\hat{\nabla}^d Z_t = \theta(B)a_t \quad (4.4)$$

de orden (p,d,q) y escribimos ARIMA(p,d,q), si p y q son las ordenes de $\phi(B)$ y $\theta(B)$, respectivamente.

En el modelo 4.3 todas las raices de $\phi(B)$ están fuera del círculo unitario. Escribir (4.4) es equivalente a escribir:

$$\varepsilon(B)Z_t = \theta(B)a_t \quad (4.5)$$

donde $\varepsilon(B)$ es un operador auto-regresivo no estacionario, de orden $p+d$, con d raices iguales a uno (sobre el círculo

unitario) y las restantes p fuera del círculo unitario, o sea:

$$\varepsilon(B) = \phi(B)\Delta^d = \phi(B)(1 - B)^d \quad (4.6)$$

Obsérvese que es indiferente escribir $\varepsilon(B)Z_t$ o $\varepsilon(B)\bar{Z}_t$, pues $\Delta^d Z_t = \Delta^d \bar{Z}_t$, para $d > 1$.

Por tanto el modelo 4.4 supone que las d -ésimas diferencias de la serie Z_t puede ser representada por un modelo ARIMA, estacionario e invertible. En la mayoría de los casos usadas, $d=1$ o $d=2$, corresponde a dos casos interesantes y comunes de no estacionariedad homogénea:

- a) Series no estacionarias cuando el nivel: oscila al rededor de un nivel medio durante algún tiempo y después salta para otro nivel temporalmente. Para tener la estacionariedad es suficiente tomar una diferencia; este es el caso típico de series económicas;
- b) Series no estacionarias cuando la inclinación: oscila en una dirección por algún tiempo y después cambian de dirección temporalmente. Para tener estacionariedad es necesario tomar la segunda diferencia.

Casos particulares del modelo (4.4)

i) ARIMA(0,1,1): $\Delta Z_t = (1 - \theta B)a_t$

es un caso importante es también llamado modelo integrado de media móvil, IMA(1,1),

$$Z_t = Z_{t-1} + a_t - \theta a_{t-1} \quad (4.7)$$

ii) ARIMA(1,1,1): $(1 - \phi B)\Delta Z_t = (1 - \theta B)a_t$

iii) $ARIMA(p,0,0) = AR(p)$; $ARIMA(0,0,q) = MA(q)$;

$ARIMA(p,0,q) = ARMA(p,q)$

Tomar diferencias podrá ser suficiente para alcanzar la estacionariedad, en algunos casos como el de las series económicas. Algunas veces es necesario considerar alguna transformación no lineal de Z_t y $Z_t^* = \log Z_t$ y muchas veces será suficiente para obtener homogeneidad. De este modo un procedimiento usual en series temporales es modelar:

$$\Delta \log Z_t = \log Z_t - \log Z_{t-1} \quad (4.8)$$

El análisis exploratorio de los datos es muy importante en los modelos ARIMA que se verá posteriormente como complemento del modo general de transformación de datos que puede ser representado por $Z_t^{(b)}$, donde b es un vector de parámetros que definen la transformación.

Por otro lado, en el área de series de tiempo se utilizan gráficos (Morettin, 1987) para detectar si la dispersión de la variable esta relacionada con el valor promedio de la misma, caso en el cual es recomendable aplicar una transformación no lineal (TNL).

La razón principal para efectuar transformaciones es estabilizar la variancia (precisamente, hacer que los residuos del modelo ajustado tengan una variancia constante).

Para tener idea del tipo de transformación que es adecuada se puede utilizar un gráfico que traza en el eje de las abscisas medias de subconjunto de observaciones de la serie y en el eje de las ordenadas las amplitudes de cada uno de estos subconjuntos; si Z_{t_1}, \dots, Z_{t_k} es un subconjunto con k observaciones, calculamos para ello:

$$Z_t = (1/k) \sum_{i=1}^k Z_{t_i}, \quad (4.9)$$

$$w = \max(Z_{t_i}) - \min(Z_{t_i}),$$

que son medidas de posición y variabilidad, respectivamente; el par (\bar{Z}, w) será un punto del gráfico. El número de elementos en cada sub serie puede ser igual al período, en el caso de las series estacionales.

Si w es independiente de \bar{Z} , obtendremos puntos espaciados alrededor de una recta paralela al eje de las abscisas y en este caso no habrá necesidad de transformación. Si w es directamente proporcional a \bar{Z} , la transformación logarítmica es apropiada. Una clase general de transformación que puede ser utilizada es el de Box-Cox, definida por:

$$Z_t(b) = \begin{cases} \frac{Z_t^b - c}{b}, & b \neq 0 \\ \log Z_t, & b = 0, \end{cases} \quad (4.10)$$

donde b y c son parámetros a ser estimados. La fig. N^o 11 extraída de Box y Jenkins (1970), da una idea de los tipos de gráficos que puede ocurrir y los respectivos valores de b .

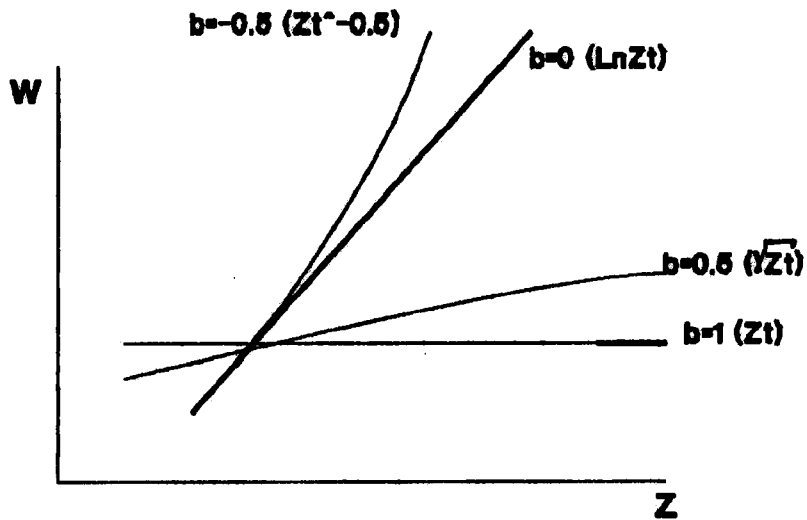


Fig. N^o 11 Amplitud Vs. media; mostrando algunos valores posibles de b ($b=r$)

Entonces la finalidad de estas transformaciones era lograr normalidad (obviamente involucra simetría) y simplicidad del modelo, así como uniformidad en la dispersión. Por simplicidad se entiende la no existencia de interacción en los modelos de Análisis de Variancia y la no inclusión de cuadrados ni productos cruzados en el caso de la regresión.

La necesidad de encontrar un b apropiado es principalmente para: lograr simetría; lograr que una relación entre dos variables sea lineal; lograr que la dispersión sea uniforme e independientemente del nivel de la variable y lograr que los efectos de dos factores sean aditivos sin interacción entre ellos. Si bien son diferentes los cálculos y gráficos hay que hacer para cada uno de estos casos.

4.2 FORMAS DEL MODELO ARIMA

Los modelos ARIMA dado en 4.4 pueden ser representados de tres formas:

- i) En términos de valores previos de Z_t y del valor actual y previo de a_t ;
- ii) En términos del valor actual y previos de a_t ;
- iii) En términos de valores previos de Z_t y los valores actuales de a_t .

Forma de la Ecuación de Diferencias

Esta es la forma usual del modelo, útil para calcular predicciones:

$$Z_t = \epsilon_1 Z_{t-1} + \epsilon_2 Z_{t-2} + \dots + \epsilon_{p+d} Z_{t-p-d} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (4.11)$$

donde:

$$\epsilon(B) = 1 - \epsilon_1 B - \epsilon_2 B^2 - \dots - \epsilon_{p+d} B^{p+d}.$$

Forma de Choques Aleatorios

Una forma para asi calcular la variancia de los errores de prediccion es:

$$\begin{aligned}
Z_t &= a_t + w_1 a_{t-1} + w_2 a_{t-2} + \dots & (4.12) \\
&= w(B)a_t
\end{aligned}$$

de esta ecuacion obtenemos:

$$\epsilon(B)Z_t = \epsilon(B)w(B)a_t$$

y usando 4.5 se consigue:

$$\epsilon(B)w(B) = \theta(B). \tag{4.13}$$

Luego, los pesos w_j de la ecuacion 4.12 pueden ser obtenidos de 4.13 identificándose los coeficientes de B , B^2 , etc.:

$$(1 - \epsilon_1 B - \dots - \epsilon_{p+d} B^{p+d})(1 + w_1 B + w_2 B^2 + \dots) = 1 - \theta_1 B - \dots - \theta_q B^q. \tag{4.14}$$

Forma invertida

de 4.12 obtenemos que $w^{-1}(B)Z_t = a_t$ ó

$$\pi(B)Z_t = \left[1 - \sum_{j=1}^{\infty} \pi_j B^j \right] Z_t = a_t \tag{4.15}$$

se consigue que:

$$\epsilon(B)Z_t = \theta(B)a_t = \theta(B)\pi(B)Z_t,$$

de donde obtenemos la relacion

$$\epsilon(B) = \theta(B)\pi(B). \tag{4.16}$$

Por tanto los pesos π_j pueden ser obtenidos de 4.16 conociendo los operadores $\epsilon(B)$ y $\theta(B)$. Es fácil ver que los pesos π_j en 4.15 suman uno, esto es $\sum_{j=1}^{\infty} \pi_j = 1$.

4.3 TERMINO CONSTANTE DEL MODELO

En el modelo ARIMA(p,d,q)

$$\phi(B)W_t = \theta(B)a_t \quad (4.17)$$

$$W_t = \Delta^d Z_t,$$

Un término constante fue omitido, implicando que $E(W_t) = \mu_w = 0$. El modelo 4.17 puede escribirse al que podríamos llamar de tendencia estocástica, en sentido que el proceso no es estacionario y nula de nivel y/o inclinación en recorrer el tiempo. La tendencia (o no estacionariedad) estocástica es caracterizada por la existencia de ceros de $\phi(B)$ sobre el círculo unitario.

Además de esta no estacionariedad estocástica, muchas series temporales pueden presentar una tendencia determinística, como en el tratamiento de modelos de descomposición. En particular podemos tener Z_t como la suma de un polinomio y de un proceso ARIMA(p,d,q), esto es:

$$Z_t = \sum_{j=0}^m \beta_j t^j + \frac{\theta(B)}{\Delta^d \phi(B)} a_t \quad (4.18)$$

en 4.18, $Z_t = T_t + Y_t$, donde Y_t sigue un modelo ARIMA(p,d,q), esto es, $\phi(B)\Delta^d Y_t = \theta(B)a_t$. Siguiendo que Z_t es no estacionario si $m > 0$ y/o $d > 0$.

tomando d diferencias, tenemos:

$$\begin{aligned} \Delta^d Z_t &= \theta_w + [\theta(B)/\phi(B)]a_t, \text{ si } m=d \\ &= [\theta(B)/\phi(B)]a_t, \text{ si } m < d, \end{aligned} \quad (4.19)$$

donde $\theta_0 = \beta_{ad}$, se obtiene entonces una serie estacionaria. Esto significa que podemos incluir una tendencia polinomial determinística de grado d en el modelo, bastando incrementar una constante θ_0 :

$$\epsilon(B)Z_t = \theta_0 + \theta(B)a_t \quad (4.20)$$

Cuando $m > d$, podemos obtener un modelo no estacionario, tomándose las d diferencias, debido a la presencia de una tendencia determinística; siendo en este caso ($m > d$) se toma m diferencias, obtendremos un proceso estacionario, más no invertible. Para otros detalles, ver Pierce (1979).

Si $\theta_0 \neq 0$, obtenemos:

$$E(W_t) = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p}, \quad (4.21)$$

y si $\bar{W}_t = W_t - E(W_t)$, tenemos $\phi(B)\bar{W}_t = \theta(B)a_t$.

En que sigue, cuando $d > 0$, suponemos $\mu_w = 0$ es por tanto $\theta_0 = 0$.

4.4 MODELOS ESTACIONALES

Los modelos de Box y Jenkins estacionales, llamados modelo estacional auto-regresivo integrado de media móvil (SARIMA), de la forma:

$$\phi(B)\phi(B^s)(1-B)^d(1-B^s)^p(z_t - \mu) = \theta(B)\delta(B^s)a_t \quad (4.24)$$

las variables y parámetros que tenemos son:

s = es el periodo estacional ($s = 12$ para valores mensuales).

B = operador de atraso definido por:

$$BZ_t = Z_{t-1} \quad \text{y} \quad B^s Z_t = Z_{t-s}$$

$\phi(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps}$; es el operador auto-regresivo (AR) estacional de orden P , tal que las raíces de $\phi(B^s)=0$ están fuera del círculo unitario para estacionariedad estacional y $\phi_i, i=1,2,\dots,P$ sea los parámetros AR estacionales;

$(1 - B^s)^D = \nabla_s^D$ es el operador diferencia estacional de orden D para producir estacionariedad estacional del D -ésimo dato diferenciado, usualmente $D = 0, 1$ o 2 ;

$W_t = \nabla_s^D \nabla_s^d Z_t$, es la serie estacionaria formada por diferenciación de la serie Z_t y $n=N-d-sD$ es el número de términos de la serie W_t ;

$\delta(B^s) = 1 - \delta_1 B^s - \dots - \delta_q B^{qs}$; es el operador MA estacional de orden Q tal que las raíces de $\delta(B^s)=0$ están fuera del círculo unitario para su invertibilidad y $\delta_i, i=1,2,\dots,Q$ los parámetros MA estacionales.

Box y Jenkins (1976) usa la notación $(p,d,q) \times (P,D,Q)_s$ para representar los modelos SARIMA de la ecuación 4.24 el primer parámetro corresponde al orden del operador no estacional, sometido a la notación (p,d,q) , visto anteriormente; al segundo parámetro $(P,D,Q)_s$ la parte estacional. Cuando un modelo es estacionario y no requiere

diferencias (esto es $D=0$ y $d=0$) es referido como un modelo ARMA estacional o SARMA (estacional auto-regresivo de medias móviles) siendo usado la notación $(p,q) \times (P,Q)$. Si este modelo ARMA es no estacional, la notamos por ARMA(p,q) visto anteriormente con (p,q) que indican las ordenes de los operadores AR y MA respectivamente.

como se trata de datos mensuales, $s=12$; entonces la diferencia estacional $\hat{\Delta}_{12} = 1 - B^{12}$ es tal;

$$\hat{\Delta}_{12} Z_t = Z_t - Z_{t-12} \quad (4.25)$$

y $\hat{\Delta}_{12}^D = (1 - B^{12})^D$, D indicando el número de "diferencias estacionales".

Del mismo modo tendríamos un modelo análogo a 4.24 que relaciona un mes determinado:

$$\phi(B^{12}) \hat{\Delta}_{12}^D Z_{t-1} = \delta(B^{12}) a_{t-1} \quad (4.26)$$

donde, por ahora los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ serían los mismo que en 4.24.

Como se ha visto en la Sec. 3.6.2 un modelo estacional de media móvil puro, SMA(Q), es de la forma:

$$Z_t = a_t - \delta_1 a_{t-12} - \dots - \delta_Q a_{t-12Q} \quad (4.27)$$

y su FAC será no nula sometida a los "lags" $12, 24, \dots, 12Q$. correspondientes a la Sec. 3.3.2 para p_j , $j=1, \dots, q$, tenemos:

$$\begin{aligned} \rho_{12} &= \frac{-\delta_1 + \delta_1\delta_2 + \dots + \delta_{a-1}\delta_a}{1 + \delta^2_1 + \dots + \delta^2_a} \\ &\vdots \\ \rho_{12a} &= \frac{-\delta_a}{1 + \delta^2_1 + \dots + \delta^2_a} \end{aligned} \quad (4.28)$$

así un SMA(1) será de la forma:

$$Z_t = a_t - \delta_1 a_{t-12} \quad (4.29)$$

tendrá correlación no nula sometida a "lag" 12; y un modelo auto-regresivo estacional puro, SAR(P), es de la forma:

$$Z_t = \phi_1 Z_{t-12} + \dots + \phi_P Z_{t-12P} + a_t \quad (4.30)$$

la función de auto correlación no nulas sometidas a los "lags" múltiplos de 12. y un modelo SAR(1) será:

$$Z_t = \phi_1 Z_{t-12} + a_t \quad (4.31)$$

La FAC dada por:

$$\begin{aligned} \rho_{12} &= \phi_1 \\ \rho_{24} &= \phi^2_1 \\ &\vdots \\ \rho_{12j} &= \phi^j_1, \quad j=0,1,2,\dots \end{aligned} \quad (4.32)$$

De modo general, la estructura de correlación del proceso W_t , siguiendo un modelo SARMA y análogo al de un proceso ARMA no estacional, con correlaciones no nulas sometidas a los "lags" 12, 24, 36, etc. como se describió en la sección 3.6.2.

Observándose también, que el modelo (4.31) será estacionario si $|\phi_1| < 1$, y el efecto estacional es transitorio y se va amortiguando. Del mismo modo (4.29) es invertible si $|\delta_1| < 1$.

En principio no hay ninguna dificultad adicional en la identificación, estimación y verificación de modelos estacionales, la diferencia es que tenemos que diferenciar la serie con respecto a \hat{z} y \hat{z}_{12} (como tenemos nuestra serie mensual con período $s=12$) a fin de producir estacionariedad, obteniéndose con d y D diferencias ordinarias y estacionales, respectivamente.

Las FAC y FAP cuyos procedimientos se describe en 3.6.2 son aplicables para la variable W_t que es estacionaria con d y D diferencias.

La estimación de los valores de los parámetros identificados utilizando estimación por máxima verosimilitud de manera análoga a la efectuada en 3.5.4.

Finalmente, para verificar el modelo propuesto, si es adecuado, utilizaremos los test de autocorrelación residual (FAR), Box y Pierce, periodograma acumulado, etc. tales como se describió en la Sec. 3.7.

Las predicciones podrán calcularse para un modelo estacional multiplicativo de modo análogo al modelo $ARIMA(p,d,q)$, utilizando básicamente los procedimientos de la sec. 4.5.

4.5 PREDICCIÓN CON MODELOS ARIMA

Como sabemos que estamos interesados en conocer un valor Z_{t+h} , $h > 1$; si tenemos observaciones $\dots, Z_{t-2}, Z_{t-1}, Z_t$, hasta el instante t , que es llamado origen de las predicciones como se puede ver Fig. N° 12.

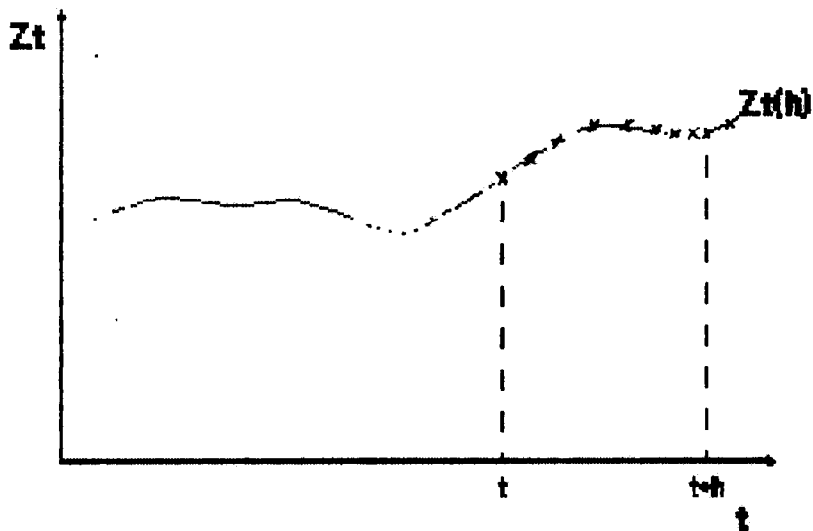


Fig. N° 12 Observaciones de una serie temporal con predicciones de origen t y horizonte h .

Si tenemos el modelo ARIMA (p,d,q) descrito por las ecuaciones 4.11, 4.12 y 4.15 con $t+h$ en lugar de t , tenemos:

a) La ecuación de diferencias,

$$Z_{t+h} = \epsilon_1 Z_{t+h-1} + \dots + \epsilon_{p+d} Z_{t+h-p-d} - \theta_1 a_{t+h-1} - \theta_2 a_{t+h-2} - \dots - \theta_q a_{t+h-q} + a_{t+h} \quad (4.33)$$

b) En forma de choques aleatorios:

$$Z_{t+h} = \sum_{j=-\infty}^{t+h} w_{t+h-j} a_j = \sum_{j=0}^{\infty} w_j a_{t+h-j} \quad (4.34)$$

donde $w_0=1$ y los demás pesos w_j son obtenidos de 4.13.

c) Forma invertida:

$$Z_{t+h} = \sum_{j=1}^{\infty} \pi_j Z_{t+h-j} + a_{t+h} \quad (4.35)$$

donde los pesos π_j son obtenidos de 4.16.

