

00361
20
2011



**UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO**

FACULTAD DE CIENCIAS

DIVISION DE ESTUDIOS DE POSGRADO

**"APLICACIONES DEL ANALISIS ESTADISTICO DE
SERIES DE TIEMPO UNIVARIADAS A PROCESOS
EVOLUTIVOS ESTOCASTICOS".**

TESIS PROFESIONAL

**QUE PARA OBTENER EL GRADO DE
MAESTRIA EN CIENCIAS
(BIOLOGIA)**

**P R E S E N T A
MARIANA NUÑEZ ZURIGA**

DIRECTORA DE TESIS: DRA. ANA BARAHONA ECHEVERRIA

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

**Dedico este trabajo
a mi Manuel y
a mi Sol y a mi Mar,
nuestras hijitas
Ana Sofi y Marianita.**

**...porque nuestro tiempo existencial
tiene una sola dirección y no regresa jamás,
lo que haces HOY amor,
queda como tu testimonio,
huella e historia de tu vida.**

**Las obras (y los recuerdos)
quedan,
las personas se van.**

**a mis padres Rafael y Conchita,
y a Rafael y Darío,
huellas imperecederas en mi corazón.**

APLICACIONES DEL ANALISIS ESTADISTICO DE SERIES DE TIEMPO UNIVARIADAS A PROCESOS EVOLUTIVOS ESTOCASTICOS.

Resumen de la tesis que para obtener el grado de Maestría en Ciencias presenta:
Biol. Mariana Núñez Zúñiga.

Se investiga la utilidad de la aplicación del Análisis estadístico de series de tiempo univariadas a procesos evolutivos estocásticos, para formalizar matemática y estadísticamente la polémica sobre el modo como ha ocurrido el cambio evolutivo. Se realiza el análisis estadístico a dos series poblacionales y a una serie de tiempo del registro fósil de radiolarios, para buscar modelos matemáticos que determinen si el cambio evolutivo ocurre de manera paulatina y gradual, o si por el contrario, los cambios observados en el registro fósil corresponden a una evolución saltatoria y discontinua como propone la teoría de los equilibrios puntuados.

Se propone la necesidad de utilizar esta herramienta estadística en los protocolos de las investigaciones sobre Biología Evolutiva, donde se tengan datos numéricos continuos, dado que el Análisis estadístico de series de tiempo es la herramienta estadística apropiada para analizar y modelar matemáticamente conjuntos de datos en el tiempo.

PRESENTACION.

La presente tesis está motivada por la aplicación de un nuevo prisma matemático y estadístico para entender el fenómeno evolutivo: la aplicación del análisis estadístico de series de tiempo univariadas a datos continuos del registro fósil, para abordar con formalismo matemático y estadístico la polémica en torno al modo como ocurrió el cambio evolutivo, dado que a la fecha de elaboración del presente trabajo, no existen investigaciones de esta índole.

Cuando se dispone de información en forma de sucesiones cronológicas numéricas a intervalos de tiempo fijos, es posible llevar a cabo un análisis estadístico de series de tiempo univariadas. La metodología de este análisis estadístico es aplicable a series de tiempo de cualquier área del conocimiento

En el presente trabajo se ejemplifican las técnicas del análisis estadístico de series de tiempo univariadas, con aplicaciones a series de tiempo de procesos estocásticos ecológicos y evolutivos. Se construyen modelos estadístico-matemáticos a partir de datos de series de tiempo observadas y se estudia su utilidad para explicar el fenómeno evolutivo.

Concretamente, se investiga si esta técnica estadística permite distinguir si el proceso de cambio evolutivo ocurre de manera gradual o de manera brusca y saltatoria. De manera mas ambiciosa se investiga si mediante esta técnica estadística es posible discernir los componentes estocásticos de los componentes deterministas en una serie de tiempo de un proceso estocástico evolutivo, y aún más, si este análisis estadístico puede explicar si el cambio se debe al azar o a selección natural y si este cambio sigue de alguna forma al ambiente.

Esta tesis es una muestra del tipo de investigaciones sobre procesos estocásticos en el área de Biología Evolutiva que se pueden realizar con esta herramienta estadística.

AGRADECIMIENTOS.

Muchas personas han contribuido conmigo, en una u otra forma en la elaboración del presente trabajo. En primer lugar quiero agradecer a mi directora de tesis, la Dra. Ana Barahona, del laboratorio de Historia de la Biología y Evolución de la Facultad de Ciencias de la UNAM, el gran interés y entusiasmo con que apoyó este trabajo desde que se lo comenté cuando aún era apenas una idea. Guardo en mis tesoros lo que de ella aprendí; no es sólo su profundo compromiso como Directora de esta tesis, es su extraordinaria maestría y profesionalismo con que solucionó un sin número de obstáculos con ese agradable, elegante e increíblemente eficiente estilo que la caracteriza ¡Muchas gracias por todo Ana!

La Dra. María Eugenia Gonsebatt Bonaparte, mi excepcional maestra y amiga, apoyó el desarrollo de esta tesis literalmente "beyond the calls of duty". Por tu ejemplo y tu valiosa amistad, muchas gracias Maru.

Quiero agradecer a la Dra. Cathy Garcia Reimbert del IIMAS, todas las horas y su desinteresado apoyo para obtener bibliografía y entender los conceptos matemáticos que se utilizan en esta tesis. Sus amistosas y pacientes explicaciones cuando revisamos juntas las ecuaciones, literalmente me revelaron el significado de ecuaciones en diferencias, raíces características y demás conceptos matemáticos, que de otra manera permanecerían oscuros o poco significativos para mí.

También quiero agradecer a la Dra. Marisol Montellano, del Instituto de Geología de la UNAM las sugerencias y facilidades bibliográficas que amablemente me brindó para la obtención de los datos que se analizan en la tercera serie de este trabajo. Ellas y mis sinodales el Dr. Enrique Buzo y el Dr. Fedro Guillén, revisaron cuidadosamente los gruesos borradores. Sus sugerencias y correcciones son un importante respaldo profesional al presente trabajo y lo hicieron más congruente e interesante. Sin embargo, si pese a todo quedan obscuridades o incongruencias en el trabajo, son responsabilidad mía.

El Dr. Victor Guerrero, especialista en Series de Tiempo del Instituto Tecnológico Autónomo de México, y autor del libro Análisis estadístico de series de tiempo económicas, amablemente revisó este trabajo y me facilitó los artículos que él ha publicado sobre pronósticos en revistas internacionales especializadas. Sus críticas y sugerencias le dieron un nivel académico mucho mayor a esta tesis, por lo cual tengo un agradecimiento muy especial para el Dr. Guerrero.

El Doctor Rodolfo Vázquez profesor titular del Instituto Tecnológico Autónomo de México, me ayudó a hacer realidad un antiguo sueño al apoyarme con la beca para estudiar el Diplomado en Estadística Aplicada, pues el horario había resultado una barrera insalvable para realizarlo en el IIMAS.

Mi extraordinario maestro de estadística no paramétrica, el Dr. Mario Cortina Borja, inmediatamente leyó el trabajo cuando aún era un primer borrador y me hizo comentarios y sugerencias sumamente pertinentes para realizar un mejor ajuste de los modelos a los datos y sobre las Pruebas para evaluar la Bondad del Ajuste. Los comentarios del Dr. Juan Núñez Farfán sobre la importancia y alcances de este trabajo para los biólogos resultaron un empujón exponencial ($n > 4$), para la culminación de esta tesis.

Tía Natasha y Babu Conchita merecen un reconocimiento muy particular por todo lo que sabemos. Sin su cariñosa cooperación, llegar a las Series de Tiempo no hubiera resultado tan factible, tranquilizador y agradable para mi.

Y a ti Manuel, quiero expresarte mi amor por compartir y respaldar este punto de vista de crear y desarrollar tantos planes y proyectos que van construyendo la historia nuestra vida. Son nuestro testimonio existencial que formamos y disfrutamos cada día y serán algún día recuerdo y refugio, quizás el cimiento para el crecimiento de nuestras queridas hijas.

ÍNDICE

	pp.
- Introducción	1
- Hipótesis de trabajo	8
- Objetivo general	8
- Objetivos específicos	9
<i>Antecedentes Biológicos:</i>	
- Capítulo 1: Hacia un modelo estocástico-determinista de la Evolución. ..	10
1.1. Evolución: Los marcos y construcciones del tiempo.	10
1.2. Procesos estocásticos lineales y series de tiempo.	14
1.3. Procesos estocásticos biológicos.	16
1.4. Procesos estocásticos en Evolución: El fenómeno evolutivo como un proceso estocástico-determinista: Las fuerzas del cambio:	19
Las fuerzas de la evolución biológica:	19
1.4.1. Componentes deterministas de la evolución biológica: La Selección Natural y las fuerzas conservadoras o de mantenimiento: La "memoria" del proceso evolutivo.	20
1.4.2. Componentes estocásticos del fenómeno evolutivo: las fuerzas generadoras del cambio: Mutaciones, recombinación genética, deriva génica y migraciones.	23
1.5. Modelos actuales sobre la Evolución biológica.	26
1.5.1. Gradualismo filético: bajo la sombra de Darwin.	26
1.5.2. La Teoría de los Equilibrios Punteados: otro modo de interpretar el registro fósil.	28
1.5.3. El paradigma actual sobre la evolución de las especies.	30

Antecedentes estadísticos y matemáticos:

- Capítulo 2: Análisis estadístico de series de tiempo: Conceptos introductorios y antecedentes.

2.1. Definición de series de tiempo.	32
2.2. Objetivos de un análisis de series de tiempo.	32
2.3. Análisis estadístico de una serie de tiempo.	33
2.4. Ecuaciones en diferencia.	34
2.5. Los modelos ARIMA de series de tiempo.	44

El método del Análisis estadístico de series de tiempo:

- Capítulo 3: La Metodología estadística del Análisis de series de tiempo: Métodos de análisis de los datos originales de la serie de tiempo: la Etapa: enfoque descriptivo.