4.5.1 CALCULO DE LA PREDICCION DEL ERROR CUADRADO MEDIO MINIMO (EQM).

Como es de suponer que $\bar{Z}_t(h)$ sea una función de las observaciones al instante t , Z_t , Z_{t-1} , Z_{t-2}, \dots , consecuentemente, por 4.34 será una función de a_t , a_{t-1} , a_{t-2}, \dots , y suponiendo que $\bar{Z}_t(h)$ sea una función lineal, se indica la mejor predicción por:

$$\bar{Z}_t(h) = w_{h+1}^* a_t + w_{h+2}^* a_{t-1} + w_{h+3}^* a_{t-2} + \dots \quad (4.36)$$

queremos determinar los pesos w_j^* que minimicen el EQM de la predicción y este es dado por:

$$E[Z_{t+h} - \bar{Z}_t(h)]^2 = E \left[\sum_{j=0}^{\infty} w_j a_{t+h-j} - \sum_{j=0}^{\infty} w_{t+j}^* a_{t-j} \right]^2$$

usando 4.34 y 4.36; notamos que la primera sumatoria puede

se escrita como $\sum_{j=-h}^{\infty} w_{t+j} a_{t-j}$, vemos que el error de la

predicción es dado por:

$$e_t(h) = Z_{t+h} - \bar{Z}_t(h) = \\ = w_0 a_{t+h} + w_1 a_{t+h-1} + \dots + w_{h-1} a_{t+1} - \sum_{j=0}^{\infty} (w_{h+j} - w_{h+j}^*) a_{t-j} \quad (4.37)$$

por tanto;

$$E[e_t(h)]^2 = (1 + w^2_1 + \dots + w^2_{h-1}) \sigma^2_a + \sum_{j=0}^{\infty} (w_{h+j} - w_{h+j}^*)^2 \sigma^2_a \quad (4.38)$$

y debido a que los a_t , son no correlacionados, se sigue que 4.38 es minimizado si: $w_{h+j}^* = w_{h+j}$.

Entonces la predicción de EQM mínimo es dada por:

$$\bar{Z}_t(h) = w_h a_t + w_{h+1} a_{t-1} + \dots = \sum_{j=0}^{\infty} w_{h+j} a_{t-j} \quad (4.39)$$

y el error de la predicción de 4.37 dada por:

$$e_t(h) = a_{t+h} + w_1 a_{t+h-1} + \dots + w_{h-1} a_{t+1} \quad (4.40)$$

también:

$$Z_{t+h} = e_t(h) + \bar{Z}_t(h) \quad , \quad h \geq 1 \quad (4.41)$$

vamos a denotar por:

$$[Z_{t+h}] = E[Z_{t+h} / Z_t, Z_{t-1}, \dots] \quad (4.42)$$

De estas ecuaciones tenemos las siguientes conclusiones:

a) $\bar{Z}_t(h) = [Z_{t+h}]$; usando 4.41 o sea la predicción de EQM, es la esperanza condicional de Z_{t+h} , dadas las observaciones pasadas de la serie.

b) de 4.40, tenemos $E[e_t(h)] = 0$ y la variancia del error de predicción es:

$$V(h) = (1 + w^2_1 + w^2_2 + \dots + w^2_{h-1}) \sigma^2_a \quad (4.43)$$

c) el error de predicción a un paso es:

$$e_t(1) = Z_{t+1} - \bar{Z}_t(1) = a_{t+1} \quad (4.44)$$

lo que nos dice que los errores de predicción a un paso no están correlacionados.

d) En tanto los errores de predicción para intervalos de tiempo mayores que uno serán correlacionados, lo mismo acontece con los errores de predicción para el mismo horizonte h , de diferentes orígenes t y $t-j$.

4.5.2 FORMAS BÁSICAS DE LA PREDICCIÓN Y ECUACION DE PREDICCIÓN.

4.5.2.1 Formas Básicas

El pronóstico $\bar{Z}_t(h)$ podemos expresar al igual que las diversas formas que se presentan los modelos ARIMA.

a) Utilizando la ecuación de diferencias; tomando esperanza condicional en 4.33 obtenemos:

$$\begin{aligned} \bar{Z}_t(h) = & \varepsilon_1[Z_{t+h-1}] + \dots + \varepsilon_{p+q}[Z_{t+h-p-q}] - \theta_1[a_{t+h-1}] - \dots - \\ & - \theta_q[a_{t+h-q} + [a_{t+h}]] \end{aligned} \quad (4.45)$$

para $h \geq 1$. Aquí debemos utilizar lo siguiente:

$$\begin{aligned} [Z_{t+k}] &= \bar{Z}_t(k) & k > 0 \\ [Z_{t+k}] &= Z_{t+k} & k \leq 0 \\ [a_{t+h}] &= 0 & k > 0 \\ [a_{t+k}] &= a_{t+k} & k \leq 0 \end{aligned} \quad (4.46)$$

b) Predicción utilizando forma de choque aleatorios de 4.34, tenemos:

$$\bar{Z}_t(h) = w_1[a_{t+h-1}] + w_2[a_{t+h-2}] + \dots + w_{h-1}[a_{t+1}] + w_h[a_t] + \dots + [a_{t+h}] \quad (4.47)$$

c) Utilizando la forma invertida, de la ecuación 4.35, obtenemos:

$$\bar{Z}_t(h) = \sum_{j=1}^{\infty} \pi_j [Z_{t+h-j}] + [a_{t+h}] \quad (4.48)$$

4.5.2.2 Ecuación de Predicción

De las ecuaciones básicas (a) la ecuación de predicción, considerada como una función de h, con origen t, satisface la ecuación de diferencias:

$$\bar{Z}_t(h) = \sum_{i=1}^{p+d} \varepsilon_i \bar{Z}_t(h-i) \quad ; \quad h > q \quad (4.49)$$

$$\varepsilon(B) \bar{Z}_t(h) = (1-B)^d \phi(B) \bar{Z}_t(h) = 0 \quad ; \quad h > q \quad (4.50)$$

con $\varepsilon(B)$ operando sobre h; mayor detalles revisar Box y Jenkins (1970) y Morettin (1987).

4.5.3 ACTUALIZACION DE LAS PREDICCIONES

Vamos a calcular las predicciones de Z_{t+h+1} a partir de dos orígenes:

$$a) \quad t+1 : \bar{Z}_{t+1}(h) = w_h a_{t+1} + w_{h+1} a_t + \dots \quad (4.51)$$

$$b) \quad t : \bar{Z}_t(h+1) = w_{h+1} a_t + w_{h+2} a_{t-1} + \dots \quad (4.52)$$

restando 4.52 de 4.51 tenemos que:

$$\bar{Z}_{t+1}(h) = \bar{Z}_t(h+1) + w_h a_{t+1} \quad (4.53)$$

Así las predicciones de Z_{t+n+1} , hecho en el instante t , puede ser actualizado cuando un nuevo dato Z_{t+1} , es observado. De este modo, haremos la predicción de Z_{t+n+1} en el origen $t+1$, adicionandose a $\bar{Z}_t(h+1)$ un múltiplo del error de predicción:

$$a_{t+1} = Z_{t+1} - \bar{Z}_t(1)$$

4.5.4 INTERVALOS DE CONFIANZA

Calcular la variancia del error de predicción que es dada por 4.43. Para poder determinar un intervalo de confianza para Z_{t+n} , será necesario hacer una suposición adicional para los residuos; o sea además de suponer que $E(a_t)=0$, $\text{Var}(a_t)=\sigma^2_a \quad \forall t$, y $E(a_t a_s)=0$, $t \neq s$, suponemos que $a_t \sim N(0, \sigma^2_a)$, para cada t .

Dado los valores pasados y presentes de la serie, Z_t, Z_{t-1}, \dots , la distribución condicional de Z_{t+n} será $N(\bar{Z}_t(h), V(h))$.

Luego:

$$U = \frac{Z_{t+n} - \bar{Z}_t(h)}{[V(h)]^{1/2}} \sim N(0,1) \quad (4.54)$$

y fijando el coeficiente de confianza γ podemos encontrar un valor μ_γ tal que $P[-\mu_\gamma < U < \mu_\gamma] = \gamma$, o sea con probabilidad γ , (nivel de confianza γ);

$$\bar{Z}_t(h) - \mu_\gamma [V(h)]^{1/2} \leq Z_{t+n} \leq \bar{Z}_t(h) + \mu_\gamma [V(h)]^{1/2} \quad (4.55)$$

En $V(h)$, el valor de σ^2_{μ} no es conocido y es sustituido por su estimado $\bar{\sigma}^2_{\mu}$ obtenida en la fase de estimación del modelo. De este modo obtenemos:

$$\bar{Z}_t(h) - \mu_V \bar{\sigma}_{\mu} \left[1 + \sum_{j=1}^{h-1} w^2_j \right]^{1/2} \leq Z_{t+h} \leq \bar{Z}_t(h) + \mu_V \bar{\sigma}_{\mu} \left[1 + \sum_{j=1}^{h-1} w^2_j \right]^{1/2} \quad (4.56)$$

TRANSFORMACIONES Y PREDICCIONES

Si Z_t es la serie original, sea $Y_t = g(Z_t)$ una transformación (instantánea, que no incluye Z_{t-j} , $j \geq 1$) de Z_t . Una de las principales razones de hacer una transformación es que Y_t puede ser gaussiana, en este caso, la predicción óptima. (en sentido de mínimos cuadrados) es una función lineal de las observaciones.

El problema que se presenta es al obtener predicciones para Z_{t+h} , dado que tenemos un modelo para Y_t y tenemos predicciones para:

$$Y_{t+h} = g(Z_{t+h}) \quad (4.57a)$$

Una manera "ingenua" de proceder es considerar la ecuación 4.57a y sustituir pronósticos por valores futuros:

$$\bar{Y}_t(h) = g(\bar{Z}_t(h)) \quad (4.57b)$$

Después se intenta obtener $\bar{Z}_t(h)$ en función de $\bar{Y}_t(h)$ a partir de 4.57b, en particular, si $g(\cdot)$ admite inversa tenemos:

$$\bar{Z}_t(h) = g^{-1}(\bar{Y}_t(h)).$$

Por ejemplo en caso de $Y_t = \ln(Z_t)$, entonces $Z_t = e^{Y_t}$, ($Z_t = \text{Exp}(Y_t)$) y un pronóstico para Z_{t+h} será.

$$\bar{Z}_t(h) = \exp(\bar{Y}_t(h)) \quad (4.58)$$

Con esto podemos demostrar que en caso de Y_t ser gaussiana, el pronóstico óptimo es:

$$\exp[\bar{Y}_t(h) + (1/2)V_Y(h)] \quad (4.59)$$

donde $V_Y(h) = \text{Var}[e_t(h)]$, siendo $e_t(h)$ el error del pronóstico $\bar{Y}_{t+h} - \bar{Y}_t(h)$. Vemos que el procedimiento 4.58 conduce a pronósticos viciados y como consecuencia, el error cuadrado medio del pronóstico aumentará.

Si $Y_t = \ln(Z_t)$ sigue un modelo ARIMA, entonces sabemos que la distribución condicional de Y_{t+h} , dado el pasado, es $N(\bar{Y}_t(h), V_Y(h))$ y un intervalo de confianza para Y_{t+h} con un nivel de confianza del 95% será:

$$\bar{Y}_t(h) \pm 1.96[\bar{V}_t(h)]^{1/2} \quad (4.60)$$

de aquí un intervalo de confianza para Z_{t+h} , con coeficiente de confianza, 95% será:

$$\left[\exp\left[\bar{Y}_t(h) - 1.96(\bar{V}_t(h))^{1/2}\right], \exp\left[\bar{Y}_t(h) + 1.96(\bar{V}_t(h))^{1/2}\right] \right] \quad (4.61)$$

llamando a $\bar{V}_Y(h)$, es el estimado de $V_Y(h)$, con σ^2_ϵ sustituido por su estimado $\bar{\sigma}^2_\epsilon$, en el ajuste del modelo para Y_t .

V. MATERIALES Y METODOS

5.1 MATERIALES Y EQUIPOS

Los materiales y equipos usados en el presente trabajo de tesis fueron:

- 1) Serie histórica de descargas medias mensuales (en m³/s) del río Pisco, que se muestra en el Tabla N^o 05, ubicado en el Distrito de Humay, Provincia de Pisco; Región Los Libertadores Wari; Estación Letrayoc con Latitud 13°40', Long. 75°46', Alt. 630 msnm. Registrado por el Servicio Nacional de Meteorología e Hidrología (SENAMHI) correspondiente al período setiembre 1974 a agosto 1988.
- 2) Una microcomputadora AT/286-386.
- 3) Softwares; pertenecientes a los "paquetes" Análisis Estadístico "Forecast - Plus"; Time Series Processor (TSP); Statgraphics; Hoja de cálculo, etc.
- 4) Utiles de escritorio y computación.

**TABLA N° 05 Descargas Medias mensuales (m³/s) del río
Pisco.**

ANO	SET	OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO
74-75	2.003	4.053	7.505	8.386	26.670	28.670	74.190	17.530	9.880	3.590	2.380	1.810
75-76	1.654	1.583	6.606	10.080	60.281	86.931	62.199	20.023	6.767	3.463	2.520	1.751
76-77	2.428	2.558	6.324	5.725	8.424	80.178	66.461	15.790	5.706	2.653	1.977	1.877
77-78	0.990	0.797	16.727	6.903	20.522	27.775	9.894	11.296	2.574	1.440	1.119	0.716
78-79	0.363	1.177	8.325	12.585	10.455	50.679	71.229	11.256	2.548	1.590	1.158	0.758
79-80	0.800	0.609	4.240	2.087	15.640	18.140	37.196	12.500	2.225	1.180	0.550	0.712
80-81	0.610	5.021	2.338	9.765	19.316	100.070	61.914	29.747	4.231	1.337	1.020	0.809
81-82	1.851	2.943	6.594	19.520	39.612	78.789	46.932	13.304	3.555	1.918	1.464	1.187
82-83	1.440	3.718	9.556	5.982	19.125	6.789	47.032	77.726	3.746	1.124	0.685	0.773
83-84	0.625	0.580	0.564	10.220	48.120	295.328	118.983	40.910	9.495	4.149	2.059	1.495
84-85	1.182	10.592	12.355	39.850	15.099	77.199	76.718	52.518	10.777	5.972	2.069	1.873
85-86	1.918	3.412	5.545	22.735	115.845	222.370	81.003	50.074	17.492	7.111	3.414	3.164
86-87	2.861	4.282	6.310	30.926	79.438	55.644	23.403	7.936	1.614	1.337	1.293	0.720
87-88	0.972	2.295	4.678	8.363	33.753	105.264	47.568	49.697	14.984	4.910	2.149	0.863

FUENTE: Servicio Nacional de Meteorología e Hidrología (SENAMHI)

5.2 MÉTODOS Y PROCEDIMIENTOS

Convencionalmente los pasos iterativos de los modelos de Box y Jenkins son: Identificación, estimación, verificación y predicción; en el presente trabajo se

plantea una metodología descriptiva para poder tener una base más sólida del análisis de series temporales ya que en el área de Ingeniería Agrícola no es muy conocida estos tipos de modelos.

La sistemática adoptada para el tratamiento de los datos de descargas medias mensuales e identificación de los modelos conforme se ha descrito en las Sec. 3 y 4, son los siguientes:

a) Adquisición de datos mensuales por el período de 14 años (información existente).

De la serie en estudio se analizó las primeras 156 observaciones para obtener el modelo adecuado para que en base a él se obtengan los pronósticos para el año hidrológico de set.87 a ago.88, (enciso de verificación del modelo).

b) Realizar el Análisis Exploratorio de los Datos y estimación de los estadísticos mensuales, y formación de la serie anual.

Este análisis revelará la posición, dispersión y simetría existente y la necesidad de aplicación alguna transformación no lineal a los datos originales; Hoaghin D. (1982) y Seier E. (1987).

c) Estimación del correlograma de la serie mensual; para la serie en estudio.

La estimación de la función de autocorrelación (ρ_k) de los datos originales para retardos equivalentes a 1/4 del total de observaciones, como se recomienda, se desarrolla según las fórmulas y pasos de la sec. 2.4 para modelos lineales.

- d) Aplicar el suavizamiento robusto de la serie u otro método de suavizamiento; para el análisis del comportamiento de la serie a ser modelada.
- e) Estudiar otras posibles transformaciones de la serie para obtener una serie homogénea.

Si fuera necesario remover la periodicidad de la serie mediante el método no paramétrico o transformación lineal:

$$Z_{1j} = \frac{X_{1j} - \bar{X}_j}{S_j}, \quad j = 1, 2, \dots, 12$$

donde:

- Z_{1j} : serie desestacionalizada
- \bar{X}_j, S_j : media y desviación estandar de j-ésimo mes
- X_{1j} : variable transformada de Box y Cox de la serie temporal del i-ésimo año y j-ésimo mes.

- f) Estimación del correlograma de la serie Z_t y ploteo de la FAC y FAP, para la serie transformada, y determinar los órdenes del modelo Autoregresivo (p) y de media móvil (q), tal que sean considerados dentro de un intervalo de confianza.

Los cálculos de las estimaciones de las FAC y FAP son en base a las ecuaciones de las Secs. 2.4, 3.3.1, 3.3.5 y 3.6.2 para los procesos planteados, metodología que se empleará muchas veces, ya que es uno de los pasos más importantes del análisis de series de tiempo.

Se puede corroborar este paso con la tabla N^o 06 del resumen del comportamiento de la FAC y FAP para un modelo ARIMA(p,d,q).

Así mismo se puede resumir que las variancias aproximadas de la variable transformada (W_t); donde $W_t = \Delta^d Z_t$, $n = N-d$; para algunos modelos se muestra en la tabla N^o 07.

- g) Estudiar la presencia o no de estacionariedad; efectuar las diferencias necesarias y/o transformaciones para obtener una nueva serie estacionaria.

Las transformaciones son utilizadas para todo proceso lineal, en cambio se utilizarán las diferencias ordinarias o estacionales para aquellos modelos no estacionarios.

Tabla N^o 06:

Comportamiento de la FAC y FAP de un proceso ARIMA (p,d,q)

ORDEN	(1,d,0)	(0,d,1)
comportamiento de ρ_k	decae exponencialmente	solamente $\rho_1 \neq 0$
comportamiento de ϕ_{kk}	solamente $\phi_{11} \neq 0$	decaimiento exponencial dominante
estimados iniciales	$\phi_1 = \rho_1$	$\rho_1 = -\theta/(1+\theta^2)$
región de admisibilidad	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$
ORDEN	(2,d,0)	(0,d,2)
comportamiento de ρ_k	mezcla de exponenciales y ondas sinusoidales amortiguadas	solamente $\rho_1 \neq 0$ y $\rho_2 \neq 0$
Comportamiento de ϕ_{kk}	solamente $\phi_{11} \neq 0$ y $\phi_{22} \neq 0$	Dominada por mezcla exponenciales o senos amortiguados
estimados iniciales	$\phi_1 = \rho_1(1-\rho_2)/(1-\rho_2^2)$ $\phi_2 = (\rho_2 - \rho_1^2)/(1-\rho_2^2)$	$\rho_1 = -\theta_1(1-\theta_2)/(1+\theta_1^2+\theta_2^2)$ $\rho_2 = -\theta_2/(1+\theta_1^2+\theta_2^2)$
región de admisibilidad	$-1 < \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $\phi_2 + \phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2 - \theta_1 < 1$ $\theta_2 + \theta_1 < 1$
ORDEN	(1,d,1)	
comportamiento de ρ_k	decae exponencialmente después de un atraso (de lag 1).	
comportamiento de ϕ_{kk}	dominada por el decaimiento exponencial después del 1 ^{er} atraso.	
estimados iniciales	$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$, $\rho_2 = \rho_1\phi_1$	
región de admisibilidad	$-1 < \phi_1 < 1$; $-1 < \theta_1 < 1$	

Tabla N207 Variancias aproximadas para $W_t = \hat{a}Z_t$, $n=N-d$

AR(1)	MA(1)	ARMA (1,1)
$\frac{C_0(1+r_1)}{n(1-r_1)}$	$\frac{C_0(1+2r_1)}{n}$	$\frac{C_0}{n} \left[1 + \frac{2r_1^2}{r_1 - r_2} \right]$
AR(2)	MA(2)	
$\frac{C_0(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)}$	$\frac{C_0(1+2r_1+2r_2)}{n}$	

C_0 es el estimado de γ_0 y r_1 el estimado de ρ_1 .

- h) Obtener las estimaciones preliminares para los parámetros de los modelos AR(p), MA(q), ARMA(p,q), ARIMA(p,d,q), SARMA(p,q)x(P,Q), SARIMA(p,d,q)x(P,D,Q) y metódicamente eliminar los órdenes del modelo que no resultan por causa de estacionariedad o de invertibilidad.

En la sec. 3.5 se describió los diferentes procedimientos para la estimación de los parámetros y los errores estandar de dichos parámetros estimados. Existen también otros procedimientos como el no condicional que puede revisarse en Box y Jenkins (1970) por el método de máxima verosimilitud.

Si tenemos un modelo ARIMA (p,d,q) y los parámetros ϕ y θ a estimar, tanto el procedimiento condicional y no condicional, formarán una suma de cuadrados que

serán funciones de ϕ y θ ; estas sumas deberán ser calculadas para diversos valores de (ϕ, θ) sobre un reticulado conveniente; esto es considerando que $-1 < \phi < 1$ y $-1 < \theta < 1$ y calculando $S(\phi, \theta)$ sobre un reticulado como el de Fig. N^o 13 ; se unen los puntos que tienen suma de cuadrados aproximadamente iguales.

Mediante la observación de las curvas de nivel, se encontrará los pares $(\bar{\phi}, \bar{\theta})$ que determinan el menor valor de $S_x(\phi, \theta)$ lo cual equivale a maximizar $L(w)$.