3.1. Análisis gráfico de una serie de tiempo.	46
3.2. Análisis estadístico de una serie.	50
3.3. Análisis de la estructura de autocorrelación de una serie.	54

- Capítulo 4: La Metodología estadística del Análisis de series de tiempo:	
2a. Etapa: Metodología para la construcción del modelo ARIMA de la serie de tiempo.	55
Estrategia para la construcción de modelos ARIMA: cuadro 1	56
4.1. Identificación del modelo ARIMA.	59
Tipos de modelos ARIMA :	
4.1.1.- Modelos autorregresivos (AR).	59
4.1.2.- Modelos de promedios móviles (MA).	63
4.1.3.- Modelos autorregresivos de promedios móviles (ARMA).	68
4.1.4.- Modelos autorregresivos integrados y de promedios móviles (ARIMA).	71
Cuadro 2: identificación del modelo.	72
Cuadro 3: comportamiento de la ACF y de la PACF para una serie estacionaria.	73
4.2. Estimación de los parámetros del modelo:	74
Tablas de resultados.	
4.3. Verificación de supuestos:	79
4.3.1. Análisis de residuales: Tablas de resultados.	79
4.3.2. Otras pruebas de la Bondad del Ajuste:	88
4.3.3. Tablas de resultados.	90
4.4. Uso del modelo: explicación del fenómeno en estudio, simulación, control y/o pronósticos.	93
 <i>...siguen Antecedentes estadísticos:</i>	
- Capítulo 5: Aplicación del modelo: pronósticos.	94
5.1. Pronósticos con Modelos ARIMA: Generalidades.	94
5.1.1. Tres formas de expresar al modelo ARIMA.	94
5.1.2. Proceso para la obtención del pronóstico.	97
5.2. Pronósticos ARIMA en condiciones óptimas:	100
Series estacionarias.	
5.3. Pronósticos ARIMA en casos especiales:	102
5.3.1. Series con tendencia.	
5.3.2. Series estacionales.	
5.4. Pronósticos ARIMA restringidos por cambios en los parámetros del modelo debido a factores exógenos: Las intervenciones.	104
5.5. Análisis de intervención.	107
5.6. Pronósticos ARIMA bajo diferentes condiciones	109

Las series:

**- Capítulo 6: primera serie.-Aplicación del método de Análisis estadístico de series de tiempo a un ejemplo clásico en Biología:
El clásico ejemplo de series de tiempo biológicas: Los linces canadienses.**

Aplicación del análisis estadístico de las series de tiempo a los datos originales de la compañía exportadora de pieles "Hudson Bay Company" desde 1821 hasta 1913.	111
6.1. Gráfica de los cambios en la abundancia anual.	112
6.2. Tabla de datos.	113
6.3. Análisis de los datos y construcción del modelo ARIMA mediante el uso del paquete PEST: Metodología. la etapa: Análisis descriptivo:	116
6.3.1. Gráfica de los datos originales.	116
6.3.2. Histograma y estadísticos.	117
6.3.3. Autocorrelograma y autocorrelograma parcial de los datos originales.	118
6.4. Metodología para la construcción del modelo ARIMA de la serie de tiempo:	
Ila. etapa: Construcción del modelo:	119
6.4.1. Identificación del modelo:	
a) Determinar una serie estacionaria en función de los datos originales:	119
Estabilización de la varianza, comparando diferencias.	
Determinar el valor de "d".	119
b) Sustracción de la media.	122
c) Análisis de estacionalidad.	124
d) Obtener una representación ARMA para la serie estacionaria.	127
e) Fijar los valores de "p" y "q": ARIMA (2,0,0).	128
6.4.2. Estimación de los parámetros del modelo: Tablas de resultados.	128
6.4.3. Verificación de supuestos:	129
a) Análisis de residuales: Gráficas.	129
b) Pruebas de Bondad del ajuste: Tablas de resultados.	132
6.4.4. Usos del modelo: pronósticos.	133
6.4.5. Otro modelo: ARIMA (2,0,19)	136
6.5. Tabla comparativa de los modelos estimados para la serie de linces.	149
6.6. Discusión de resultados.	150
6.7. Conclusiones.	151

- Capítulo 7: segunda serie.-

Fase Experimental 1: Comprobación de hipótesis.

Aplicación del análisis estadístico de series de tiempo univariadas a un conjunto de datos de un proceso estocástico ecológico: serie estacionaria: El caso de las garzas inglesas y galesas "<u>Ardea cinerea</u>".	152
7.1. Gráfica de los cambios en la abundancia anual.	152
7.2. Tabla de datos.	153
7.3. Análisis de los datos y construcción del modelo ARIMA mediante el uso del paquete PEST: Metodología. la etapa: Análisis descriptivo.	154
7.3.1. Gráfica de los datos originales.	154
7.3.2. Histograma y estadísticos.	155
7.3.3. Autocorrelograma y autocorrelograma parcial de los datos originales.	156
7.4. Metodología para la construcción del modelo ARIMA de la serie de tiempo:IIa.etapa: Construcción del modelo:	157
7.4.1. Identificación del modelo:	
a) Determinar una serie estacionaria en función de los datos originales: Estabilización de la varianza, comparando diferencias. Determinar el valor de "d".	157
b) Sustraer la media	158
c) Obtener una representación ARMA para la serie estacionaria.	160
d) Fijar los valores de "p" y "q".	160
7.4.2. Estimación de los parámetros del modelo:ARIMA (0,1,1) Tablas de resultados.	161
7.4.3. Verificación de supuestos:	
a) Análisis de residuales: Gráficas.	163
b) Pruebas de Bondad del ajuste: Tablas de resultados.	163
7.4.4. Usos del modelo: pronósticos.	165
7.4.5. Resultados obtenidos con otros modelos.	168
7.5. Tabla comparativa de los modelos estimados	169
7.6. Discusión de resultados.	169
7.7. Conclusiones.	

- Capítulo 8: tercera serie.-

Fase experimental 2: Comprobación de hipótesis.

Aplicación del análisis estadístico de series de tiempo univariadas a un conjunto de datos de un proceso estocástico evolutivo: serie con tendencia: Los radiolarios <u>Pseudocubus vema</u>	170
8.1. Gráfica del cambio filogenético de los radiolarios <u>Pseudocubus vema</u> observado en su registro fósil.	172
8.2. Tabla de datos.	173
8.3. Análisis de los datos y construcción del modelo ARIMA mediante el uso del paquete PEST: Metodología. la etapa: Análisis descriptivo:	
8.3.1. Gráfica de los datos originales.	174
8.3.2. Histograma y estadísticos.	175
8.3.3. Autocorrelograma y autocorrelograma parcial de los datos originales.	176
8.4. Metodología para la construcción del modelo ARIMA de la serie de tiempo: la etapa: Construcción del modelo:	177
8.4.1. Identificación del modelo:	
a) Determinar una serie estacionaria en función de los datos originales: Estabilización de la varianza, comparando diferencias. Determinar el valor de "d".	177
b) Sustraer la media.	
c) Obtener una representación ARMA para la serie estacionaria.	182
d) Fijar los valores de "p" y "q".	185
8.4.2. Estimación de los parámetros del modelo ARIMA (1,1,1)	186
8.4.3. Verificación de supuestos.	
a) Análisis de residuales: Gráficas	187
b) Pruebas de Bondad del ajuste: Tablas de resultados.	190
8.4.4. Usos del modelo: pronósticos.	191
8.4.5. Resultados con otros modelos ARIMA (1,1,0)	193
8.5. Tabla comparativa de los modelos estimados para la serie de los radiolarios.	202
8.6. Análisis de intervenciones: Comparación de la serie de datos originales con una serie "patrón" sin intervenciones.	203
Análisis de la serie patrón invertida.	214
8.7. Discusión de resultados.	224
8.8. Conclusiones.	226

- Capítulo 9: Análisis de los resultados generales del trabajo:	229
- Capítulo 10: Conclusiones generales.	234
- Bibliografía.	235
- Anexo a: Comportamientos típicos de la FAC y de la FACP.	238
- Anexo b: Método de Máxima verosimilitud	244

INTRODUCCION.

En la historia de la ciencia hay muchos ejemplos en los que la aplicación de herramientas nuevas y más sofisticadas al realizar la investigación científica de distintos fenómenos, ha descubierto nuevos niveles de complejidad (Rostand, 1942). En ocasiones no se trata tan solo de una nueva dimensión que plantea nuevas preguntas sobre un área del conocimiento científico, sino de todo un replanteamiento que revoluciona los modelos de problemas y las soluciones que una comunidad científica reconocía como los cánones aceptados del pensamiento científico de esa época (Kuhn, 1970).

Las realizaciones científicas universalmente reconocidas que, durante cierto tiempo, proporcionan modelos de problemas y sus soluciones a una comunidad científica, han sido denominados "paradigmas" (Kuhn, 1970).

Una de las polémicas contemporáneas medulares en Evolución es el modo como han surgido nuevas especies. Diversos autores señalan que el paradigma actual en la teoría de la evolución de las especies es el modo como se ha dado el proceso evolutivo (Gould, Eldredge, 1977; Eldredge, 1985; Olea, 1986).

El enfoque gradualista clásico postula que la evolución ha sido un proceso gradual y continuo, que se da desde las poblaciones hasta la macroevolución. La visión gradualista de la naturaleza está siendo reemplazada por una concepción saltatoria o discontinua y se ha propuesto un enfoque alternativo: La Teoría de los Equilibrios Puntuados.

La teoría de los equilibrios puntuados fué, propuesta a principios de los setentas por Gould, Eldredge y Stanley, para explicar la gran cantidad de patrones evolutivos que observaron en el registro fósil: millones de años sin cambio, y repentinamente cambios rápidos, asociados a la especiación. Este nuevo enfoque, propone una visión jerárquica de la evolución, considerando que la visión gradualista tiene un enfoque reduccionista que quiere extrapolar el comportamiento de la evolución de las poblaciones, hacia el

surgimiento de nuevas especies y la macroevolución (Martínez, 1986; Soberón, 1986).

Una manera de abordar la polémica sobre el modo como ocurrió el cambio evolutivo, es tratar de plantear y formalizar el proceso evolutivo en términos matemáticos y estadísticos. Entender a la evolución como procesos estocásticos y estudiar en términos estadísticos y matemáticos a los procesos evolutivos estocásticos-deterministas.

Ahora bien, ya que nunca es posible abarcar la realidad en su totalidad, el conocimiento científico selecciona un fenómeno o una serie de fenómenos, que siempre son abstracciones de un orden elemental (Rosenblueth, 1971). El conocimiento científico busca regularidades en la naturaleza, para establecer relaciones, entre objetos o fenómenos que son invariantes en el tiempo y en el espacio. La ciencia busca modelos abstractos o teóricos lógicos, que representen fielmente las relaciones funcionales invariantes que existen en la naturaleza. Dado que la probabilidad de que un fenómeno dado se reproduzca en su totalidad es nula, la búsqueda de uniformidades y reproducibilidades implica una abstracción. En una situación compleja dada, se seleccionan algunos aspectos, algunas variables como importantes o pertinentes y se hace caso omiso de otras variables que se consideran incidentales o contingentes, desde el punto de vista de la relación particular que se explora (Rosenblueth, 1971).

El lenguaje apropiado para establecer las relaciones funcionales entre las variables de interés, es el simbolismo de las matemáticas. Las teorías científicas que pueden formularse a través de ecuaciones diferenciales son más importantes y precisas. Estos modelos parten de experiencias pasadas y deben ser aplicables a situaciones similares futuras. La ciencia empieza y acaba siempre en la naturaleza, en los hechos, en una realidad externa que se postula (Rosenblueth, 1971).