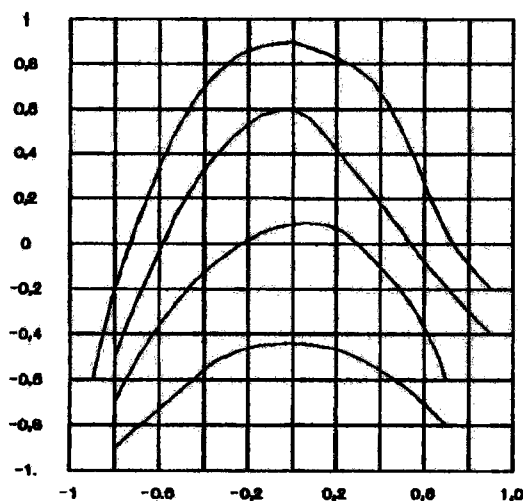


Fig. N^o 13 Curvas con la misma Suma Cuadrados (SQ) para un modelo ARMA(1,1).

Para tener una idea de la precisión de los estimadores encontrados debemos construir intervalos de confianza para los parámetros y se analiza según la sec. 3.4.5, que a partir de las estimados de las

variancias podemos obtener los intervalos de confianza y para los modelos más comunes se presenta en la tabla N^o 08 que muestra estas variancias aproximadas de los estimadores.

Tabla N^o 08 Variancias aproximadas para los parámetros de los modelos más comunes.

MODELO	VARIANZA
AR(1)	$\text{Var}(\bar{\phi}) \approx (1-\phi^2)/n$
AR(2)	$\text{Var}(\bar{\phi}_1) = \text{Var}(\bar{\phi}_2) \approx (1-\phi^2_2)/n$
MA(1)	$\text{Var}(\bar{\theta}) \approx (1-\theta^2)/n$
MA(2)	$\text{Var}(\bar{\theta}_1) = \text{Var}(\bar{\theta}_2) \approx (1-\theta^2_2)/n$
ARMA(1,1)	$\text{Var}(\bar{\phi}) \approx \frac{(1-\phi)^2}{n} + \frac{(1-\phi\theta)^2}{(\phi-\theta)^2}$
	$\text{Var}(\bar{\theta}) \approx \frac{(1-\theta)^2}{n} + \frac{(1-\phi\theta)^2}{(\phi-\theta)^2}$

- i) Eliminar los coeficientes preliminares, aquellos que poseen mayor error estandard; y los que son no significativos.
- j) Verificar si los residuos es un ruido blanco, caso contrario obtener otros modelos o alterarlo, adicionando o retirando parámetros con valor próximo a dos veces la desviación estandar del estimador.

Una técnica a ser usada, si se tiene un modelo con muchos parámetros es ajustado para ellos, estimaremos un modelo con parámetros extras y examinaremos si estos son significativos y si su inclusión disminuye significativamente la variancia residual. Este método es útil cuando sabemos a priori en que dirección puede estar ocurriendo la inecuabilidad del modelo.

Esta fase se basa en lo descrito en la sec. 3.7 incluyendo el análisis del periodograma para la variable estocástica a lo que serán usados en el software Statgraphics y Systac.

- k) Como es de nuestro interés saber cuales y cuantos son las predicciones en un tiempo futuro, Z_{t+h} , $h > 1$, y teniendo las observaciones Z_{t-2} , Z_{t-1} , Z_t en el instante t , origen de las predicciones; los diferentes procedimientos para obtener estas predicciones se presentaron en la sec. 4.5 con sus respectivos intervalos de confianza.

Los horizontes de análisis serán elejidos de acuerdo a las fuertes variaciones y a los años normales para así obtener un buen pronóstico.

VI RESULTADOS Y DISCUSION6.1 RESULTADOS

La serie en estudio se presenta en el tabla N° 09 cuyos datos originales de descargas en m³/s, para el análisis de series de tiempo, con 156 observaciones.

TABLA N° 09 SERIE EN ESTUDIO CAUDAL (m³/s)

	SET	OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO
74-75	2,003	4,053	7,505	8,386	26,670	28,670	74,190	17,530	9,800	3,590	2,380	1,810
75-76	1,654	1,583	6,606	10,080	60,281	86,931	62,199	20,023	6,767	3,463	2,520	1,751
76-77	2,428	2,558	6,324	5,725	8,424	80,178	66,461	15,790	5,706	2,653	1,977	1,877
77-78	0,990	0,797	16,727	6,903	20,522	27,775	9,894	11,296	2,574	1,440	1,119	0,716
78-79	0,363	1,177	8,325	12,585	10,455	50,679	71,229	11,256	2,548	1,590	1,158	0,758
79-80	0,800	0,609	4,240	2,087	15,648	18,148	37,196	12,500	2,225	1,180	0,550	0,712
80-81	0,610	5,021	2,338	9,765	19,316	100,078	61,914	29,747	4,231	1,337	1,028	0,809
81-82	1,851	2,943	6,594	19,520	39,612	78,789	46,932	13,304	3,555	1,918	1,464	1,187
82-83	1,440	3,718	9,556	5,982	19,125	6,789	47,032	77,726	3,746	1,124	0,685	0,773
83-84	0,625	0,588	0,564	10,220	48,120	295,328	118,983	40,910	9,495	4,149	2,059	1,495
84-85	1,182	10,592	12,355	39,850	15,099	77,199	76,718	52,518	10,777	5,972	2,869	1,873
85-86	1,918	3,412	5,545	22,735	115,845	222,378	81,003	50,074	17,492	7,111	3,414	3,164
86-87	2,861	4,282	6,910	30,926	79,438	55,644	23,403	7,936	1,614	1,337	1,293	0,720

Se realizó el ploteo de la serie original que se muestra en la figura N° 14, de acuerdo a la información de la tabla 09.

Observándose la Fig. N° 14, la serie, presenta un comportamiento periódico, lo que indica que la serie es de carácter estacional; serie con una media mensual de 20,77 m³/s y una desviación estandar de 37,70 m³/s, teniéndose

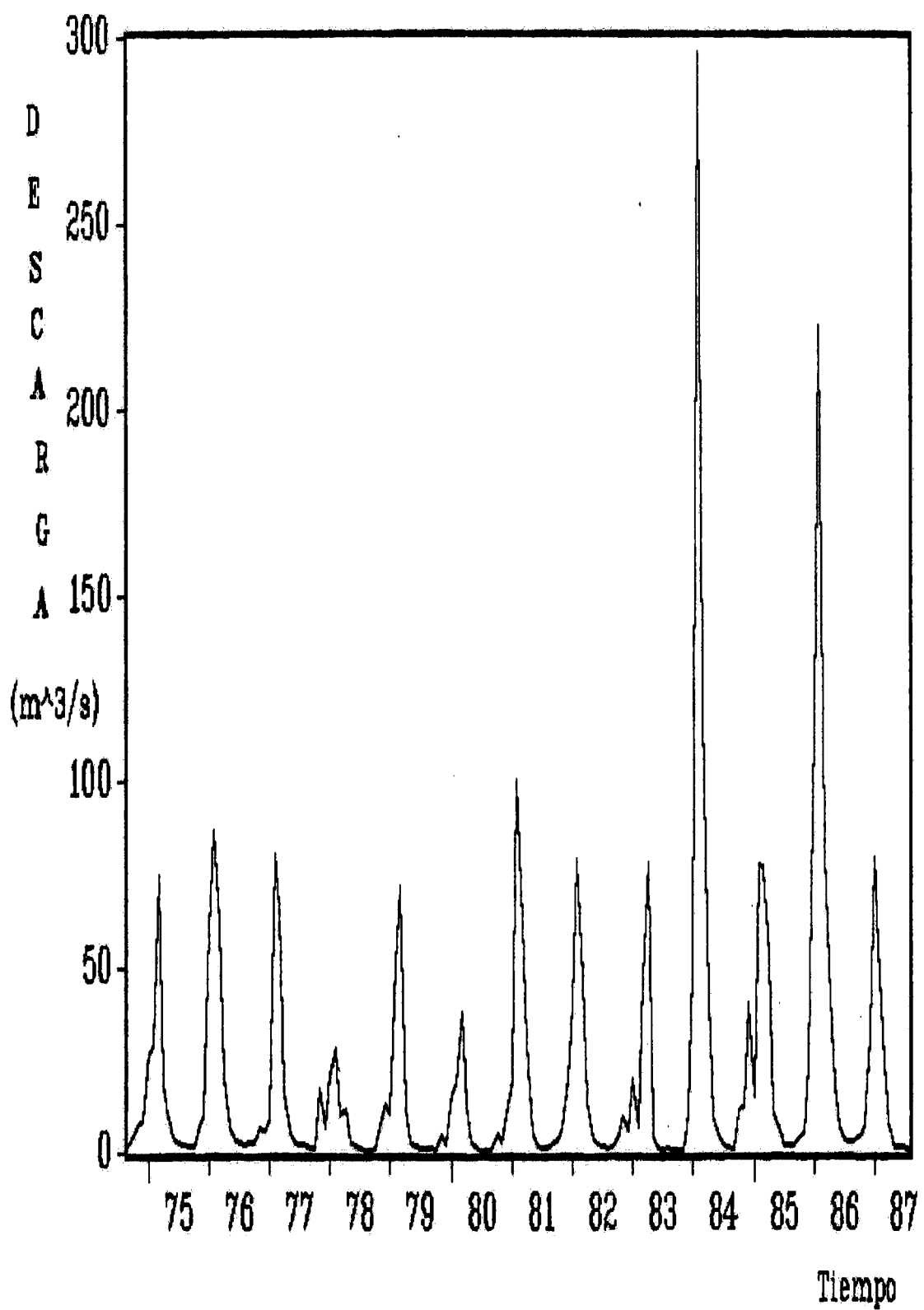


Fig. N 14 Descargas Medias Mensuales
Rio Pisco (m³/s)

caudales picos generalmente en los meses de febrero y un atraso o adelanto, como la presencia de pico en abril de 1983 (77,73 m³/s). También presenta algunos picos muy resaltantes como en febrero de 1984 y 1986 (de 295,33 y 222,38 m³/s respectivamente).

La presencia de estos picos no han sido muy bien definidos porque la información con que se cuenta es sólo a partir de 1974 y probables efectos de un atraso a la presencia de el fenómeno "antiniño". Para efecto del análisis consideramos estas como observaciones irregulares y que el modelo tratará de absorber.

Análisis Exploratorio de los datos

Para el análisis del comportamiento de la serie; primero, se realizó el suavizamiento robusto 4253HT. y el ploteo de estos resultados se presenta en la Fig. 15; mostrando un comportamiento similar al de los datos originales, con picos, presencia de estacionalidad y cierto crecimiento estacional; lo que indica no estacionariedad periódica.

Continuando con el análisis exploratorio de los datos (AED) se realizó el análisis de cajas mostrado en la figura N° 16 agrupando las observaciones anuales, observandose las dispersiones algo mayores para los años hidrológicos del 83 - 84 - 85, ligeramente mayores con

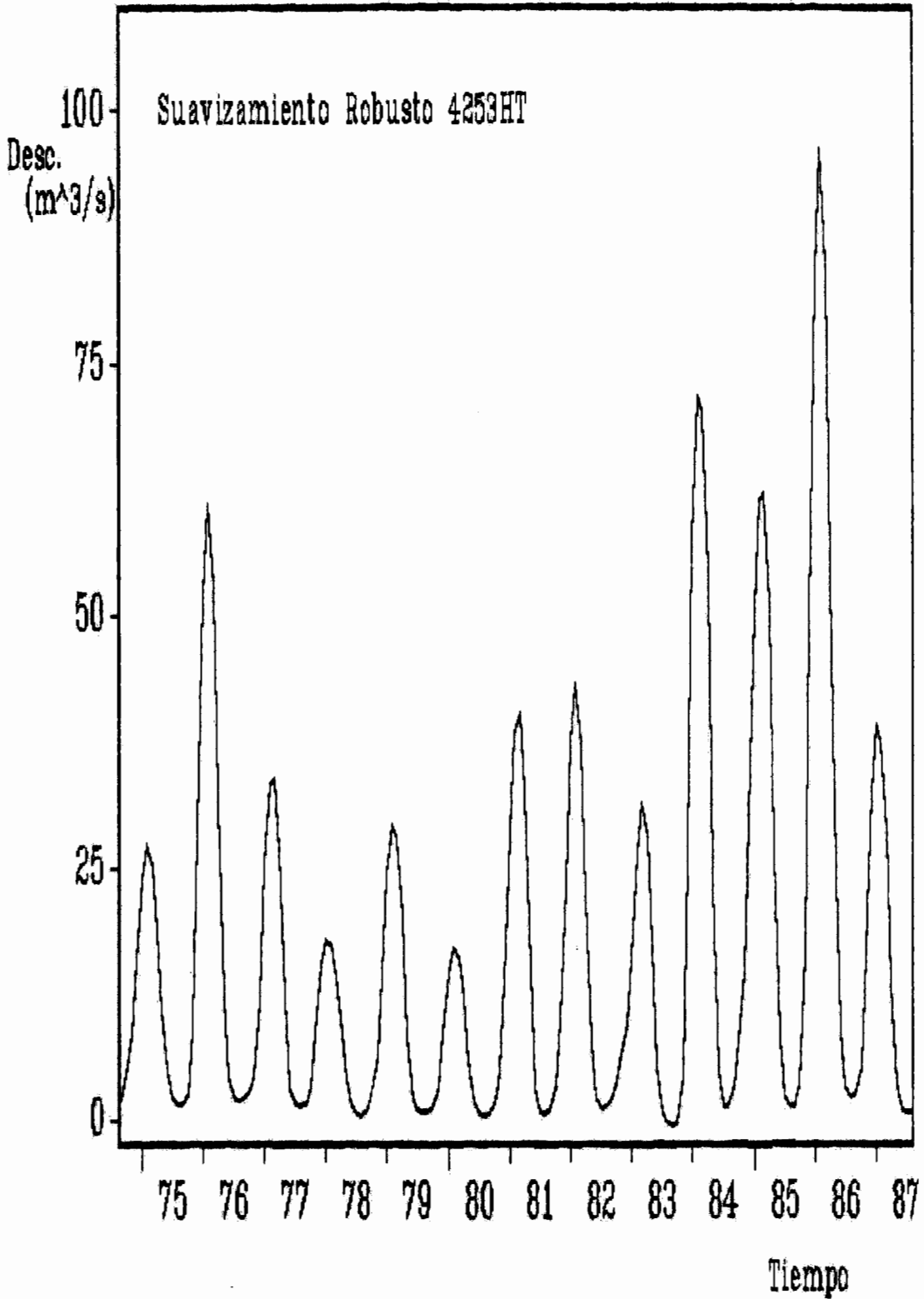


Fig. N 15 Descargas Medias Mensuales Suavizadas

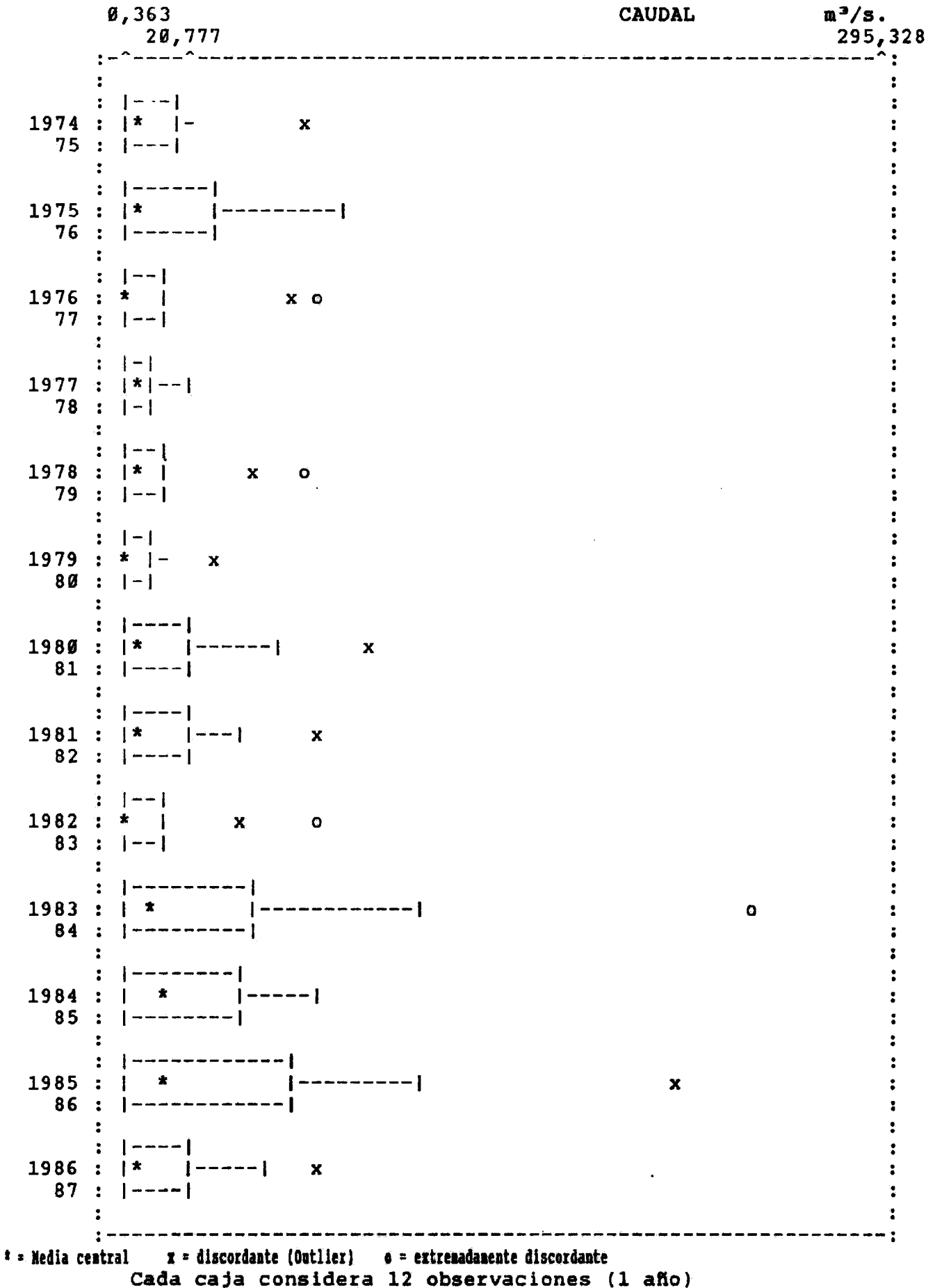


FIG. N°16 PLOTEO DE CAJAS (BOX PLOT)

respecto a los años anteriores y disminuyendo en 1986, presentando algunas observaciones discordantes ("outlier") y extremadamente discordantes "extreme outlier" así como cierto incremento de la media (*). Indicando de esta manera una serie asimétrica y presencia de carácter no estacionario, siendo necesario alguna transformación no lineal de los datos de tal manera que se estabilice la dispersión de los datos.

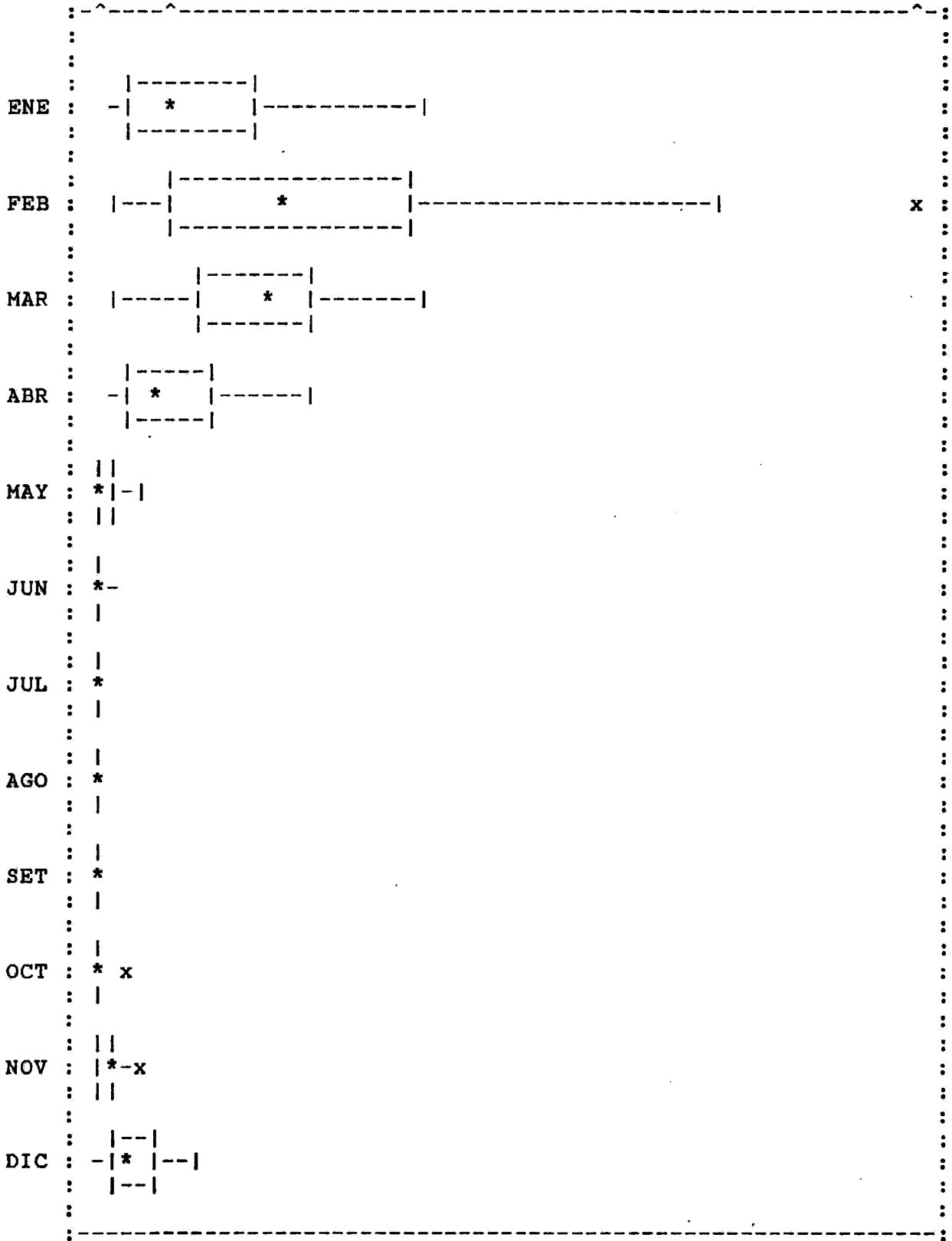
Para analizar la presencia de estacionalidad de la muestra, se realizó el ploteo de caja agregada cuyo resultado se presenta en la figura N^o 17 correspondiente a los meses de enero a diciembre, confirmandose que la presencia de descargas máximas son generalmente en el mes de febrero (2) y las épocas de poca descarga corresponden a los meses de julio, agosto y setiembre, así mismo los meses de marzo presenta una baja dispersión en sus datos salvo una discordante, lo que indica una descarga casi uniforme en el tiempo.