La presente tesis está animada por la aplicación de una herramienta estadística concreta, el análisis estadístico de las series de tiempo univariadas, al estudio del proceso evolutivo, con la

finalidad de encontrar nuevos argumentos estadísticos y matemáticos sobre el modo en que han evolucionado las especies.

El análisis estadístico de las series de tiempo univariadas es una metodología matemática que se aplica a información recopilada sobre un fenómeno particular en forma de registros metódicos u observaciones numéricas efectuadas a intervalos de tiempo fijos, con fines de análisis; es decir, para explicar el fenómeno en estudio y posteriormente, hacer planeación y toma de decisiones (Guerrero, 1991a).

Esta herramienta estadístico-matemática es muy utilizada para el análisis de datos en investigaciones de muy diversas disciplinas científicas donde se estudian procesos estocásticos. El análisis de series de tiempo se aplica para investigar procesos estocásticos en muy diversas disciplinas científicas, desde investigaciones sismológicas, oceanográficas, biológicas, físicas, químicas, económicas, etc. (Wei, 1990).

Sobre las aplicaciones del análisis estadístico de series de tiempo en estudios sobre la variación de la abundancia de poblaciones biológicas, hay trabajos considerados como "clásicos" con los que se ejemplifica esta técnica estadística. Tal es el caso de los trabajos sobre las poblaciones de lince canadienses. (Elton y Nicholson, 1942, Schaffer, 1984). Sin embargo, el análisis estadístico de series de tiempo univariadas no ha sido aplicado para investigar el fenómeno de la evolución biológica, entendiendo a la evolución de las especies, como procesos estocásticos deterministas biológicos evolutivos, pese a que en esta área se tienen conjuntos de datos numéricos o sucesiones cronológicas de poblaciones del registro fósil.

El objetivo inicial de esta tesis es destacar que el análisis estadístico de las series de tiempo univariadas es una herramienta estadística aplicable a procesos evolutivos biológicos estocásticos, en los que se tienen registros numéricos metódicos de una variable de interés a intervalos de tiempo fijos. A la fecha de presentación de esta tesis, no hay trabajos publicados sobre este tema.

El resultado del análisis es un modelo matemático "ARIMA" que resume la información de la serie y explica el comportamiento de la variable de interés en relación al tiempo (Guerrero, 1991a).

La importancia de realizar este tipo de análisis estadístico en estudios de procesos evolutivos estocásticos, radica en que al aplicar el análisis estadístico de las series de tiempo univariadas a los datos, no solo se visualiza mejor el comportamiento presente de la variable en estudio mediante el modelo matemático obtenido, sino que además de explicar la situación actual del fenómeno, el modelo ARIMA permite hacer simulaciones y pronósticos sobre el comportamiento futuro de la variable en estudio y hacer inferencias sobre su historia, es decir, sobre los valores probables que tuvo la variable de interés en los tiempos pasados.

Ahora bien, como ya se mencionó, actualmente las interrogantes centrales en la teoría de la evolución son: ¿cómo ha sido el cambio?; ha sido un cambio gradual o el cambio se dió por "saltos" de un estado de equilibrio inicial a un nuevo estado de equilibrio?, y sobre todo, ¿a qué se debe el cambio? ¿a la selección natural o al azar?, ¿este cambio sigue de alguna forma al ambiente?.

Dado que los modelos matemáticos ARIMA obtenidos mediante la aplicación del análisis estadístico de las series de tiempo univariadas a conjuntos de datos numéricos del registro fósil explican el comportamiento de la variable de interés, las preguntas más ambiciosas que se plantean en esta tesis son:

1. ¿Es posible argumentar matemáticamente a favor del cambio gradual o por el contrario, el modelo matemático ARIMA indica que la variable tiene un comportamiento "por saltos", es decir a favor de la teoría de los equilibrios puntuados?.
2. ¿Es posible mediante dicho análisis estadístico "separar" o entender los componentes azarosos o estocásticos, de los componentes deterministas en la serie de datos considerada?.
3. ¿Es posible entender al proceso evolutivo mediante un modelo estocástico-determinista que considere como fuerzas

antagónicas a las fuerzas generadoras del cambio y a las fuerzas de conservación o mantenimiento?

4. ¿Este análisis estadístico permite distinguir si el cambio se debió a selección natural o a deriva génica? ¿cómo?

En suma, se trata de investigar la "finura" y posibilidades de esta técnica estadística para estudios sobre evolución biológica.

Para responder a estas preguntas medulares, se realizó una investigación siguiendo las etapas del método científico: 1. Observación y formulación de hipótesis de trabajo, en la que se estudian los antecedentes y conceptos fundamentales tanto de los procesos evolutivos estocásticos, como del método estadístico, es decir del análisis estadístico de las series de tiempo univariadas. 2. Experimentación, donde se aplica la metodología estudiada a dos series de datos de procesos estocásticos biológicos evolutivos. 3. Discusión de resultados y corroboración de las hipótesis de trabajo. 4. Conclusiones: respuesta a las proposiciones centrales de esta tesis.

Para poder establecer las hipótesis de trabajo, la investigación de los antecedentes se dividió en tres áreas. La primer área, los antecedentes biológicos, se presenta en el capítulo 1, y se refiere a la Evolución de las especies. En este capítulo se busca realizar un análisis de las fuerzas involucradas en el proceso evolutivo, con la idea de proponer un modelo estocástico-determinista de dicho proceso. Se presentan también las posiciones, modelos o escuelas actuales que se ocupan del fenómeno evolutivo.