La presencia de dispersiones y asimetría de los datos como se observan en el análisis de cajas y en el ploteo de la serie hara necesario encontrar el valor de b de la ec. 4.10; apropiado para lograr la simetría, lograr la dispersión uniforme e independiente del nivel de la variable, etc. Para encontrar el valor de b adecuado se efectuó el ploteo del Nivel vs. la Dispersión de los

VARIABLE: CAUDAL

0,363
20,777

m³/s.
295,328



* = Media central
x = Dato discordante
Cada caja considera 13 observaciones (de 13 años)

FIG. Nº 17 GRAFICO DE CAJAS AGREGADO (AGGREGATE BOX PLOT)

caudales; como se muestra en la figura N^o 18; estimando una pendiente de $m=1,0506$ por mínimo cuadrados lo que indica que la familia de transformaciones, la existencia de una dependencia entre el nivel y la dispersión de tipo logarítmica, encontrándose de $1-m = 1-1,0506 = -0,0506 \approx 0$, por lo tanto los caudales serán transformados por medio de una transformación logarítmica ($b=0$); definido por el Forecast Plus; por $Z_{t'} = \ln(Z_t + 0,1)$; $\forall Z_t \geq 0$.

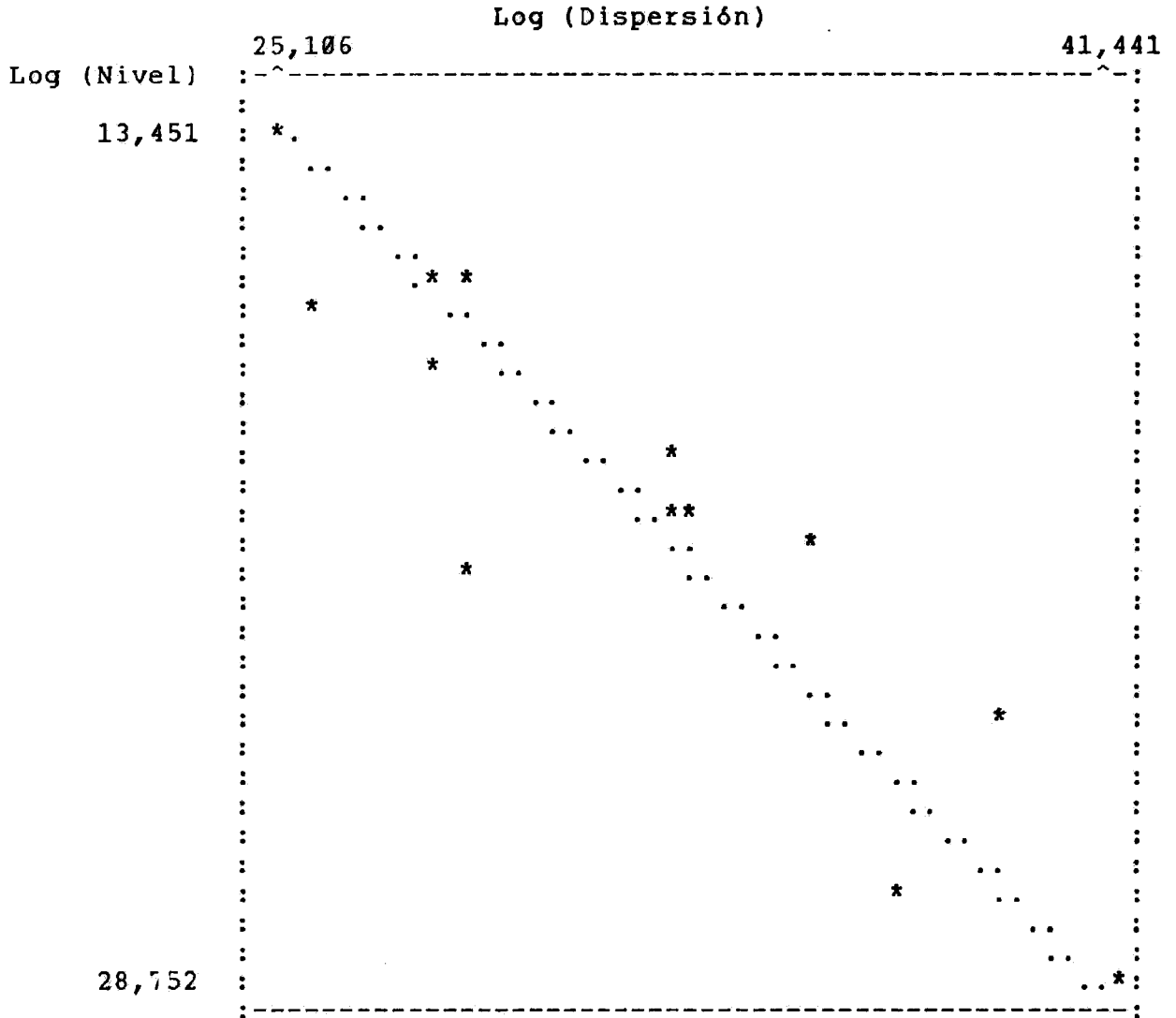
Una vez identificado el tipo de transformación se procedió a aplicar logaritmo a cada una de las observaciones. luego se procedió nuevamente al ploteo del Nivel VS. la Dispersión de los datos transformados obteniéndose la figura N^o 19; así encontrándose una pendiente de $m=-0,0368$, por lo tanto un valor de b igual a $1-m=1,0368 \approx 1$. indicando que ya no existe dependencia entre el nivel y la dispersión. De esta manera se obtiene la serie con cierto grado de homogeneidad.

La serie transformada ($\ln(\text{Caudal})=Z_{t'}$) es nuestra nueva serie a ser utilizada, estos resultados se presenta en la figura N^o 20 de acuerdo a la transformación utilizada la serie presenta simetría y una variación alrededor de su media, etc.

Nuevamente podemos efectuar el ploteo de caja y caja agregada para la serie transformada ($b=0$) ya que esta transformación debe eliminar toda las dispersiones,

VARIABLE: CAUDAL

valores de los datos a Escala 1×10



* Considera subconjuntos de 12 Observaciones consecutivas
Estimado mínimo cuadrado de la pendiente = 1,0506
Transformación aproximada para estabilizar la dispersión
del tipo LOGARITMICO.

FIGURA Nº 18 PLOTEO DEL NIVEL VS. LA DISPERSION

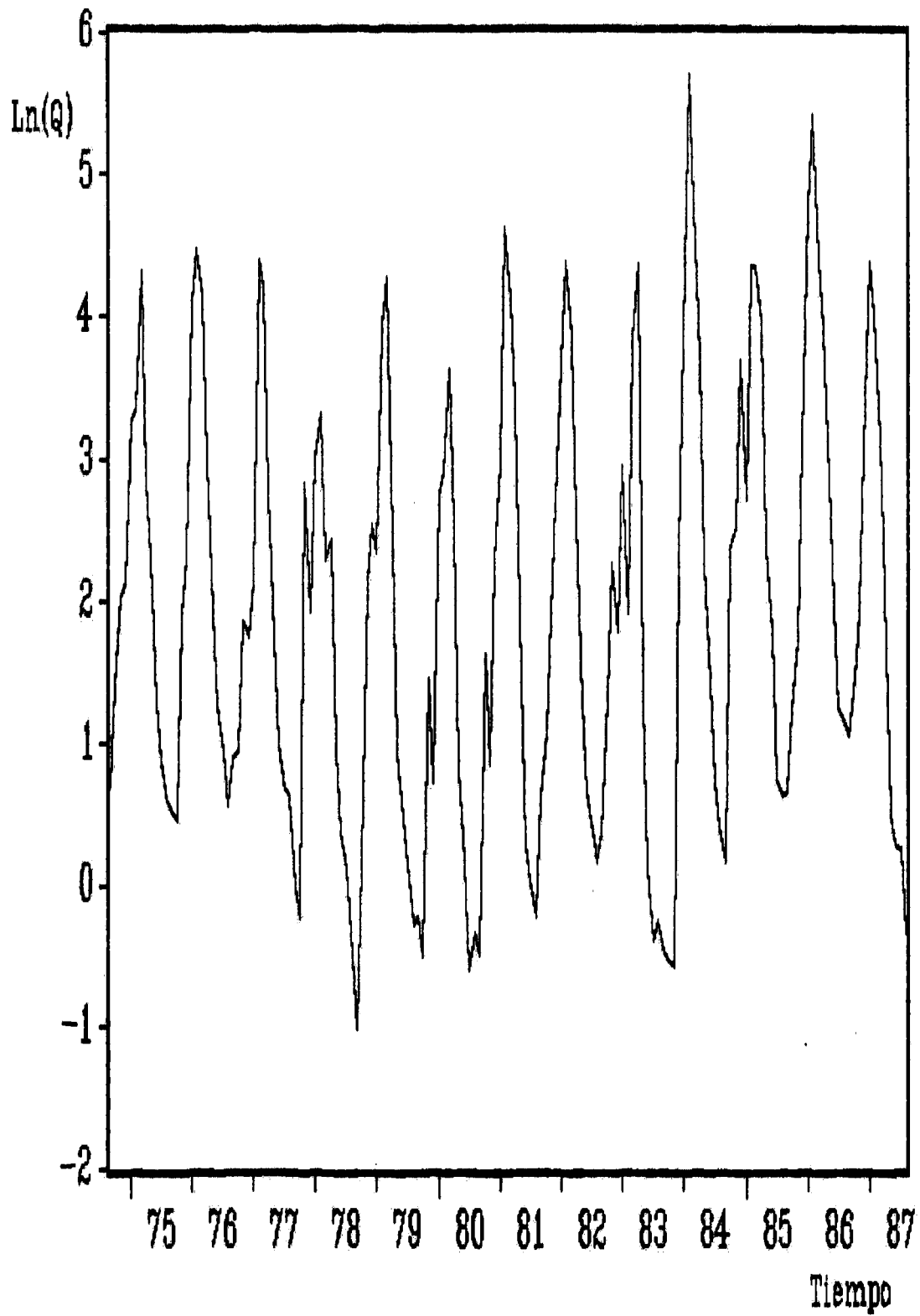


Fig. N 20 Ploteo de serie Transformada - $\text{Ln}(\text{Descarga})$

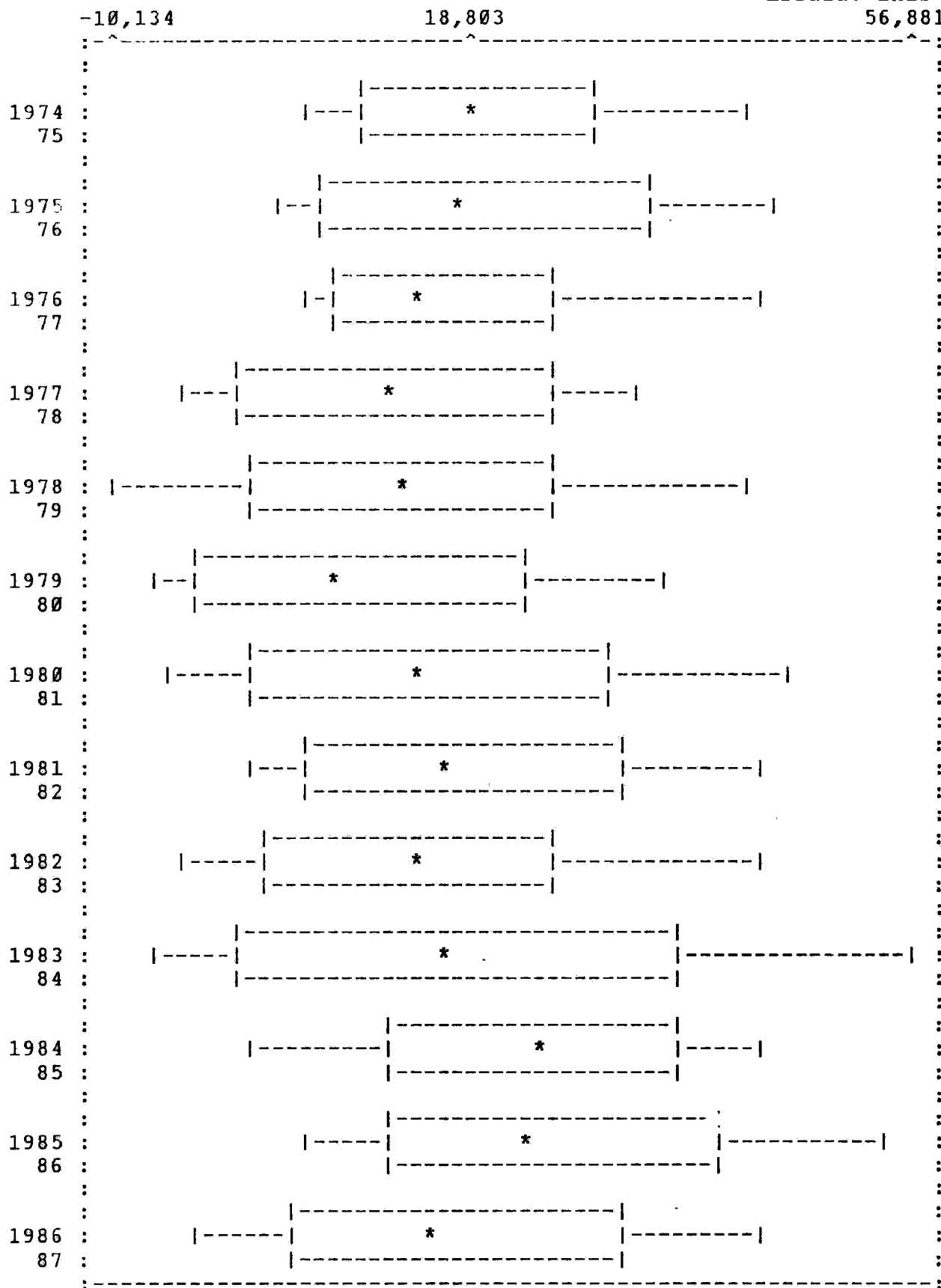
asimetría, etc.; así mostrándose en las figuras Nº 21 y 22 respectivamente. En la Fig. 21 se puede observar la presencia de cierta tendencia lo que indica no estacionariedad; en la Fig. 22 se observa presencia de estacionalidad, con período $s=12$, bien definido, en ambas figuras no se presentan datos de extremo discordantes.

Para verificar la variabilidad de nuestra serie se procedió a efectuar el análisis de consistencia de nuestros datos transformados y obteniéndose que estadísticamente no son significativos la variancia como se muestra en la tabla Nº 10 estos resultados de variancia para 13 años.

La prueba de comparación de la variancia, entre 1974 y de 1983. obteniéndose un $F_c=3,347$; con valores críticos para un nivel de confianza del 99% el F tabla con (11,11) G.L. es 4,46; por lo tanto no existe evidencia estadística para rechazar la hipótesis entre la homogeneidad de variancia. Pero más no presenta homogeneidad los datos originales.

Como podemos haber visto el análisis exploratorio de los datos es una herramienta para analizar el comportamiento de nuestros datos y que deben cumplir ciertas condiciones para poder ser modelado por un modelo de Box y Jenkins ya que sino fuera este análisis estaríamos asumiendo los

Escala: 1x10
56,881



* = Media (Midmean)
Cada caja considera 12 observaciones (de 1 año)

FIG. Nº 21 PLOTEO DE CAJAS (BOX PLOT)
VARIABLE: Ln(CAUDAL)

TABLA Nº 10: VARIANCIAS ANUALES DE Ln (CAUDAL)

ANO	VARIANCIA		ANO	VARIANCIA
74-75	1,284 *		81-82	2,017
75-76	2,030		82-83	2,190
76-77	1,516		83-84	4,297 *
77-78	1,754		84-85	1,917
78-79	2,586		85-86	2,377
79-80	2,066		86-87	2,310
80-81	2,763		87-88	2,402

datos como si tuvieran una distribución normal, sin dispersión, etc. sin embargo no siempre los datos hidrometeorológicos presentan características homogéneas.

Fase de Identificación y planteamiento del Modelo

Una vez que tenemos los datos con ciertas condiciones podemos efectuar su identificación dentro de los modelos de Box y Jenkins y plantear los modelos probables que pueden representar a nuestras descargas medias mensuales y poder predecir valores futuros.

De acuerdo a las ecuaciones 3.10, 3.11 y según a las sec. 2.4, 3.3.1 y 3.3.3 para la estimación de las funciones de autocorrelación y autocorrelación parcial se calculó para los datos originales y para los transformados; presentando

en las figuras Nº 23 y 24 respectivamente a los datos originales porque es necesario comparar que con la transformación mantenga la estructura de la información; mostrando las bandas del límite de confianza.

Como puede observarse en las Figs. Nº 23 y 24 el comportamiento de la FAC y FAP son muy similares y muestran cierta estacionariedad ordinaria; sin embargo hay presencia de estacionalidad y no estacionariedad por efecto estacional. Para poder completar con el análisis se procedió a efectuar el análisis del periodograma y periodograma acumulado de acuerdo a la sec. 3.7.3 para una variable Z_t ($\ln(\text{caudal})$), mostrándose en las figuras Nº 25 (a y b) obteniéndose que el efecto estacional es muy importante en nuestra serie para período $s=12$, e igual a una frecuencia de $1/12 \approx 0,083$ y con menos importancia para otros períodos, en la fig. 25.b en $0,083$ se produce un salto lo que indica estacionalidad. Esta presencia de estacionalidad nos indica que para caudales de enero, febrero, etc. del año j tienen cierta o estrecha relación con los caudales de enero, febrero, etc. del $j-1$, $j-2$; respectivamente.

Las ondas sinusoidales que presenta la FAC presenta estacionariedad regular, pero para el período estacional no presenta ondas ni decaimiento exponencial, sino un

Número de observaciones : 156 (1974.09-1987.08)
 Media de la serie = 20,77661
 Desviación estandar de la serie = 37,70373

Autocorrelaciones	Autocorrelación Parcial	FAC	FAP
. *****	. *****	1	0.531 0.531
. **	*** .	2	0.127 -0.214
. * .	. * .	3	-0.090 -0.097
*** .	. * .	4	-0.195 -0.092
*** .	** .	5	-0.239 -0.121
*** .	** .	6	-0.249 -0.117
*** .	. * .	7	-0.220 -0.100
** .	. * .	8	-0.161 -0.066
. * .	. .	9	-0.047 -0.006
. ***	. ***	10	0.220 0.264
. *****	. *	11	0.370 0.111
. *****	. **	12	0.400 0.164
. ***	. .	13	0.267 0.016
. *	. .	14	0.075 -0.003
. * .	. * .	15	-0.107 -0.050
** .	. .	16	-0.105 0.014
*** .	. * .	17	-0.227 -0.045
*** .	. .	18	-0.223 -0.017
** .	. .	19	-0.109 -0.008
** .	. * .	20	-0.137 -0.095
. * .	. * .	21	-0.059 -0.005
. **	. .	22	0.115 0.020
. *****	. ****	23	0.424 0.329
. *****	. ***	24	0.563 0.232
. ****	** .	25	0.290 -0.140
. *	. *	26	0.076 0.102
. * .	. .	27	-0.090 -0.012
** .	. *	28	-0.162 0.041
** .	. .	29	-0.192 0.015
** .	. .	30	-0.109 0.030
** .	. .	31	-0.173 0.007
** .	. .	32	-0.135 0.007
. * .	. * .	33	-0.039 -0.007
. **	. * .	34	0.154 -0.001
. ****	. .	35	0.292 0.021
. ****	. .	36	0.325 0.014

Box-Pierce Q-Stat 337,02 Prob 0,0000 SE of Correlations 0.004
 Ljung-Box Q-Stat 305,70 Prob 0.0000

FIGURA Nº 23 FUNCION DE AUTOCORRELACION (FAC) Y FUNCION DE AUTOCORRELACION PARCIAL (Datos originales - Caudal)

Número de observaciones : 156
 Media de la serie = 1,88031
 Desviación estandar de la serie = 1,556896

Autocorrelaciones	Autocorrelación Parcial	PAC	PAP
. *****	. *****	1 0.772 0.772	
. *****	***** .	2 0.441 -0.305	
. .	***** .	3 0.024 -0.438	
**** .	*** .	4 -0.335 -0.195	
***** .	*** .	5 -0.598 -0.227	
***** .	** .	6 -0.691 -0.124	
***** .	. *	7 -0.589 0.045	
**** .	. ***	8 -0.293 0.236	
. *	. **	9 0.079 0.165	
. *****	. *	10 0.422 0.112	
. *****	. ***	11 0.677 0.243	
. *****	. *	12 0.747 0.094	
. *****	. .	13 0.624 0.008	
. ****	. .	14 0.340 0.016	
. .	. .	15 -0.015 0.012	
**** .	. *	16 -0.360 -0.077	
***** .	** .	17 -0.606 -0.116	
***** .	. *	18 -0.662 0.061	
***** .	. *	19 -0.544 -0.047	
**** .	** .	20 -0.296 -0.150	
. .	. *	21 0.021 -0.057	
. ****	. *	22 0.344 0.054	
. *****	. **	23 0.603 0.142	
. *****	. *	24 0.699 0.07.	
. *****	. *	25 0.569 -0.088	
. ****	. .	26 0.292 -0.034	
. .	. .	27 -0.038 0.034	
**** .	. .	28 -0.353 -0.011	
***** .	. .	29 -0.570 0.009	
***** .	. .	30 -0.643 -0.011	
***** .	. *	31 -0.555 -0.102	
**** .	. *	32 -0.312 -0.040	
. .	. .	33 0.007 -0.002	
. ****	. .	34 0.322 0.007	
. *****	. .	35 0.546 -0.011	
. *****	. .	36 0.614 -0.018	

Box-Pierce Q-Stat 1345,16 Prob 0,0000 SE of Correlations 0,000
 Ljung-Box Q-Stat 1544,47 Prob 0,0000

FIGURA Nº 24 FUNCION DE AUTOCORRELACION Y FUNCION DE AUTOCORRELACION PARCIAL (Variable=Ln(caudal))

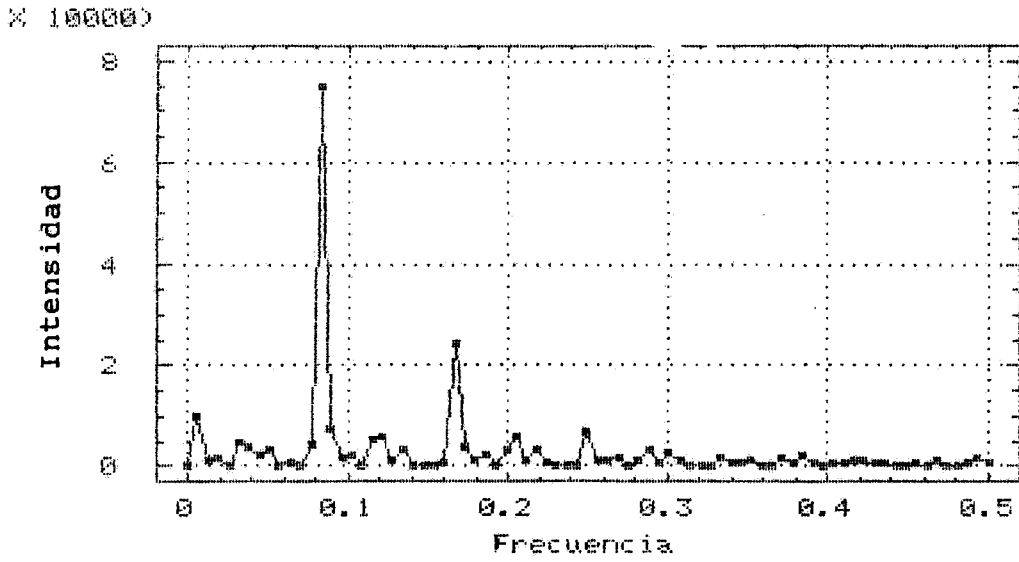


Fig. Nº 25.a Periodograma - Rio Pisco

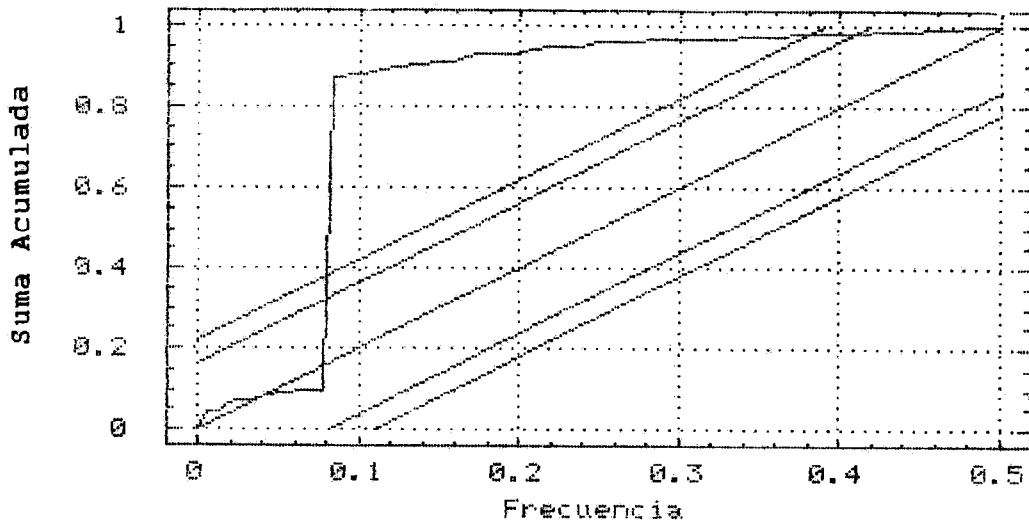


Fig. Nº 25.b Periodograma Acumulado - Rio Pisco
Ln(Caudal)

crecimiento tal como se ha visto en los análisis anteriores esto requiere llevar a ser estacionaria para lo cual se tomarán diferencias de orden 12 ($D=1$) para lograr estacionariedad, lo mismo puede efectuarse para la serie original que presenta este mismo comportamiento a pesar de no gozar de homogeneidad de varianza.

Tomando las diferencias de orden 12 se procedió a plotear las FAC y FAP tal como se presenta en las figuras N^o 26; presentando una serie estacionaria donde se puede identificar el orden para los términos: AR, MA, SAR y/o SMA.

Inicialmente se puede plantear un modelo de orden 1 para el término autorregresivo regular ($AR(1)$) ya que en la FAP es significativa para un atraso; y un orden 1 para el término de media móvil regular ($MA(1)$) por lo que presenta una correlación significativa; para la parte estacional se puede diferenciar la FAP un orden 2 para el término autorregresivo estacional ($SAR(2)$) presentando la FAP correlaciones significativas de atrasos 12 y 24; así mismo en la función de autocorrelación presenta para la parte estacional una correlación significativa para atraso 12 por lo que le corresponde orden de 1 para la media móvil estacional ($MA(1)$).

Esto indica que nuestro primer modelo identificado según las funciones de autocorrelación y autocorrelación

Número de observaciones : 144
 Media de la serie = -0,02085
 Desviación estandar de la serie = 1,010666
 Diferenciación estacional de primer orden (D=1)

Autocorrelaciones	Autocorrelación Parcial	PAC	PAP
. *****	. *****	1	0.417 0.417
. ***	. .	2	0.190 0.030
. .	** .	3	-0.018 -0.133
. *	. **	4	0.000 0.155
. **	. **	5	0.151 0.117
. *	** .	6	0.039 -0.133
. *	. *	7	-0.061 -0.067
. *	. ***	8	0.057 0.203
. **	. **	9	0.187 0.128
. *	*** .	10	0.042 -0.237
** .	** .	11	-0.130 -0.130
***** .	***** .	12	-0.510 -0.427
*** .	. ***	13	-0.193 0.227
. *	. *	14	-0.039 0.113
. *	. .	15	0.065 -0.033
. .	. *	16	0.002 0.070
. *	. .	17	-0.074 -0.015
. *	. .	18	0.042 0.037
. **	. *	19	0.138 0.077
. .	. .	20	0.018 -0.003
. *	. .	21	-0.100 0.020
. *	** .	22	-0.106 -0.122
. .	. .	23	0.009 -0.033
. **	*** .	24	0.144 -0.208
. *	. **	25	0.066 0.136
. .	. *	26	-0.020 0.084
. *	. *	27	-0.040 -0.056
. *	. *	28	-0.059 -0.110
. .	. *	29	-0.023 -0.048
. *	. .	30	-0.113 0.011
*** .	. *	31	-0.226 -0.091
** .	. *	32	-0.100 -0.086
. *	. *	33	-0.062 0.004
. .	. *	34	0.010 -0.054
. .	. .	35	-0.014 0.014
. *	. *	36	-0.004 -0.000

Box-Pierce Q-Stat 114,57 Prob 0,0000 SE of Correlations 0,003
 Ljung-Box Q-Stat 127,80 Prob 0,0000

FIGURA Nº 26 FUNCION DE AUTOCORRELACION Y AUTO-CORRELACION PARCIAL (Variable=Ln(caudal; D=1)

parcial sería SARIMA(1,0,1)x(2,1,1)₁₂ donde la variable sería la transformada de Box y Cox; $Z_t = \ln(\text{caudal})$; para el software del Forecast Plus trabaja con $\ln(\text{caudal} + 0,1)$, 0,1 para poder considerar algunos datos de valores 0 que podría presentarse. y que los valores de caudales siempre tendrán que ser mayores o iguales a cero. El modelo SARIMA(1,0,1)x(2,1,1)₁₂ para Z_t es equivalente a un modelo SARMA(1,1)x(2,1)₁₂ para W_t ; con $W_t = Z_t - Z_{t-12}$.

Fase de Estimación Verificación y selección de modelos

El modelo identificado, SARIMA (1,0,1)x(2,1,1)₁₂ se procedió a estimar los parámetros ϕ_1 , ϕ_2 , θ_1 y δ_1 siguiendo las ecuaciones de YW y las de la sec. 3.5, 4.2 y 4.4; estos resultados se presenta en la tabla N° 11; así también se muestra la matriz de correlación entre los parámetros estimados lo que nos indica que θ_1 tiene fuerte correlación con ϕ_1 , más no con los demás parámetros; δ_1 corelacionado con ϕ_1 y ϕ_2 , esto nos indica que el modelo puede ser explicado con un menor número de parámetros, y nos permite plantear nuevo modelo.

La validación del modelo planteado nos permite encontrar la función de residuos según Ecs. de la sec. 3.7 para $d=0$ y $D=1$, obteniendose así la Fig. N° 27 donde se muestra la FAR encontrandose que los residuos se puede considerar como ruido ruido blanco por lo que todas las

Número de observaciones : 143
 Media de la serie residual = 0,07023
 Desviación estandar de la serie residual = 0,716802
 Diferenciación estacional de primer orden (D=1; s=12)

Lag	Value	T-Value	-1.0	0.0	+1.0
1	0.00	0.02	:	[]	:
2	0.02	0.28	:	[*]	:
3	-0.10	-1.21	:	[***]]	:
4	0.05	0.65	:	[*]	:
5	0.04	0.49	:	[*]	:
6	0.07	0.79	:	[**]	:
7	-0.06	-0.68	:	[*]	:
8	-0.00	-0.05	:	[]	:
9	0.14	1.61	:	[***]	:
10	-0.03	-0.38	:	[*]	:
11	0.07	0.81	:	[**]	:
12	-0.01	-0.17	:	[]	:
13	-0.06	-0.63	:	[*]	:
14	0.01	0.07	:	[]	:
15	0.09	1.07	:	[**]	:
16	-0.02	-0.18	:	[]	:
17	-0.12	-1.30	:	[***]]	:
18	0.07	0.83	:	[**]	:
19	0.00	0.05	:	[**]	:
20	-0.06	-0.62	:	[*]	:
21	-0.01	-0.06	:	[]	:
22	-0.19	-2.13	:	[****]]	:
23	0.06	0.59	:	[*]	:
24	-0.02	-0.25	:	[*]	:
25	0.01	0.10	:	[]	:
26	-0.02	-0.18	:	[]	:
27	0.09	0.95	:	[**]	:
28	-0.06	-0.60	:	[*]	:
29	0.01	0.16	:	[]	:
30	-0.00	-0.05	:	[]	:

[] = Límites 2-error estándar.

Estadístico Box-Pierce Chi-Cuadrado con 25 G.L. = 20,2
 Probabilidad = 0,739 (73,9%)

FIGURA Nº 27 FUNCION DE AUTOCORRELACION RESIDUAL
 MODELO: SARIMA (1,0,1)x(2,1,1)₁₂

autocorrelaciones de los residuos se encuentra dentro del intervalo de confianza lo que se dice que el ruido esta incorrelacionado. La prueba de bondad de Box y Pierce según indica un $QB=20,2$ con 25 G.L., afirmándose que se acepta los residuos como ruido blanco con un nivel de confianza de 73,9% para Chi-cuadrado con 25 G.L., indicando un valor muy aceptable.

Como podemos ver que nuestro modelo puede ser modificado y considerar un parámetro autorregresivo estacional de orden 1 porque este SAR(1) tiende a representar al SAR(2) como se vio en el modelo inicial; es así que se plantea en modelo SARIMA $(1,0,1) \times (1,1,1)_{12}$ para lo cual se efectuó la estimación de los parámetros correspondientes como se indica en la tabla N° 12; donde la correlación MA(1) con AR(1) se mantiene y resultando parámetros consistentes como muestra la tabla 12.

El análisis de los residuos se efectuó, obteniéndose la FAR y que también presenta como un ruido blanco lo que se presenta en la Fig. N° 28; para lo cual se procede a examinar estos residuos con la prueba de Box y Pierce; encontrándose el valor de $QB=24,8$ con 29 G.L. correspondiendo a aceptar a los residuos, con un nivel de confianza del 68,8%, como puede observarse que numéricamente hay error o es menos aceptar los residuos en comparación con el modelo inicial; sin embargo el número

TABLA Nº 11 ESTIMACION DE PARAMETROS Y ESTADISTICOS

Modelo: SARIMA (1,0,1)x(2,1,1)₁₂
 Diferenciación estacional de primer orden (D=1; s=12)
 Suma de cuadrados del error = 61,556
 Variancia Residual = 0,407656

```

=====
Parameter      Value      T-Statistic
-----
AR(1)          0,6497      6,18
MA(1)          0,1793      1,28
SAR(1)         -0,4050     -2,10
SAR(2)         -0,1014     -0,65
SMA(1)         0,5195      2,96
=====
    
```

Correlation Matrix

```

AR(1)  MA(1)  SAR(1)  SAR(2)  SMA(1)
-----
AR(1) | 1,00
MA(1) | 0,82  1,00
SAR(1) | -0,05 -0,00  1,00
SAR(2) | -0,00  0,04  0,84  1,00
SMA(1) | -0,02 -0,00  0,90  0,83  1,00
=====
    
```

TABLA Nº 12 PARAMETROS ESTIMADOS Y ESTADISTICOS

Modelo: SARIMA (1,0,1)x(1,1,1)₁₂
 Diferenciación estacional de primer orden (D=1; s=12)
 Suma de cuadrados del error = 63,213
 Variancia Residual = 0,415876

```

=====
Parameter      Value      T-Statistic
-----
AR(1)          0,8343      10,97
MA(1)          0,4221      3,52
SAR(1)         -0,2431     -2,41
SMA(1)         0,7419      10,33
=====
    
```

Correlation Matrix

```

AR(1)  MA(1)  SAR(1)  SMA(1)
-----
AR(1) | 1,00
MA(1) | 0,79  1,00
SAR(1) | -0,08 -0,01  1,00
SMA(1) | 0,10  0,13  0,58  1,00
=====
    
```

Número de observaciones : 143
 Media de la serie residual = 0,03477
 Desviación estandar de la serie residual = 0,6963052
 Diferenciación estacional de primer orden (D=1; s=12)

Lag	Value	T-Value	-1.0	0.0	+1.0
1	0.05	0.55	:	[*]	:
2	-0.05	-0.65	:	[*]	:
3	-0.17	-2.05	:	[***]	:
4	0.07	0.84	:	[**]	:
5	0.05	0.52	:	[*]	:
6	0.07	0.83	:	[**]	:
7	-0.05	-0.53	:	[*]	:
8	0.11	1.24	:	[***]	:
9	0.14	1.63	:	[***]	:
10	-0.01	-0.07	:	[]	:
11	0.03	0.33	:	[*]	:
12	0.01	0.15	:	[]	:
13	-0.05	-0.51	:	[*]	:
14	0.04	0.41	:	[*]	:
15	0.05	0.50	:	[*]	:
16	0.06	0.66	:	[**]	:
17	-0.07	-0.78	:	[**]	:
18	0.11	1.25	:	[***]	:
19	0.13	1.46	:	[***]	:
20	0.06	0.66	:	[**]	:
21	-0.05	-0.57	:	[*]	:
22	-0.12	-1.28	:	[***]	:
23	-0.00	-0.03	:	[]	:
24	0.06	0.60	:	[*]	:
25	0.04	0.41	:	[*]	:
26	0.03	0.33	:	[*]	:
27	0.03	0.32	:	[*]	:
28	0.02	0.24	:	[*]	:
29	0.05	0.52	:	[*]	:
30	0.04	0.42	:	[*]	:
31	-0.10	-1.02	:	[**]	:
32	-0.04	-0.41	:	[*]	:
33	-0.01	-0.15	:	[]	:

[] = Límites 2-errores estándar.

Estadístico Box-Pierce Chi-Cuadrado con 29 G.L. = 24,8
 Probabilidad = 0,688 (68,8%)

FIGURA Nº 28 FUNCION DE AUTOCORRELACION RESIDUAL
 MODELO: SARIMA (1,0,1)x(1,,1,1)₁₂

de parámetros es menor y que debe tenerse en cuenta por el principio de parsimonia. lo que nos lleva también a plantear un nuevo modelo según como se indica en la sec. 3.7.2; y este nuevo modelo a plantearse es sin la participación del parámetro MA(1) y esto sería un SARIMA(1,0,0)x(1,1,1)₁₂ lo que pasamos a estimar los parámetros correspondientes siguiendo la misma metodología anterior. Estos parámetros estimados se presentan en la tabla N° 13 al igual que las correlaciones entre los parámetros. este modelo presenta ciertas características de simplicidad de un modelo por tener un reducido número de parámetros. Pero al efectuar el análisis de residuos como se presenta en la Fig. N° 29 muestra el residuo dentro del intervalo de confianza con un indicador de Box y Pierce de QB=29,5 para 30 G.L. con un nivel de confianza de aceptar a los residuos como un ruido blanco de 49% . Lo que nos indica que el modelo no es el más indicado estadísticamente por lo que los probables modelos ya no serán con disminuir parámetros del modelo identificado.