La segunda área de los antecedentes del trabajo se refiere a los antecedentes propiamente estadísticos y matemáticos, y se presentan en los capítulos 2, 3, 4 y 5. Por tanto, en el capítulo 2, se presentan los antecedentes y conceptos generales del Análisis estadístico de las series de tiempo univariadas. En los capítulos 3 y 4 se estudia la metodología del propio método, es decir, los pasos para construir el modelo ARIMA. La aplicación y los usos del modelo ARIMA se estudian en el capítulo 5: "Aplicación del modelo: pronósticos", donde se presentan tres situaciones bajo las cuales es posible obtener pronósticos, en primer lugar, los pronósticos para

series de tiempo estacionarias, después los pronósticos bajo situaciones especiales, para series de tiempo con tendencia, y para series de tiempo cíclicas o estacionales, y finalmente, se estudian los pronósticos que explican cambios en la estructura determinística del proceso, es decir, cambios en los parámetros del modelo ARIMA debidos a eventos exógenos: las intervenciones. Este capítulo es particularmente importante para entender los alcances del método estadístico. Aquí se estudia de manera especial el caso de las intervenciones, debido a la importancia de estos eventos en el estudio de procesos estocásticos evolutivos.

En la tercer área para establecer los antecedentes de trabajo, se ejemplifica la aplicación del análisis estadístico de series de tiempo univariadas, para lo cual, en el capítulo 6 se presenta el ejemplo clásico de series de tiempo biológicas: los registros de la abundancia anual de linces canadienses realizados por la "Hudson's Bay Company" desde 1821 hasta 1913. Un proceso estocástico ecológico sumamente estudiado que permite investigar los alcances de esta técnica y analizar los componentes estocásticos y deterministas de dicho proceso.

Una vez establecidos los antecedentes y las hipótesis de trabajo, la segunda etapa de esta tesis, la fase experimental se presenta en los capítulos 7 y 8, donde se aplica la metodología estadística de las series de tiempo univariadas a dos conjuntos de datos de procesos estocásticos biológicos, para comprobar experimentalmente las hipótesis de trabajo. Así, en el capítulo 7, se presenta el primer conjunto de datos: una serie estacionaria y en el capítulo 8 se presenta el segundo grupo de datos, que es una serie con tendencia.

La serie estacionaria del capítulo 7 es un registro de los cambios anuales de la abundancia de garzas "Ardea cinerea" en Inglaterra y Gales, durante más de 40 años. Este trabajo fué realizado por S.Stafford en 1971 y reeditado por Begon, Harper y Townsend en 1986. Esta serie de datos si bien ha sido muy estudiada por distintos autores, ilustra dos conceptos muy importantes en series de tiempo, la estacionariedad y las intervenciones.

La serie con tendencia del capítulo 8 es un proceso estocástico evolutivo. En este caso, se eligió un extraordinario estudio paleontológico donde se exponen las variaciones en la morfología de los radiolarios fósiles, durante un período de más de 2 millones de años. Se trata de un conjunto de mediciones torácicas realizadas en muestras poblacionales de los radiolarios Pseudocobus vema, obtenidas en el registro fósil de la Antártida, por la Dra. D.E. Kellogg, publicadas en 1975. En 1977 los paleontólogos Gould y Eldredge publican un interesante artículo donde discuten las interpretaciones que la Dra. Kellogg dió a sus datos, y proponen un nuevo paradigma dentro de la Teoría de la Evolución: La teoría de los Equilibrios Puntuados. Cabe mencionar que en Paleontología es muy difícil obtener registros fósiles tan detallados con datos continuos, debido tanto a diferentes factores técnicos y económicos, como a imperfecciones del propio registro fósil.

Al final de cada serie analizada, es decir al final de los capítulos 6, 7 y 8 respectivamente, se desarrolla la tercera etapa de esta investigación, que consiste en la discusión de los resultados. La discusión sobre la hipótesis de trabajo y la evaluación de los objetivos de la investigación, se presentan en el capítulo 9.

Finalmente, las conclusiones de esta investigación, se presentan en el capítulo 10, donde se responde a las preguntas centrales por las que se inició este trabajo, mediante las proposiciones fundamentales que resultaron del estudio. Se discute la utilidad de esta herramienta estadística en el estudio de procesos estocásticos evolutivos biológicos, tanto sus alcances, como sus limitaciones.

HIPOTESIS DE TRABAJO

Los modelos ARIMA obtenidos mediante el análisis estadístico de series de tiempo aplicado a conjuntos de datos numéricos de series de tiempo evolutivas, explican matemáticamente cómo ocurrió el proceso de evolución de las especies; si hubo una tendencia de cambio gradual, o por el contrario si el cambio ocurrió "por un salto" debido a una intervención que alteró la secuencia histórica de la serie.

OBJETIVO GENERAL.

El objetivo de la presente tesis es conocer los alcances y limitaciones del análisis estadístico de series de tiempo univariadas en el estudio de procesos evolutivos estocásticos.

Concretamente, el objetivo de esta tesis es aplicar esta metodología estadística a una serie de datos numéricos del registro fósil de los radiolarios Pseudocobus vema, con el fin de modelar matemáticamente el comportamiento de dicha serie. Al modelar esta muestra del fenómeno evolutivo, se buscan argumentos matemáticos y estadísticos sobre el modo como ocurrió el proceso evolutivo, si se dió de manera gradual, o como un proceso de "equilibrios puntuados".