Ahora podemos cambiar de modelo a partir del modelo inicial identificado, añadiendo o cambiando parámetros que pueden explicar con mayor claridad el fenómeno; así tenemos un modelo SARIMA (2,0,1)x(1,1,1)₁₂; modelo con doble efecto regresivo de orden regular y 1 de orden estacional; la estimación de estos parámetros se presentan en la tabla N° 14 lo que nos indica que los

TABLA Nº 13 PARAMETROS ESTIMADOS Y ESTADISTICOS

Modelo: SARIMA (1,0,0)x(1,1,1)₁₂
 Diferenciación estacional de primer orden (D=1; s=12)
 Suma de cuadrados del error = 65,097
 Variancia Residual = 0,425469

Parameter	Value	T-Statistic
AR(1)	0,5629	8,27
SAR(1)	-0,2315	-2,18
SMA(1)	0,6879	8,48

Correlation Matrix

	AR(1)	SAR(1)	SMA(1)
AR(1)	1,00		
SAR(1)	-0,12	1,00	
SMA(1)	-0,01	0,64	1,00

TABLA Nº 14 PARAMETROS ESTIMADOS Y ESTADISTICOS

Modelo: SARIMA (2,0,1)x(1,1,1)₁₂
 Diferenciación estacional de primer orden (D=1; s=12)
 Suma de cuadrados del error = 63,316
 Variancia Residual = 0,419308

Parameter	Value	T-Statistic
AR(1)	1,2639	11,23
AR(2)	-0,2636	-2,34
MA(1)	0,8040	11,32
SAR(1)	-0,2431	-2,42
SMA(1)	0,7659	11,02

Correlation Matrix

	AR(1)	AR(2)	MA(1)	SAR(1)	SMA(1)
AR(1)	1,00				
AR(2)	-0,80	1,00			
MA(1)	0,62	-0,62	1,00		
SAR(1)	-0,01	0,01	0,06	1,00	
SMA(1)	-0,13	0,13	-0,05	0,57	1,00

Número de observaciones : 143
 Media de la serie residual = -0,01678
 Desviación estandar de la serie residual = 0,7070534
 Diferenciación estacional de primer orden (D=1; s=12)

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.08	-0.95	:	[**]	:
2	0.09	1.12	:	[**]	:
3	-0.06	-0.71	:	[**]	:
4	0.17	2.04	:	[***]	:
5	0.09	1.05	:	[**]	:
6	0.12	1.40	:	[***]	:
7	-0.02	-0.21	:	[]	:
8	0.12	1.36	:	[***]	:
9	0.18	1.94	:	[****]	:
10	-0.00	-0.03	:	[]	:
11	0.10	1.04	:	[**]	:
12	0.00	0.03	:	[]	:
13	0.00	0.03	:	[]	:
14	0.07	0.70	:	[**]	:
15	0.00	0.00	:	[**]	:
16	0.09	0.91	:	[**]	:
17	-0.06	-0.60	:	[**]	:
18	0.11	1.14	:	[***]	:
19	0.12	1.20	:	[***]	:
20	0.05	0.52	:	[*]	:
21	-0.00	-0.05	:	[]	:
22	-0.00	-0.06	:	[**]	:
23	0.06	0.56	:	[*]	:
24	0.03	0.30	:	[*]	:
25	0.05	0.49	:	[*]	:
26	0.02	0.17	:	[]	:
27	0.05	0.49	:	[*]	:
28	-0.00	-0.01	:	[]	:
29	0.05	0.49	:	[*]	:
30	0.04	0.43	:	[*]	:
31	-0.00	-0.78	:	[**]	:
32	-0.02	-0.18	:	[]	:
33	-0.03	-0.26	:	[*]	:

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 30 Degrees of Freedom = 29,5
 Probability = 0,49

FIGURA Nº 29 FUNCION DE AUTOCORRELACION RESIDUAL
 MODELO: SARIMA (1,0,0)x(1,,1,1)₁₂

parámetros son consistentes y muestran cierta correlación entre ellos; los resultados del análisis de la FAR se muestra en la Fig. N° 30 lo que nos indica en comportamiento de los residuos más cercano al de un ruido blanco. La prueba de Box y Pierce para los residuos indica un $QB=20,4$ y con 28 G.L. corresponden a un nivel de confianza del 85% de aceptar los residuos como un ruido blanco, modelo que tendremos en cuenta para poder elegir el mejor modelo;

Finalmente podemos estimar los parámetros del modelo SARIMA $(2,0,1) \times (2,1,1)$, modelo que aparentemente nos indicará mayor explicación de la serie por tener un mayor número de parámetros; sin embargo los resultados de los estimados que se presentan en la tabla N° 15, y la Figura N° 31 presentan relativamente alta la correlación entre los parámetros y la prueba de los residuos con un Box y Pierce de $QB = 26,4$ con 24 G.L.; correspondiente a un chi-cuadrado de 33,3% de aceptar a los residuos como un ruido blanco. Este resultado estadísticamente es muy bajo ya que podemos decir que el modelo no es aceptado.

Número de observaciones : 143
 Media de la serie residual = -0,03407
 Desviación estandar de la serie residual = 0,6980929
 Diferenciación estacional de primer orden (D=1; s=12)

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.00	-0.01	:	[]	:
2	0.01	0.13	:	[]	:
3	-0.15	-1.76	:	[***]	:
4	0.08	0.93	:	[**]	:
5	-0.00	-0.02	:	[]	:
6	0.06	0.66	:	[*]	:
7	-0.08	-0.93	:	[**]	:
8	0.08	0.92	:	[**]	:
9	0.11	1.21	:	[***]	:
10	-0.04	-0.48	:	[*]	:
11	-0.01	-0.06	:	[]	:
12	0.02	0.23	:	[]	:
13	-0.08	-0.92	:	[**]	:
14	0.01	0.10	:	[]	:
15	0.01	0.10	:	[]	:
16	0.04	0.48	:	[*]	:
17	-0.09	-1.02	:	[**]	:
18	0.10	1.14	:	[***]	:
19	0.09	1.00	:	[**]	:
20	0.04	0.39	:	[*]	:
21	-0.07	-0.74	:	[**]	:
22	-0.14	-1.51	:	[***]	:
23	-0.03	-0.32	:	[*]	:
24	0.05	0.50	:	[*]	:
25	0.01	0.16	:	[]	:
26	0.01	0.06	:	[]	:
27	0.01	0.11	:	[]	:
28	0.02	0.16	:	[]	:
29	0.02	0.19	:	[]	:
30	0.02	0.27	:	[*]	:
31	-0.12	-1.29	:	[***]	:
32	-0.05	-0.50	:	[*]	:
33	-0.06	-0.62	:	[*]	:

[] = Límites 2-errores estándar.

Estadístico Box-Pierce Chi-Cuadrado con 28 G.L. = 20,9.
 Probabilidad = 0,85 (85%).

FIGURA Nº 30 RESIDUAL AUTOCORRELATION FUNCTION
 Modelo: SARIMA(2,0,1)x(1,1,1)₁₂

TABLA Nº 15 PARAMETROS ESTIMADOS Y ESTADISTICOS

Modelo: SARIMA (2,0,1)x(2,1,1)₁₂
 Diferenciación estacional de primer orden (D=1; s=12)
 Suma de cuadrados del error = 60,84
 Variancia Residual = 0,4056

Parameter	Value	T-Statistic
AR(1)	-0,1567	-0,80
AR(2)	0,4265	3,75
MA(1)	-0,6397	-3,02
SAR(1)	-0,4956	-2,21
SAR(2)	-0,1561	-0,90
SMA(1)	0,3995	1,87

Correlation Matrix

	AR(1)	AR(2)	MA(1)	SAR(1)	SAR(2)	SMA(1)
AR(1)	1,00					
AR(2)	-0,66	1,00				
MA(1)	0,93	-0,77	1,00			
SAR(1)	-0,01	-0,05	0,01	1,00		
SAR(2)	-0,00	-0,03	0,03	0,88	1,00	
SMA(1)	-0,03	-0,02	-0,03	0,92	0,86	1,00

Número de observaciones : 143
 Media de la serie residual = 0,05917
 Desviación estandar de la serie residual = 0,710435
 Diferenciación estacional de primer orden (D=1; s=12)

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.00	-0.03	:	[]	:
2	-0.03	-0.32	:	[*]	:
3	-0.00	-0.02	:	[]	:
4	0.08	0.91	:	[**]	:
5	0.05	0.58	:	[*]	:
6	0.09	1.07	:	[**]	:
7	-0.07	-0.82	:	[**]	:
8	-0.01	-0.06	:	[]	:
9	0.16	1.07	:	[***]	:
10	-0.03	-0.35	:	[*]	:
11	0.07	0.76	:	[**]	:
12	-0.03	-0.20	:	[*]	:
13	-0.10	-1.19	:	[***]	:
14	0.01	0.10	:	[]	:
15	0.13	1.45	:	[***]	:
16	-0.04	-0.40	:	[*]	:
17	-0.13	-1.48	:	[***]	:
18	0.12	1.35	:	[***]	:
19	0.06	0.65	:	[**]	:
20	-0.05	-0.53	:	[*]	:
21	-0.01	-0.07	:	[]	:
22	-0.19	-2.00	:	[****]	:
23	0.06	0.66	:	[**]	:
24	0.00	0.02	:	[]	:
25	-0.02	-0.16	:	[]	:
26	-0.03	-0.35	:	[*]	:
27	0.12	1.23	:	[***]	:
28	-0.08	-0.79	:	[**]	:
29	-0.00	-0.00	:	[]	:
30	0.05	0.56	:	[*]	:

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 24 Degrees of Freedom = 26.4
 Probability = 0,333

FIGURA Nº 31 FUNCION DE AUTOCORRELACION RESIDUAL
 MODELO: SARIMA(2,0,1)x(2,1,1)₁₂

Como podemos haber visto los resultados de los modelos planteado y selección del modelo al aceptar a los residuos como ruido blanco con 68,8, 73,8 y 85% de probabilidad y que son representados por 4, 5 y 5 parámetros respectivamente, lo que indica que el principio de parsimonia no es muy importante en esta selección y más aún ahora que las computadoras están al alcance, entonces, nuestro modelo que representa la serie de las descargas medias mensuales serán: el modelo SARIMA (1,0,1)x(2,1,1)₁₂ y el SARIMA(2,0,1)x(1,1,1)₁₂; ambos modelos representados por 5 parámetros y que nuevamente serán estimados utilizando el Micro T.S.P.

Modelo SARIMA(1,0,1)x(2,1,1)₁₂:

para $W_t = Z_t - Z_{t-12}$; y Z_t la transformada de Box y Cox; donde: W_t se define estacionaria, según Ec.4.1 y las Ecs. 3.69, 4.3 y 4.24; con $d=0$ y $D=1$ al reemplazar Z_t por W_t el modelo queda:

$$\phi(B)\phi(B)W_t = \theta(B)\delta(B)a_t \quad (6.1)$$

con: $\phi(B) = 1 - \phi_1 B$; $\phi(B) = 1 - \phi_1 B^{12} - \phi_2 B^{24}$; $\theta(B) = 1 - \theta_1 B$; $\delta(B) = 1 - \delta_1 B^{12}$; reemplazando $\phi(B)$, $\phi(B)$, $\theta(B)$ y $\delta(B)$ en 6.1 y ordenando se tiene la siguiente expresión:

$$W_t = \phi_1 W_{t-1} + \phi_1 W_{t-12} - \phi_1 \phi_1 W_{t-13} + \phi_2 W_{t-24} - \phi_1 \phi_2 W_{t-25} + a_t - \theta_1 a_{t-1} - \delta_1 a_{t-12} + \theta_1 \delta_1 a_{t-13} \quad (6.2)$$

reemplazando los valores de la tabla N^o 11, estimados con

el Forecast Plus, para:

$$\begin{aligned} \phi_1 &= 0,6497 & ; & & \theta_1 &= 0,1793 \\ \phi_1 &= \phi_{12} = -0,4050 & ; & & \delta_1 &= \theta_{12} = 0,5195 \\ \phi_2 &= \phi_{24} = -0,1014 \end{aligned}$$

$$\begin{aligned} W_t = & 0,6497W_{t-1} - 0,4050W_{t-12} + 0,2631W_{t-13} - 0,1014W_{t-24} + \\ & + 0,0659W_{t-25} + a_t - 0,1793a_{t-1} - 0,5195a_{t-12} + 0,0931a_{t-13} \end{aligned} \quad (6.3)$$

para poder estimar con el Micro T P por el método de mínimo cuadrado se considera el modelo equivalente:

SARMA(1,1)x(2,1)₁₂ y la Ec. 6.2 con los parámetros: MA(1), MA(12), MA(13), AR(1), AR(12), AR(13), AR(24) y AR(25).

Observándose que la interacción de $\phi_1\phi_1 = \phi_{13}$ así $\theta_1\delta_1 = \theta_{13}$; obteniéndose así valores muy próximos a los estimados por el Forecast Plus y el Statgraphics.

El modelo 6.2 en función de Z_t viene dado por la siguiente relación:

$$\begin{aligned} Z_t = & \phi_1 Z_{t-1} + (1+\phi_1)Z_{t-12} - (\phi_1\phi_1 + \phi_1)Z_{t-13} + (\phi_2 - \phi_1)Z_{t-24} + (\phi_1\phi_1 - \\ & \phi_1\phi_2)Z_{t-25} - \phi_2 Z_{t-36} + \phi_1\phi_2 Z_{t-37} + a_t - \theta_1 a_{t-1} - \delta_1 a_{t-12} + \theta_1 \delta_1 a_{t-13} \end{aligned} \quad (6.4)$$

de acuerdo a los valores de la tabla N° 11 y reemplazando en la Ec. 6.4 se tiene el modelo:

$$\begin{aligned} Z_t = & 0,6497Z_{t-1} + 0,595Z_{t-12} - 0,38657Z_{t-13} + 0,3036Z_{t-24} - \\ & 0,1973Z_{t-25} + 0,1014Z_{t-36} - 0,06588Z_{t-37} + a_t - 0,1793a_{t-1} - \\ & 0,5195a_{t-12} + 0,09314a_{t-13} \end{aligned} \quad (6.5)$$

El modelo equivalente a 6.4 y estimado por el TSP es:
 MA(1), MA(12), MA(13), AR(1), AR(12), AR(13), AR(24),
 AR(25), AR(36) y AR(37) estos valores se presenta en la
 tabla N° 16 para el modelo SARIMA (1,0,1)x(2,1,1)₁₂ y la
 función de los residuos para aceptar a los residuos como
 ruido blanco se presenta en la figura N° 32 aceptándolos
 con un Box y Pierce Chi-cuadrado con 25 GL. igual a QB =
 18,29 de 95,38% de probabilidad es de notar que el valor
 de QB para la misma muestra y G.L. es mayor y en varios
 modelos arroja valores relativamente altos, esto puede
 deberse al no considerar las interacciones de los
 parámetros y el algoritmo de mínimos cuadrados muchas
 veces no es el más recomendable.

TABLA N° 16 PARAMETROS ESTIMADOS DEL
 MODELO : SARIMA(1,0,1)x(2,1,1)₁₂

LS // Variable Dependiente : Ln(caudal)
 Convergencia despues de 8 iteraciones

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
MA(1)	-0,2315515	0.1636735	-1.4147157	0.1600
MA(12)	-0,7399161	0.0672967	-10.994036	0.0000
MA(13)	0,0613190	0.1754042	0.3495915	0.7273
AR(1)	0,7001009	0.1387692	5.0451330	0.0000
AR(12)	0,6859576	0.0980540	6.9957135	0.0000
AR(13)	-0,3657515	0.1175011	-3.1127490	0.0024
AR(24)	0,2975937	0.0845567	3.5194506	0.0006
AR(25)	-0,1975609	0.1130753	-1.7472327	0.0834
AR(36)	-0,0419212	0.0971306	-0.4315608	0.6669
AR(37)	-0,0715651	0.0991904	-0.7214922	0.4722
R-squared	0,823917	Mean of dependent var	1.046591	
Adjusted R-squared	0,809370	S.D. of dependent var	1.627937	
S.E. of regression	0,710762	Sum of squared resid	55.06490	
Log likelihood	-123,0024	P-statistic	56.66944	

Number of observations: 121

Autocorrelaciones	Autocorrelación Parcial	PAC	PAP
. *	. *	1 0.042	0.042
. *	. *	2 0.070	0.068
.* .	.* .	3 -0.100	-0.177
. *	. *	4 0.045	0.050
.* .	. .	5 -0.039	-0.030
. .	. .	6 -0.018	-0.033
.* .	.* .	7 -0.083	-0.067
. .	. .	8 0.024	0.026
. **	. **	9 0.143	0.152
. *	. .	10 0.060	0.030
. *	. *	11 0.100	0.089
. .	. .	12 0.023	0.033
. .	.* .	13 -0.031	-0.056
. .	. .	14 -0.021	-0.003
. .	. .	15 0.013	0.028
.* .	.* .	16 -0.079	-0.065
*** .	*** .	17 -0.203	-0.200
. .	. .	18 -0.002	0.020
. .	. .	19 0.032	0.033
.* .	** .	20 -0.069	-0.161
. .	. .	21 -0.014	-0.008
.* .	.* .	22 -0.069	-0.055
. *	. *	23 0.093	0.066
. .	.* .	24 -0.026	-0.047
. .	. .	25 -0.003	0.014
. .	. *	26 -0.019	0.006
. .	. .	27 0.011	-0.013
.* .	.* .	28 -0.092	-0.053
.* .	.* .	29 -0.076	-0.042
.* .	.* .	30 -0.065	-0.056
Box-Pierce Q-Stat 18,29	Prob 0,9537	SR of Correlations	0.091
Ljung-Box Q-Stat 21,33	Prob 0.8773		

FIGURA Nº 32 FUNCION DE AUTOCORRELACIONES DEL RESIDUAL

MODELO: SARIMA(1,0,1)x(2,1,1)₁₂

El modelo equivalente a 6.4 encontrado según el TSP será:

$$Z_t = \phi_{11}Z_{t-1} + \phi_{12}Z_{t-12} + \phi_{13}Z_{t-13} + \phi_{24}Z_{t-24} + \phi_{25}Z_{t-25} + \\ + \phi_{36}Z_{t-36} + \phi_{37}Z_{t-37} + a_t - \theta_{11}a_{t-1} - \theta_{12}a_{t-12} + \theta_{13}a_{t-13} \quad (6.4b)$$

reemplazando valores de la tabla N° 16 se tiene:

$$Z_t = 0,700Z_{t-1} + 0,686Z_{t-12} - 0,3657Z_{t-13} + 0,2976Z_{t-24} - \\ 0,1976Z_{t-25} - 0,042Z_{t-36} - 0,07156Z_{t-37} + a_t - 0,232a_{t-1} - \\ 0,7399a_{t-12} + 0,0613a_{t-13} \quad (6.6)$$

Similarmente con los parámetros estimados se procede a la representación del modelo SARIMA(2,0,1)x(1,1,1)₁₂ :

$$W_t = 1,2635W_{t-1} - 0,2639W_{t-2} - 0,2431W_{t-12} + 0,3072W_{t-13} - \\ 0,0641W_{t-14} + a_t - 0,804a_{t-1} - 0,7659a_{t-12} + 0,61578a_{t-13}.$$

para $W_t = Z_t - Z_{t-12}$, estacionario e invertible.

reemplazando a W_t , se tiene:

$$Z_t = 1,2635Z_{t-1} - 0,2639Z_{t-2} + 0,7569Z_{t-12} - 0,9564Z_{t-13} + \\ + 0,1997Z_{t-14} + 0,2431Z_{t-24} - 0,3072Z_{t-25} + 0,0641Z_{t-26} + a_t \\ - 0,804a_{t-1} - 0,7659a_{t-12} + 0,61578a_{t-13} . \quad (6.7)$$

Predicción con los Modelos Seleccionados

Según la sec. 4.5 de predicción con modelos ARIMA; se efectúa la predicción para encontrar valores futuros de set. 1987 a ago.1988 según las ecuaciones 6.5 y 6.7 y en base a una selección adecuada del horizonte de análisis,

para ello, por lo general el origen está constituido por la última observación con la finalidad de utilizar toda la información posible; pero en algunas veces es necesario considerar otros orígenes debido a la presencia de aleatoriedad y a causas naturales que pueda presentarse (años secos, húmedos y/o normales); el pronóstico será el promedio de los pronósticos de estos diferentes orígenes y además nos permite reducir la aleatoriedad.

Consideramos como origen de las predicciones los períodos de tiempo $t = 156, 132$ y 120 . Estos orígenes seleccionados; 132 porque los modelos consideran parámetros con atrasos hasta 37 meses y considerando un año normal como inicio del pronóstico; origen 120 debido que es el final del período de presencia del pico ó año irregular que puede llegar a ser significativa su inclusión, considerando que los efectos son mayores que un año; así la descarga estimada para setiembre de 1987 está en función de las descargas de ago.87, set.86, ago.86, set.85, ago.85, set.84, ago.84 para el modelo $SARIMA(1,0,1) \times (2,1,1)_{12}$ y para el modelo $SARIMA(2,0,1) \times (1,1,1)_{12}$ estará en función de ago.87, jul.87, set.86, ago.86, jul.86, set.85, ago.85, jul.85. En base a estas consideraciones se procedió a efectuar los pronósticos para los diferentes orígenes y calculando el promedio del pronóstico se presenta en la tabla N° 17 para los dos modelos.

TABLA Nº 17 PREDICCIÓN PARA ORIGENES 156, 132 y 120.

MODELO: SARIMA(1,0,1)x(2,1,1)₁₂

Mes\Origen	t =			PROMEDIO
	156	132	120	
SETIEMBRE	0,943	1,160	1,096	1,066
OCTUBRE	2,857	3,148	1,882	2,629
NOVIEMBRE	4,480	4,335	3,198	4,004
DICIEMBRE	18,788	15,427	9,179	14,465
ENERO	40,263	24,837	28,625	31,242
FEBRERO	104,061	71,191	53,581	76,278
MARZO	69,105	73,430	63,768	68,768
ABRIL	40,016	41,779	34,418	38,738
MAYO	9,103	6,890	4,852	6,948
JUNIO	4,276	3,085	1,954	3,105
JULIO	2,082	1,511	1,189	1,594
AGOSTO	1,798	1,314	1,042	1,384

MODELO: SARIMA(2,0,1)x(1,1,1)₁₂

Mes\Origen	156	132	120	PROMEDIO
SETIEMBRE	0,775	1,825	1,213	1,271
OCTUBRE	1,784	3,920	1,998	2,567
NOVIEMBRE	3,231	7,583	4,629	5,147
DICIEMBRE	9,223	18,977	9,600	12,600
ENERO	21,996	40,476	29,507	30,660
FEBRERO	53,172	100,621	57,752	70,515
MARZO	41,024	97,279	59,080	65,795
ABRIL	21,387	46,451	26,535	31,458
MAYO	4,466	8,210	4,486	5,721
JUNIO	1,955	3,619	1,909	2,494
JULIO	1,012	2,001	1,194	1,402
AGOSTO	0,869	1,664	0,982	1,171

Como se indicó en la Sec. 4.5.4. sobre los intervalos de confianza y las Ecs. 4.54, 4.55 y 4.56; además considerando la transformación de Box y Cox efectuada a la serie según la Ec. 4.59 se tiene los intervalos de confianza a un 95%, la Ec. 4.61, donde $v_{\epsilon}(h)$, es la variancia residual estimada por cada modelo; presentándose en las Tablas N^o 11 y 14 para los modelos $SARIMA(1,0,1) \times (2,1,1)_{12}$ y $SARIMA(2,0,1) \times (1,1,1)_{12}$ respectivamente y el valor de $Y_{\epsilon}(h)$ es el logaritmo natural del caudal estimado para origen t y horizonte h ; los resultados del intervalo de confianza se presenta en la tabla N^o 18.

Errores de la Predicción

Para comparar los modelos seleccionados se calculan los errores de la predicción con los datos observados para set.87 a ago.88 así encontrándose los errores y mostrándose en la tabla N^o 19 ; de igual manera se calculan la suma de los errores al cuadrado para cada modelo obteniéndose la tabla N^o 20 lo que indica que el modelo que presentaba un Box y Pierce de aceptar los residuos con 85% presenta mayor error que el modelo que presenta un 73,9% indicando que este último modelo sería el mejor para los pronósticos y presentado buen pronóstico para los orígenes 156 y 120; para poder encontrar en

TABLA Nº 18 INTERVALOS DE CONFIANZA AL 95%

SARIMA(1,0,1)x(2,1,1)₁₂

Var (R) = 0,407656

	PRONOSTICO		L I M I T E S	
		OBSERVACION	INFERIOR	SUPERIOR
SET	1,066	0,972	0,305	4,077
OCT	2,629	2,295	0,752	9,538
NOV	4,004	4,678	1,146	14,346
DIC	14,465	8,363	4,138	50,908
ENE	31,242	33,753	8,938	109,549
FEB	76,278	105,264	21,823	266,963
MAR	68,768	47,568	19,674	240,714
ABR	38,738	49,697	11,083	135,749
MAY	6,948	14,984	1,988	24,635
JUN	3,105	4,910	0,888	11,203
JUL	1,594	2,149	0,456	5,920
AGO	1,384	0,863	0,396	5,189

SARIMA(2,0,1)x(1,1,1)₁₂

Var (R) = 0,419308

	L I M I T E S			
	PREDICION	OBSERVACION	INFERIOR	SUPERIOR
SET	1,271	0,972	0,357	4,521
OCT	2,567	2,295	0,722	9,134
NOV	5,147	4,678	1,447	18,314
DIC	12,600	8,363	3,541	44,831
ENE	30,660	33,753	8,617	109,085
FEB	70,515	105,264	19,819	250,887
MAR	65,795	47,568	18,492	234,092
ABR	31,458	49,697	8,842	111,924
MAY	5,721	14,984	1,608	20,353
JUN	2,494	4,910	0,701	8,875
JUL	1,402	2,149	0,394	4,989
AGO	1,171	0,863	0,329	4,168

TABLA Nº 19 ERRORES DE LA PREDICCIÓN SET.87 AGO.88

	OBSERVADA	SARMA(1,1)x(2,1)		SARMA(2,1)x(1,1)	
		PROM1	ERROR1	PROM2	ERROR2
SETIEMBRE	0,972	1,066	-0,094	1,271	-0,299
OCTUBRE	2,295	2,629	-0,334	2,567	-0,272
NOVIEMBRE	4,678	4,004	0,674	5,147	-0,469
DICIEMBRE	8,363	14,465	-6,102	12,600	-4,237
ENERO	33,753	31,242	2,511	30,660	3,093
FEBRERO	105,264	76,278	28,986	70,515	34,749
MARZO	47,568	68,768	-21,200	65,795	-18,227
ABRIL	49,697	38,738	10,960	31,458	18,239
MAYO	14,984	6,948	8,036	5,721	9,264
JUNIO	4,910	3,105	1,805	2,494	2,416
JULIO	2,149	1,594	0,555	1,402	0,747
AGOSTO	0,863	1,384	-0,521	1,171	-0,308
		SUM =	25,275	SUM =	44,695

TABLA Nº 20 SUMA DE CUADRADOS DE LOS ERRORES DE PREDICCIÓN

	SARMA(1,1)(2,1)		% A/B
	SARMA(1,1)(2,1)	SARMA(2,1)x(1,1)	
SETIEMBRE	0,01	0,09	10,002
OCTUBRE	0,12	0,16	73,706
NOVIEMBRE	0,57	0,38	149,694
DICIEMBRE	37,81	18,34	206,161
ENERO	44,11	27,91	158,073
FEBRERO	884,33	1235,41	71,582
MARZO	1333,77	1567,61	85,083
ABRIL	1453,88	1900,29	76,508
MAYO	1518,45	1986,10	76,454
JUNIO	1521,71	1991,94	76,393
JULIO	1522,02	1992,49	76,388
AGOSTO	1522,29	1992,59	76,398
	(A)	(B)	
SECMP	1522,290	1992,589	

cuanto es bueno uno y el otro se calculó el porcentaje de los errores al cuadrado acumulativos, presentándose en la tabla N^o 20

Puede verse que el modelo SARIMA(1,0,1)x(2,1,1)₁₂ presenta resultados más satisfactorios para los orígenes tomados y muy confiables para los meses de setiembre 87 y de febrero 1988 a agosto 88; en cambio para los meses de noviembre, diciembre (87) y enero (88) es superior el modelo SARIMA(2,0,1)x(1,1,1) siendo más confiable para el período de origen 132; de esta manera determinamos los pronósticos y podemos ver el comportamiento de estos resultados y de los observados (sin incluir los dos picos existentes) en la figura N^o 33, hasta agosto de 1988.

Con el modelo obtenido SARIMA(2,0,1)x(1,1,1)₁₂ y los datos hasta Ago.88 encontramos las predicciones para Set.88 a Ago.93 para orígenes de 168, 156 y 144, el promedio de las predicciones y sus respectivos intervalos de confianza (95%) que se muestra en la Fig. N^o 34 y los valores estimados se presenta en sec. Anexo.

Como puede observarse que las descargas estimadas presento para un período muy grande a pesar que por lo general los modelos en series de tiempo son aplicables a períodos cortos; es por ello tomar estos resultados con cautela y con un intervalo de confianza de 95%, y actualizarlos estas predicciones cuando se obtienen los datos de registro.

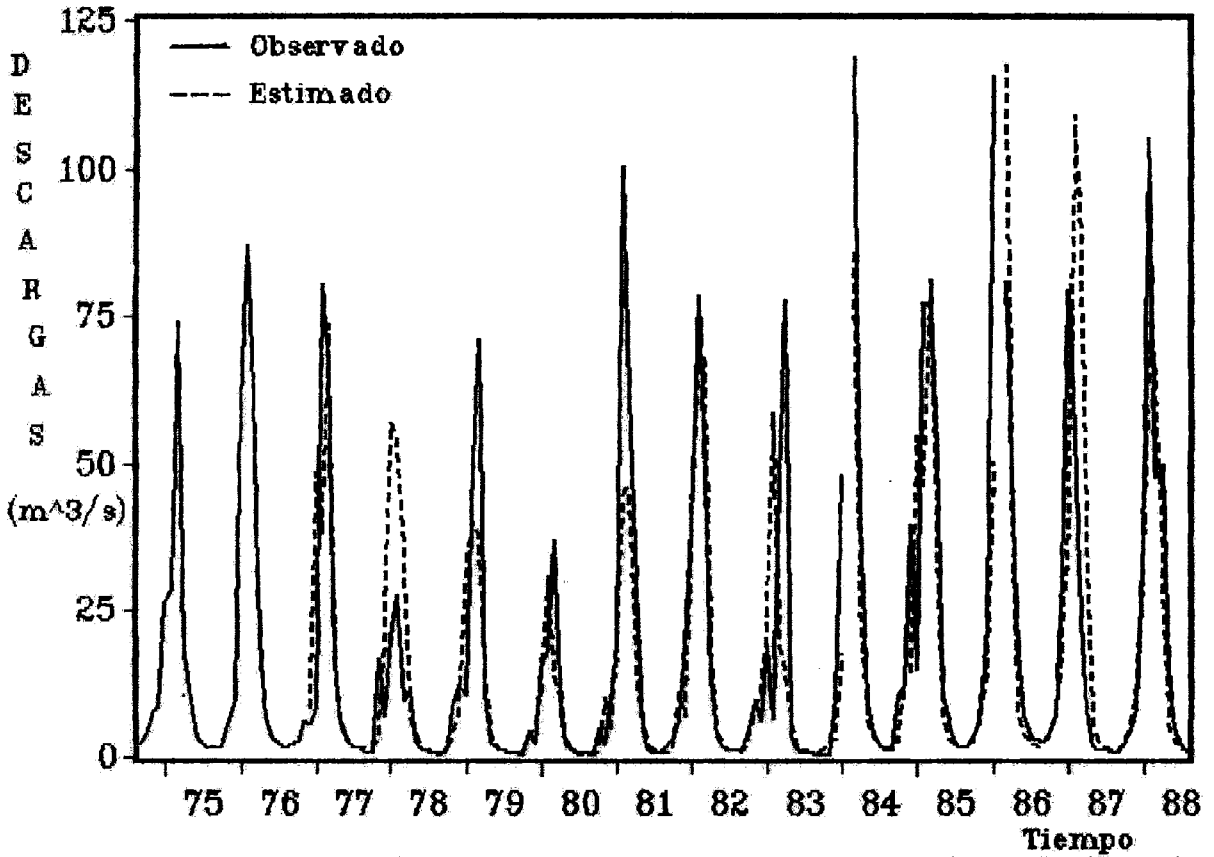


Fig. N 33a Pronósticos con el Modelo SARIMA(2,0,1)x(1,1,1)

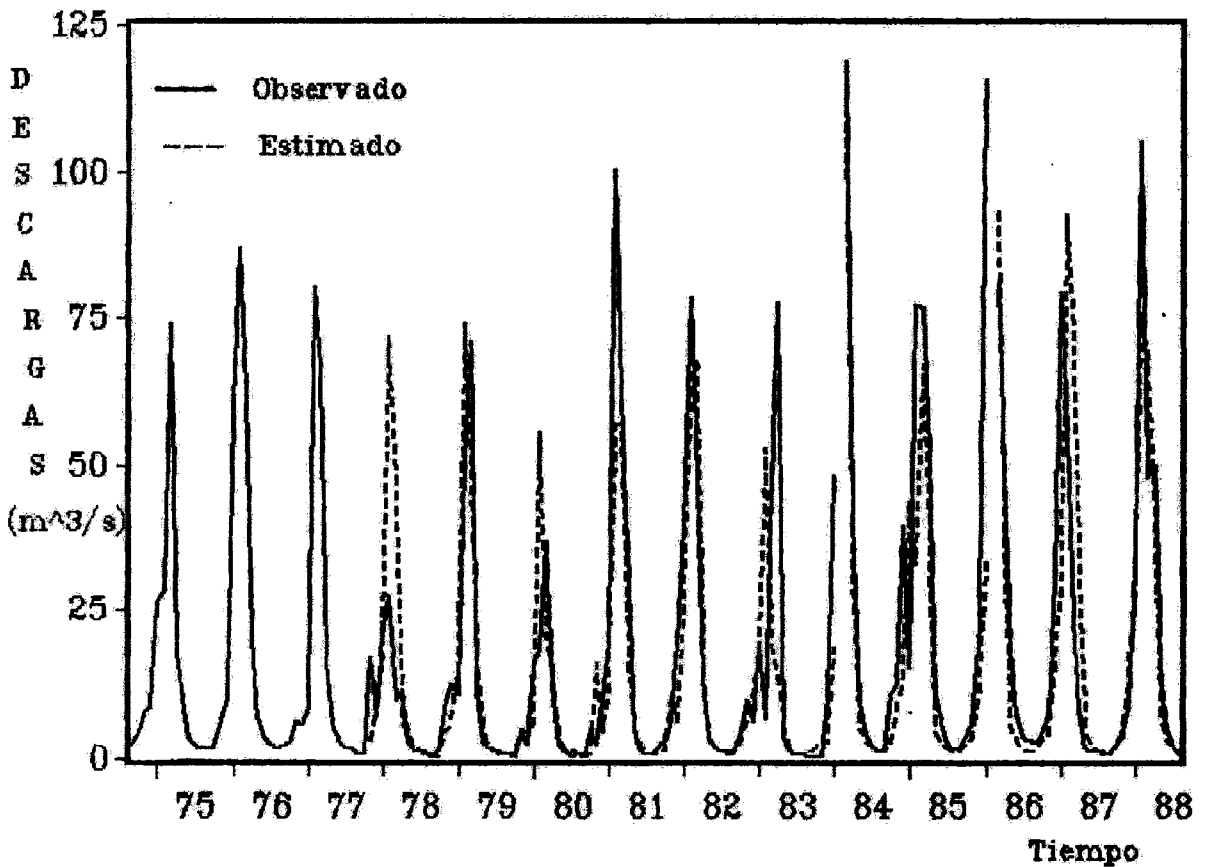


Fig. N 33b Pronósticos con el Modelo SARIMA(1,0,1)x(2,1,1)

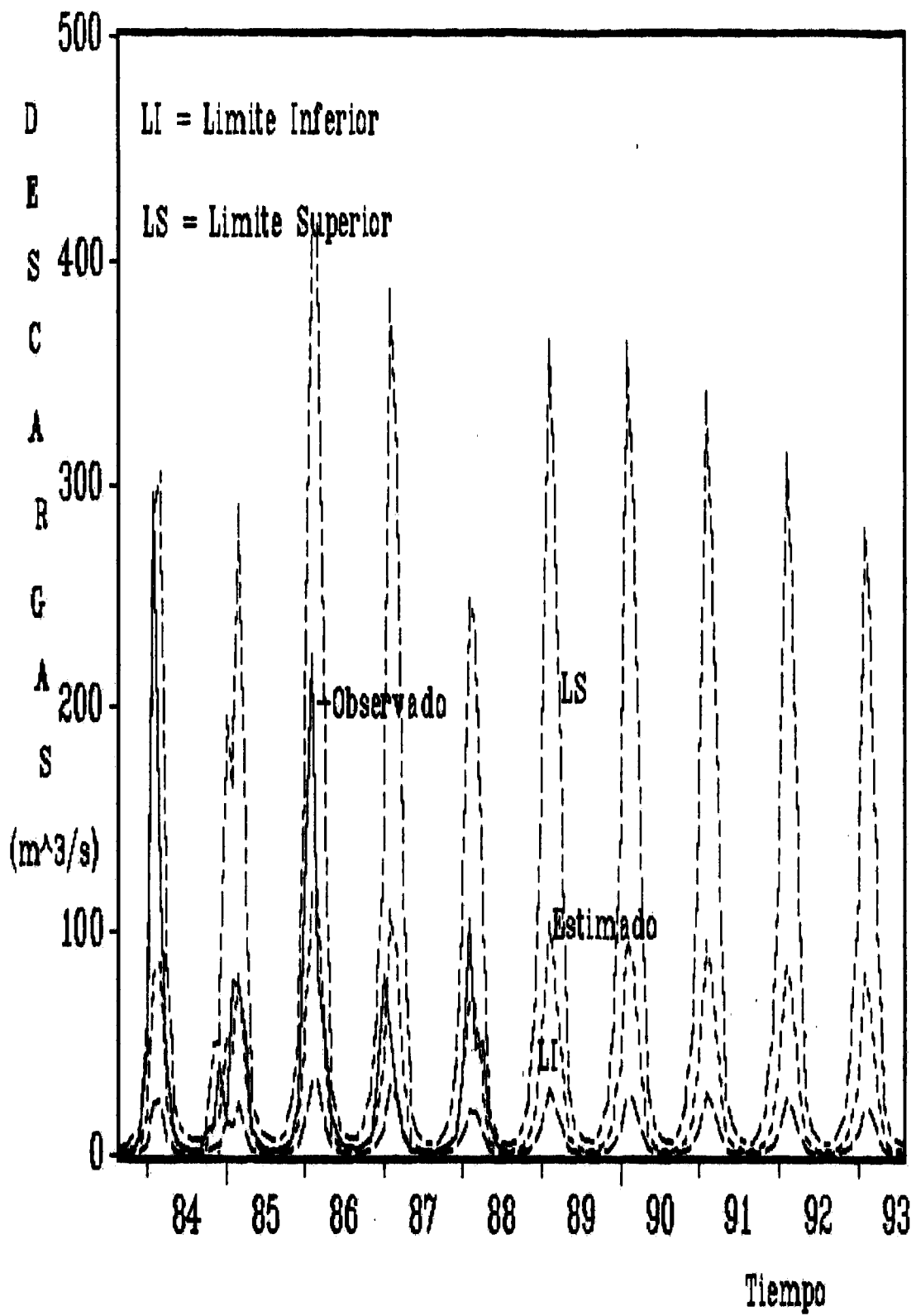


Fig. N 34 Pronósticos hasta Ago.93-Modelo SARIMA(2,0,1)(1,1,1)

VII. CONCLUSIONES Y RECOMENDACIONES

7.1 CONCLUSIONES

Las conclusiones son las siguientes:

- 1) De la serie de las descargas medias mensuales del río Pisco; en el Análisis Exploratorio de los Datos, presenta un comportamiento estacional, presentando datos no homogéneos y necesidad de aplicar una transformación no lineal de Box y Cox de tipo logarítmica obteniéndose una serie simétrica (que involucra normalidad), dispersión uniforme logrando que los efectos de dos factores sean aditivos sin interacción entre ellos (no inclusión de cuadrados ni productos cruzados).
- 2) La serie es diferenciada en relación al atraso unitario estacional ($D=1$), para producir estacionariedad.
- 3) El mejor modelo que representa a las descargas medias mensuales del río Pisco es el $SARIMA(2,0,1) \times (1,1,1)_{12}$ y el $SARIMA(1,0,1) \times (2,1,1)_{12}$.
- 4) De la verificación (set.97 a ago.88), para setiembre 87 y para los meses de febrero 88 a agosto 88 pronostica mejor el modelo $SARIMA(1,0,1)(2,1,1)_{12}$ en relación al modelo $SARIMA(2,,0,1) \times (1,1,1)_{12}$ el cual da mejores resultados para los meses de octubre, noviembre, diciembre (87) y enero (88).

- 5) Con los modelos obtenidos y con los datos actualizados se predice las descargas medias mensuales del Río Pisco con un intervalo de confianza del 95%, para tiempos futuros.

7.1 RECOMENDACIONES

- 1) Para realizar nuevas predicciones es necesario tener en cuenta la transformación logarítmica, actualizar la información y considerar un período mayor de información (unos 10 a 15 años más) para el estudio ciclos.
- 2) Realizar predicciones para períodos cortos (1 a 2 años) y para años normales, secos y húmedos.
- 3) Determinar Modelos de predicción para los demás ríos del Perú con la metodología utilizada en la presente tesis.
- 4) Utilizar modelos multivariados teniendo como base estos modelos de Box y Jenkins, a las descargas, precipitación, operación de embalse, etc.

VIII BIBLIOGRAFIA

- 1.- ANDERSON T.W. (1971), "The Statistical Analysis of Time Series" Ed. Jhon Wiley & Sons; Inc New York
- 2.- BOX G.P. y JENKINS G.M. (1976), "Time Series Analysis" Forecast and control hydrology. Holden-Day San Francisco; 2^a Ed. 1976.
- 3.- BANON G.J.F. (1971), "Etude d'algorithmes d'estimation des parametres pour l'idenfication adaptative en temps reel des processus linéares pertubés par un bruité corrélé. (These de Docteur-Ingénieur). Université Paul Sabatier, Toulouse, Jan. 1971;
- 4.- BENJAMIN J.R. y C.C. CORNELL (19), "Probabilidad y Estadística Aplicado a la Ingenieria Civil" Mc. Graw Hill;
- 5.- BROCKWELL, P.J. y DAVIS R.A. (1987), "Time Series: Theory and Methods"; Berlin, Springer-Verlag;
- 6.- CIRIANI T.A.; et al (1974), "Matemtical Models for Surface" Water Hydrology; IBM Scientific Center, Pisa Italy;
- 7.- CLARKE B.R.T. (1973), "Matemtical models in hydrology"; Institute of Hidrology of the United Nation; Roma (F.A.O.);
- 8.- DUDGEON D.E; MERSEREAN R.M. (1984), "Multidimensional Digital Signal Processing"; Englewood Cliffs, NJ, Prentice Hall;
- 9.- GRANGER, C.W.J; NEWBOLD P. (1977), "Forecasting Economic Time Series"; New York, NY Academic Press;
- 10.- HOAGLIN D.C; MASTELLER F.; TUKEY JHON W. (1983), "Understanding Robust and Exploratory Data Analysis" Jhon Willey & Sons, Inc New York.
- 11.- HALTINER J.P. y SALAS J.D. (1988); "Development and Testing of A Multivariate, Seasonal ARMA(1,1) Model. Hydrology and Water Resources Program; Departament of Civil Engineering Colorado U.S.A.