OBJETIVOS ESPECIFICOS

1. Dar un argumento matemático y estadístico mediante el modelo ARIMA obtenido a favor de un cambio gradual, si el análisis indica que hay una tendencia en la serie de datos evolutivos analizada.

2. Dar un argumento matemático y estadístico si el modelo matemático ARIMA obtenido mediante el análisis de la serie de datos evolutivos indica que en la historia de la serie hay una intervención exógena que altera la secuencia del comportamiento histórico de la serie. En este caso, el comportamiento de la variable de interés presenta "saltos" , y entonces el cambio ocurrió conforme lo explica la teoría de los equilibrios puntuados.

3. Investigar si es posible mediante dicho análisis estadístico separar a los componentes azarosos o estocásticos de los componentes deterministas de la serie de datos considerada.

4. Investigar si es posible explicar al proceso evolutivo mediante un modelo estocástico-determinista que considere como fuerzas antagónicas a las fuerzas generadoras del cambio y a las fuerzas de conservación o mantenimiento.

5. Investigar si el análisis estadístico de series de tiempo permite distinguir si el cambio se debió a selección natural o a deriva génica.

CAPITULO 1.

HACIA UN MODELO ESTOCASTICO-DETERMINISTA DE LA EVOLUCION.

Evolucionar: desenvolverse, desarrollarse, pasar de un estado a otro. Cambio de forma.

Real Academia de la lengua Española

1.1. Evolución: Los marcos y las construcciones del tiempo.

Estudiar la evolución de las especies es tratar de comprender la vida desde una perspectiva muy peculiar; es entender la vida desde su complejidad y diversidad actual y verla de manera retrospectiva. Es entender su situación actual en términos de su historia, de su desarrollo en la abismal perspectiva del tiempo enmarcado en el registro fósil.

Evolución significa cambio en los seres vivos a través de largos periodos en el tiempo. Significa cambios en la forma y conducta de los organismos de una generación a la siguiente. "Descendencia con modificaciones" en palabras de Charles Darwin (Darwin, 1859; Ridley, 1993).

La evolución de las especies es un fenómeno natural que se ha desarrollado de manera continua en el tiempo, mediante procesos estocásticos y deterministas. El registro fósil y las evidencias paleontológicas han sido los testimonios más contundentes del fenómeno evolutivo, y actualmente se realizan muchas investigaciones "de campo" y en el laboratorio para estudiar al proceso evolutivo (Smith, 1982; Eldredge, 1985; Piñero, 1987; Egúarte y Souza, 1993).

La vida en nuestro planeta ha sido un continuo desde sus inicios. En palabras de Theodosius Dobzhansky "En biología, nada tiene sentido excepto a la luz de la evolución" (Dobzhansky 1980). Sin embargo, si bien la evolución biológica en nuestro planeta ha sido el resultado de un largo proceso de cambios, también es el resultado de un importantísimo mecanismo de memoria, es decir, de la conservación de las características que resultaron "exitosas" ante los problemas o distintas circunstancias que el medio les ha presentado a los organismos para permitir o no su existencia, desarrollo y

permanencia. Este proceso es el resultado en buena medida de circunstancias azarosas, es decir cambios impredecibles, sin embargo tiene también un fuerte componente conservador, determinista, de mantenimiento de las características no sólo morfológicas, sino también fisiológicas y etológicas de los organismos, probadas en muchísimas generaciones. Los organismos actuales somos descendientes de organismos que, en su momento, fueron buenos solucionadores de problemas (Soberón, 1987), y de cuya experiencia somos herederos.

La evolución es el cambio neto direccional o acumulado, en las características de los organismos, poblaciones y especies en el tiempo. Incluye el origen de alelos, variantes (individuos), valores de características (por ejemplo tamaño) o estados (función, forma por ejemplo). Por lo tanto, incluye el cambio en las poblaciones y especies. Los genetistas de poblaciones emplean la definición de evolución como el cambio en la frecuencia de los genes en el tiempo, por azar o direccionalmente (Endler, 1986; citado en Núñez y Cordero, 1993).

Ahora bien hemos que dicho que una de las polémicas contemporáneas medulares en Evolución es el modo como han surgido nuevas especies. El enfoque gradualista clásico que postula que la evolución ha sido un proceso gradual y continuo, que se da desde las poblaciones hasta la macroevolución; está siendo reemplazado por una concepción saltatoria o discontinua y se ha propuesto un enfoque alternativo: La Teoría de los Equilibrios Puntuados. Hemos mencionado también que este nuevo enfoque, propone una visión jerárquica de la evolución, y considera reduccionista a la visión gradualista que quiere extrapolar el comportamiento de la evolución de las poblaciones, hacia el surgimiento de nuevas especies y la macroevolución (Martínez, 1986; Soberón, 1986).

La teoría de los equilibrios puntuados considera que la adaptación se da asociada a estos procesos saltatorios, de cambios bruscos y sugieren que posiblemente sea consecuencia de algo que llaman "selección de especies": se forman especies al azar y algunas se extinguen y otras llegan por azar a algún pico adaptativo y son las que perduran. Entonces el proceso adaptativo se daría por

esta formación aleatoria y extinción diferencial de especies (Eguiarte, 1986).