- 12.- GENOVEZ A.M. y CHAUDHRY F.H.; (1983) "Modelo Estocástico Das Vazões Mensais do rio Tiete em Cumbica. RBE. V Simposio Brasileiro de Hidrologia e Recursos Hídricos; Blumenau, SC Brasil Nov.
- 13.- JIMENES D.L. (1991); "Modelos Económicos y de Series de Tiempo mediante el uso del TSP (Time Series Processor); Centro de Investigación del Dpto. de Economía y Planificación UNALM
- 14.- LANNA A.E. y INDURSKY A.B. (1982); "Utilização de Séries Sintéticas Geradas por Modelos Multivariados Auto-regressivo-AR(1) e Auto-regressivo com Média Móvel - ARMA(1,1) no Dimensionamento da Capacidade de Reservatório Fluvial; Revista Brasileira de Engenharia. Asc. Brasilina de Hidrologia e Recursos Hídricos E.R.H. Vol. 1 Nº 1 -Dezembro 1982.
- 15.- MARPLE S.L. (1987); "Digital Spectral Analysis" Englewood Cliffs, N.J. Prentice-Hall
- 16.- MORETTIN P.A. y TOLDO C.M.C. (1987); "Previsão de Séries Temporais"; Sao Paulo 2^{da} Edic. (1^{ra} Edición 1981);
- 17.- NELSON C.R. (1973); "Applied Time Series Analysis for Managerial Forecasting" San Francisco, CA. Holden Day;
- 18.- PEREIRA B.B. (1984); "Series Temporais Multivariadas" Curso apresentado durante o V Simpósio Nacional de Probabilidade e Estatística, Rio de Janeiro, realizado na UFRJ, 1984;
- 19.- PRIESTLEY, M.B. (1981) "Spectral analysis and time series" London, Academic Press. 1981. (Univariate Series, v.1);
- 20.- PULIDO Antonio, (1980); "Modelos Económicos"; España.
- 21.- ROBINSON, E.A.; TREITEL S. (1980); "Geophysical signal analysis"; Englewood Cliff, NJ, Prentice-Hall,

- 22.- SOTOMAYOR R.R. (1989); "Análisis de Series de Tiempo Mediante Modelos ARIMA con fines de pronóstico, comparación con los Modelos de Descomposición"; Tesis para optar el título de Ing. Estadístico, UNALM.,
- 23.- STEDINGER J.R; LETTENMAIER D.P; VOGEL R.M. (1985); "Multisite ARMA(1,1) and Disaggregation Models for Annual Streamflow Generation"; Water Resources Vol. 21 No 4 abril. 1985
- 24.- SEIER EDITH, (1987); "Análisis Exploratorio de Datos" ; Quinto Coloquio Sociedad Matematica Peruana; Trujillo-Perú; nov.1987
- 25.- SEIER EDITH, (1992); "Que es el Analisis Espectral y como se Aplica"; Laboratorio de Series de Tiempo Univ. Mayor de San Marcos; Lima
- 26.- TORRES S. D. (1979); " Ajuste de Modelos de Simulación de los Volúmenes de Descargas de Agua del río Chicama; tesis para optar el título de Ing. Estadístico. UNALM.
- 27.- URIEL EZEQUIEL, (1985); "Análisis de Series Temporales Modelos ARIMA" ; colección Abaco - Paraninfo; Madrid España;
- 28.- VELASQUEZ B.J.R. (1982); " Modelo Autorregresivo Armónico para Generación de Series Sintéticas Mensuales" Prov. de Huallaga; tesis para optar el título de Ing. Agrícola.
- 29.- WALONICK ASSOCIATES (1984-1985); "Software y manual de Forecast Plus" for tru IBM.
- 30.- YEVJEVICH VUJICA (1972); "Probability and Statistics in Hidrology, Water Resources Publications Fort Collins, Colorado USA.

A N E X O S

ANEXO A ESTACIONARIEDAD E INVERTIBILIDAD, AUTOCOVARIANCIA Y AUTOCORRELACION PARA ALGUNOS MODELOS

1) Sea el modelo AR(1) : $Z_t = 0,8Z_{t-1} + a_t$

$$\sigma^2_a = 2,0$$

Es estacionario puesto que $\phi_1 = |0,8| < 1$

Es invertible por definición, AR de orden finito

Cálculo de Covariancia

Se obtiene:

$$\gamma_0 = \frac{\sigma^2_a}{1 - \phi_1^2} = \frac{2}{1 - (0,8)^2} = 5,56$$

Tomando γ_0 como valor inicial, se tiene:

$$\begin{array}{lll} \gamma_1 = 4,45 & \gamma_2 = 2,84 & \gamma_3 = 1,82 \\ \gamma_4 = 3,56 & \gamma_5 = 2,28 & \gamma_6 = 1,45 \end{array}$$

Los coeficientes de autocorrelación se obtienen directamente de la forma recursiva:

$$\rho_k = \phi_1 \rho_{k-1}$$

Tomando $\rho_1 = 1$, como valor inicial:

$$\begin{array}{lll} \rho_1 = 0,80 & \rho_2 = 0,512 & \rho_3 = 0,328 \\ \rho_4 = 0,64 & \rho_5 = 0,410 & \rho_6 = 0,262 \end{array}$$

El cálculo de los coeficientes v_1, v_2, \dots, v_3 del modelo MA(∞), como puede transformarse en AR(1) para un idéntico retardo, coinciden con los ρ_k : $\rho_1=0,8$; $\rho_2=0,64$; ...

2) Para un modelo AR(2)

$$Z_t = 0,6Z_{t-1} + 0,3Z_{t-2} + a_t$$

$$\sigma^2_a = 3$$

Estacionariedad; las raíces de la ecuación característica deben ser menor que 1, en valor absoluto; ó si para el cálculo se utiliza el polinomio característico, las

raíces deben ser en valor absoluto mayor que 1, ó lo que es lo mismo deben estar situado fuera del círculo unitario:

A) Utilizando la ecuación característica:

$$G^2 - 0,6G - 0,3 = 0$$

cuyas raíces son: $(G_1, G_2) = 1/2(0,6 \pm \sqrt{0,6^2 + 4*0,3})$
 $G_1 = 0,92$; $G_2 = 0,32$; son raíces reales y distintas donde:

$|G_1| = |0,92| < 1$ y $|G_2| = |0,32| < 1$. entonces el proceso es estacionario.

B) Utilizando el polinomio característico:

$$1 - 0,6B - 0,3B^2 = 0$$

Obteniéndose las raíces $B_1 = 1,08$ y $B_2 = -3,08$; entonces:

$|B_1| > 1$ y $|B_2| > 1$, están fuera del círculo unitario, luego el proceso es estacionario y comprobándose que:

$$G_1 = 1/B_1 \quad \text{y} \quad G_2 = 1/B_2$$

C) En un proceso AR(2) se puede comprobar si es estacionario o no examinando si los coeficientes ϕ_1 y ϕ_2 cumplen las tres desigualdades:

$$\phi_1 + \phi_2 < 1 \quad 0,6 + 0,3 = 0,9 < 1$$

$$-\phi_1 + \phi_2 < 1 \quad -0,6 + 0,3 = -0,3 < 1$$

$$\phi_2 > -1 \quad 0,3 > -1.$$

El proceso AR(2) es invertible ya que es finita:

Cálculo de la Covariancia

para Y_0 se tiene: $Y_0 = \phi_1 Y_1 + \phi_2 Y_2 + \sigma^2_0$
 también se tiene:

$$Y_1 = \phi_1 Y_0 + \phi_2 Y_1$$

$$Y_2 = \phi_1 Y_1 + \phi_2 Y_0$$

$$Y_1 = \frac{\phi_1}{1 - \phi_2} Y_0 \quad ; \quad Y_2 = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2} Y_0$$

reemplazando valores se tiene:

$$\gamma_0 = 12,42$$

también se tiene:

$$\gamma_1 = \gamma_0[\phi_1/(1-\phi_2)] = 10,64$$

los demás valores se tiene de:

$$\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 = 0,6*10,64 + 0,3*12,42 = 10,11$$

y

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} ; k \geq 2$$

$$\gamma_3 = 9,26$$

$$\gamma_4 = 8,59$$

$$\gamma_5 = 7,93$$

Los coeficientes de correlación correspondientes son:

$\rho_k = \gamma_k/\gamma_0$; obteniéndose los siguientes resultados:

$$\rho_1 = 0,86 , \quad \rho_2 = 0,81 , \quad \rho_3 = 0,75 , \quad \rho_4 = 0,69 ,$$

$$\rho_5 = 0,64.$$

otro método es $\rho_1 = \phi_1/(1-\phi_2) = 0,86$ y los demás valores de: $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$.

3) Sea el modelo MA(2):

$$z_t = a_t - 0,4a_{t-1} + 1,2a_{t-2}$$

$$\sigma^2_a = 2$$

Estacionariedad

Por definición todo proceso MA de orden finito es estacionario.

Invertibilidad

A) Utilizando las raíces de la ecuación característica:

$$G^2 - 0,4G + 1,2 = 0$$

Resolviendo se tiene:

$$G_1 = 0,2 + i*1,08$$

$$G_2 = 0,2 - i*1,08$$

son raíces complejas conjugadas;

cuyos módulos es:

$$r = \sqrt{0,2^2 + 1,08^2} = 1,10$$

como $|r| = |1,10| > 1$, el proceso es no invertible.

B) Utilizando el polinomio característico:

$$1 - 0,4B + 1,2B^2 = 0$$

resolviendo se tiene: $B_1 = 0,166 + i0,90$;

$$B_2 = 0,166 - i0,90$$

cuyo módulo $r = \sqrt{0,166^2 + 0,90^2} = 0,91$
 Como $|r| < 1$; el modelo es no invertible.

C) Para el modelo MA(2) las condiciones de invertibilidad aplicadas directamente sobre los parámetros se tiene:

$$\theta_1 + \theta_2 < 1 \quad ; \quad 0,4 - 1,2 = -0,8 < 1 \quad \text{Se cumple}$$

$$-\theta_1 + \theta_2 < 1 \quad ; \quad -0,4 - 1,2 = -1,6 < 1 \quad \text{Se cumple}$$

$$\theta_2 > -1 \quad ; \quad -1,2 < -1 \quad \text{No se cumple}$$

luego el modelo es no invertible:

@ Aunque el proceso no sea invertible siguen siendo válidas las ecuaciones para el cálculo de la variancia y de las autocovariancias.

$$\gamma_0 = (1 + \theta^2_1 + \theta^2_2)\sigma^2_\epsilon = 5,2$$

$$\gamma_1 = (-\theta_1 + \theta_1\theta_2)\sigma^2_\epsilon = -1,76$$

$$\gamma_2 = -\theta_2\sigma^2_\epsilon = 2,4$$

$$\gamma_k = 0 \quad ; \quad k = 3, 4, 5, \dots$$

* Las autocorrelaciones se calculan a partir de la variancia y de las autocovariancias:

$$\rho_1 = -0,34$$

$$\rho_2 = 0,46$$

$$\rho_k = 0 \quad \text{para } k = 3, 4, 5, \dots$$

4) Para el Modelo ARMA(1,1):

$$Z_t = 0,9Z_{t-1} + a_t - 0,8a_{t-1}$$

$$\sigma^2_a = 5$$

- Como $|\phi_1| = 0,9 < 1$, el proceso es estacionario.
- análogamente como $|\theta_1| = 0,8 < 1$ el proceso es invertible.

* Cálculo de la variancia y autocovariancia:

$$\gamma_0 = \frac{1 - \theta_1\phi_1 + \theta^2_1}{1 - \phi^2_1} \sigma^2_a = 5,26$$

también se calcula para:

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma^2_a = 0,73$$

$$\gamma_2 = \phi_1\gamma_1 = 0,66$$

$$\gamma_3 = \phi_1\gamma_2 = 0,59$$

$$\gamma_4 = \phi_1\gamma_3 = 0,53$$

$$\gamma_5 = \phi_1\gamma_4 = 0,48$$

* así los coeficientes de autocorrelación se obtiene en forma inmediata:

$$\rho_1 = 0,14 , \rho_2 = 0,13 , \rho_3 = 0,11 , \rho_4 = 0,10 , \rho_5 = 0,09$$

ANEXO B ALGUNOS COMANDOS Y USO DEL FORECAST PLUS Y DEL TSP

En esta sección voy a describir en forma breve y resumida el contenido del Software Forecast Plus y algunas limitaciones de este y del T.S.P.

Forecast Plus

Consta de un menú principal conformado de cuatro componentes que se muestran en la pantalla como:

- 1) Data Management
- 2) Exploratory Package
- 3) Forecasting Analysis
- 4) Batch Mode
- 5) End the Program

Which ?

De los cuales podemos describir los contenidos principales de cada uno de estos componentes;

1) Data Management

Donde está incluido programas para crear, editar, transformar e imprimir files; es esta sección introducimos los datos históricos al computador, completar datos por omisión, editar datos en files o archivos, etc.

Editar un New File

antes de introducir la variable y los datos mensuales se crea un file label (el etiquetado) de la siguiente forma:

Number of variables:	1
Type of labeling:	YEAR
Starting label:	1974
Starting observation:	09
Number of observation between each label:	12
Increment of the labels:	1

Este etiquetado lleva una extensión *.lbl y para cualquier error o cuando se desea interrumpir el proceso usar la tecla <ESC>.

Transformación de Datos

En esta sección se elaboran sub-programas para crear nuevas variables, normalizar la serie, etc. con el uso de los siguientes comandos NEW, SELECT, COMPUTE, IF-THEN COMPUTE, NORMALIZE, LAG, DIFFERENCE, SUM, WRITE, Etc. con la ayuda de los comandos algebraicos y trigonométricos en utilizando la sintaxis idéntico al BASIC.

Cambio de parámetros en la tabla

Importante porque omite o activa algunas funciones del sistema DOS y de las operaciones internas del programa como el camino de los datos y la salida de los resultados (por archivos o por impresora), operaciones formato de cada archivo, etc.

2) Exploratory Package

En esta sección sirve para el análisis exploratorio de los datos cuyo sub-programas que cuenta son:

- 1) Time Plot
- 2) 4253HT Robust Smoothing
- 3) Box Plot (Trend/cicle variation)
- 4) Agregate Box Plot (Seasonal variation)
- 5) Spread vs Level plot
- 6) Autocorrelation Function
- 7) Return Main Menu

3) Forecasting Analysis

En esta sección se efectúa diferentes métodos para el suavisamiento de una serie temporal, métodos de descomposición de una serie y pronósticos mediante modelos de descomposición y de Box & Jenkins, análisis de regresión, etc. cuyos sub-programas que contiene son:

- 1) Simple Moving Average
- 2) Single Exponential Smoothing
- 3) Doble Exponential Smoothing
- 4) Holt's Two-Parameter Smoothing
- 5) Harrison's Harmonic Smoothing
- 6) Brown's Quadratic Exponential Smoothing
- 7) Winter's Seasonal Smoothing
- 8) Robust Descomposition
- 9) Census X-11 Descomposition
- 10) Regression Trend Analysis
- 11) Multiple Regression Analysis
- 12) Generalized Adaptive Filtering
- 13) Box Jenkins Analysis
- 14) Return Main Menu.

Como podemos notar que dentro de ello se encuentra el análisis de Box y Jenkins (13) para series temporales el cual a su vez consta de:

- 1) Identificación
- 2) Estimación / Pronóstico
- 3) Retorna al menú

En la fase de indentificación se indicará la transformación usada, el grado de diferencias ordinarias y/o estacionales a efectuar a la serie y la salida de las funciones de autocorrelación y autocorrelación parcial. Además en la fase de estimación se indicará los términos ordinarios y/o estacionales autorregresivos y de media móvil, número de Backforecast, número máximo de iteraciones para la estimación de los parámetros, convergencia-tolerancia, valor inicial de la estimación y la función de autocorrelación de los residuos.

4) Batch Mode

Donde se crea, ejecuta o borra un file batch también desde este componente puedes retornar al menú principal o salir del Forecast Plus.

Dentro de las limitaciones y alcances encontradas dentro del Software menciono:

- Los gráficos que se obtienen, no son apropiados para el análisis, es recomendable usar otro graficador.
- Cuando una serie llega al número máximo de iteraciones o no llega a la convergencia arroja el mensaje de ser no estacionaria o invertible y que es posible extenderse o efectuar una nueva corrida.
- Cuando no se usan transformaciones es posible que cuando se tengan datos muy bajos (cerca de cero) pudiera dar pronósticos menores que cero, ya que no existe la restricción $\forall Z_t > 0$.
- Si bien es cierto que las pruebas de Box-Pierce son buenos indicadores para los residuos pero pudiera añadirse otras pruebas para modelos con igual o aprox. Box-Pierce.
- La selección de los "lag" para los residuos es automático.

TSP

El micro TSP utilizado es de la Versión 7.03 adaptable a cualquier computador AT; y dentro de las limitaciones encontradas se tiene:

- No cuenta aparentemente con un menú principal y esta orientado principalmente a Econometría; es necesario seguir cuidadosamente una secuencia del trabajo que se efectúa.

- Como no esta establecido el orden de los comandos cualquier comando es aceptado en cualquier paso; las operaciones y cálculos se hacen en muestras o sub muestras por un comando Exp. y un Smpl.
- Cualquier error durante el proceso y la ejecución de un comando puede llevar a resultados viciados y es recomendable realizar nuevamente las operaciones.
- En la fase de estimación de los parámetros acepta solo parametros linealizados (no existe producto ni cuadrados de los parámetros) es necesario tener un modelo linealizado en variable y en parámetro.
- La matriz de correlación presenta como pares ordenados que dificulta el rápido análisis.
- Las pruebas estadísticas que presenta las diferentes pruebas de autocorrelación, regresión lineal, etc. son muchas y no están de acuerdo a las funciones que desempeña cada prueba o valor estadístico.
- La prueba de los residuos y el ploteo de ello es un proceso adicional.

**ANEXO C PREDICCIONES E INTERVALOS DE CONFIANZA (95%) PARA
SEPTIEMBRE 1988 A AGOSTO DE 1993**

MODELO : SARIMA(2,0,1)x(1,1,1)

MES\ORIGEN	168	156	144	PROMEDIO	INFERIOR	SUPERIOR
88\SET	1,17	0.96	3.12	1.75	0.49	6.22
OCT	2,63	1.95	7.13	3.90	1.10	13.80
NOV	4,68	3.42	12.90	7.03	1.98	24.99
DIC	15,42	10.94	36.01	20.79	5.05	73.93
89\ENE	37,03	26.61	80.57	48.07	13.52	170.95
FEB	64,91	47.00	198.00	103.30	29.05	367.37
MAR	43,27	31.12	163.42	79.27	22.29	201.91
ABR	20,00	14.55	83.99	39.51	11.11	140.52
MAY	4,11	3.02	16.81	7.98	2.24	28.30
JUN	2,13	1.56	7.46	3.72	1.05	13.22
JUL	1,30	0.95	3.92	2.06	0.50	7.31
AGO	1,02	0.73	3.35	1.70	0.48	6.05
SET	1,06	0.73	3.23	1.67	0.47	5.95
OCT	2,23	1.53	7.33	3.70	1.04	13.15
NOV	4,02	2.70	13.28	6.67	1.87	23.77
DIC	11,11	8.33	37.11	18.85	5.30	67.04

MES\ORIGEN	168	156	144	PROMEDIO	INFERIOR	SUPERIOR
90\ENE	30,57	20,12	84,51	45,07	12,67	160,27
FEB	62,40	38,94	205,92	102,42	28,80	364,23
MAR	37,32	26,94	166,10	76,79	21,59	273,00
ABR	21,43	13,00	85,54	39,99	11,24	142,22
MAY	4,88	2,70	17,31	8,30	2,33	29,51
JUN	2,23	1,32	7,65	3,73	1,05	13,28
JUL	1,24	0,76	4,00	2,00	0,56	7,11
AGO	0,80	0,61	3,41	1,61	0,45	5,71
SET	0,90	0,60	3,28	1,59	0,45	5,67
OCT	1,92	1,24	7,44	3,53	0,99	12,57
NOV	3,43	2,18	13,44	6,35	1,79	22,58
DIC	9,93	6,77	37,37	18,02	5,07	64,10
91\ENE	26,24	16,33	84,44	42,34	11,90	150,56
FEB	51,24	30,73	205,61	95,86	26,96	340,90
MAR	31,46	20,95	166,37	72,93	20,51	259,35
ABR	16,97	9,90	85,37	37,41	10,52	133,05
MAY	3,75	2,07	17,18	7,67	2,16	27,26
JUN	1,77	1,02	7,58	3,46	0,97	12,29
JUL	1,00	0,60	4,00	1,07	0,52	6,64
AGO	0,69	0,47	3,36	1,51	0,42	5,36
SET	0,74	0,46	3,23	1,48	0,42	5,25
OCT	1,58	0,96	7,30	3,28	0,92	11,66
NOV	2,83	1,69	13,15	5,89	1,66	20,95
DIC	8,06	5,21	36,51	16,59	4,67	59,01
92\ENE	21,47	12,53	82,42	30,01	10,91	130,01
FEB	42,30	23,69	200,11	88,70	24,94	315,44
MAR	25,71	16,16	166,30	69,39	19,51	246,77
ABR	14,06	7,70	82,60	34,79	9,78	123,71
MAY	3,12	1,59	16,60	7,10	2,00	25,26
JUN	1,46	0,78	7,30	3,18	0,89	11,31
JUL	0,82	0,46	3,80	1,69	0,48	6,02
AGO	0,55	0,36	3,23	1,38	0,39	4,91
SET	0,60	0,35	3,09	1,35	0,38	4,79
OCT	1,30	0,73	6,96	3,00	0,84	10,66
NOV	2,27	1,27	12,52	5,35	1,51	19,04
DIC	6,47	3,92	34,67	15,02	4,22	53,42
93\ENE	17,14	9,41	70,05	34,07	9,80	124,00
FEB	33,59	17,70	189,06	80,12	22,53	204,92
MAR	20,41	12,04	156,08	62,04	17,67	223,49
ABR	11,00	5,72	77,68	31,49	8,06	112,00
MAY	2,45	1,18	15,57	6,40	1,80	22,76
JUN	1,14	0,60	6,83	2,06	0,80	10,16
JUL	0,64	0,34	3,55	1,51	0,42	5,37
AGO	0,43	0,27	3,00	1,23	0,35	4,39