De esta manera, el enfoque de los equilibrios puntuados considera que la mayor parte de las nuevas especies no han surgido de manera gradual a través de la acumulación paulatina de cambios adaptativos. Postula la posibilidad de que las nuevas especies hayan surgido bruscamente, de manera no adaptativa a través del rearrreglo de su material genético. Largos períodos de equilibrio o estasis son interrumpidos bruscamente por cambios repentinos, puntuados, que sustituyen al anterior equilibrio por un nuevo período de estasis. De esta manera las especies evolucionan de manera saltatoria brincando de un equilibrio a otro, de una especie a otra a través del rearrreglo de su material genético. Sería este rearrreglo el que serviría de barrera reproductiva y que aislaría a la nueva especie de la población original (Eguiarte, 1986; Olea, 1986).

La propuesta de los Equilibrios Puntuados sobre el modo puntuado de la evolución así como la posibilidad de que la gran correlación entre los organismos y su medio -la adaptación- se deba a un proceso a nivel mayor que el de las poblaciones, a nivel de especiación y extinción diferencial, es interesante. Sewall Wright afirma que los Equilibrios Puntuados pueden reducirse a la topografía adaptativa (Eguiarte, 1986).

Actualmente, la Biología evolutiva pretende explicar dos grandes temas: la diversidad biológica y la adaptación (Núñez y Cordero, 1993). La diversidad biológica, es decir, la coexistencia de diferentes "entidades" se aborda desde diferentes niveles de complejidad: molecular, celular, individual, poblacional y comunitario (Simpson 1947, Nei 1987, Wilson 1992, citados por Núñez y Cordero, op cit.). La adaptación, es decir, el ajuste fino de los organismos al medio ambiente, también ocurre a varios niveles: gene, linaje celular, organismo, familia y demo (Williams 1966, 1992, Harvey y Pagel 1991, citados por Núñez y Cordero, op cit.).

Al estudiar la diversidad y la adaptación, la mayoría de las preguntas que se plantean en torno a estos temas, caen dentro de dos categorías, origen y mantenimiento (Simmons y Parker 1989, citados en Núñez y Cordero 1993.). ¿Cómo se originan las adaptaciones?

¿Cómo se mantienen? Cuando se habla del origen de las adaptaciones se analizan las fuerzas selectivas que favorecen la propagación de un gene mutante X en el seno de una población; mientras que el mantenimiento se refiere al estudio de las fuerzas que permiten que el gene X, una vez que predomina en la población, no sea reemplazado por otro gene alternativo. Aunque este planteamiento se enfoca desde un nivel génico, es aplicable también a los demás niveles o entidades ya mencionados (Dawkins, 1982; citado en Soberón, 1986.). De manera similar, es posible analizar los procesos que generan u originan la diversidad biológica y los procesos de pérdida y ganancia de distintas entidades que afectan la diversidad ya existente, es decir su mantenimiento. (Núñez y Cordero opus cit., 1993).

Una manera de tratar tanto el modo como ha ocurrido el proceso evolutivo, así como el proceso de adaptación, es formalizar los equilibrios puntuados matemáticamente. ¿Es posible explicar en términos cuantitativos y predictivos el proceso evolutivo? ¿ Se puede formalizar matemáticamente la propuesta de la Selección de especies de los Equilibrios Puntuados? ¿Es reducible la propuesta de selección de especies a la topografía adaptativa? ¿Es posible explicar en términos cuantitativos y predictivos el proceso de adaptación?

La propuesta del presente trabajo, es formalizar matemáticamente una de las polémicas contemporáneas en torno al fenómeno evolutivo: el modo como ocurrió el cambio evolutivo. Concretamente, analizar los datos numéricos continuos disponibles en el registro fósil y modelarlos matemáticamente mediante el análisis estadístico de series de tiempo. Se propone entender a la evolución de las especies como procesos estocásticos deterministas que se han desarrollado durante largos periodos. Si se elige a alguna de las características del registro fósil que sea medible en términos de variables continuas, es posible considerar a dicha característica como una variable a la que denominaremos a partir de ahora como $Z(t)$,y estudiar su desarrollo en el tiempo mediante el análisis estadístico de series de tiempo univariadas, para explorar los alcances y limitaciones que esta herramienta estadística ofrece en el estudio de los fenómenos evolutivos.

Para ejemplificar esta propuesta, en la presente tesis se estudia un proceso estocástico biológico evolutivo: el caso de los radiolarios fósiles "Pseudocobus vema", en donde se analizan las variaciones morfológicas que presentan dichos organismos en su registro fósil. El registro fósil es un testimonio de las formas de vida que el tiempo ha permitido construir. Enmarcados en armazones geológicos, los fósiles son el elocuente testimonio, la huella de manifestaciones de vida anteriores a las actuales.

Este estudio es muy interesante y singular, porque los casos de características medibles como variables continuas en el registro fósil son extraordinarias. Sin embargo, cabe enfatizar que los procesos estocásticos evolutivos se presentan en todos los niveles de complejidad anteriormente mencionados, y casos de variación en la abundancia poblacional o variaciones en alguna característica morfológica son tan solo algunos ejemplos en los que se puede aplicar este tipo de análisis.

A continuación se explicará qué es un proceso estocástico, a diferentes niveles: lineales, biológicos y evolutivos.

1.2. Procesos estocásticos lineales y las series de tiempo.

Los procesos estocásticos son fenómenos o eventos azarosos, es decir su desarrollo depende del azar. Un proceso estocástico es lineal, si representa una familia de variables aleatorias asociadas a un conjunto índice de números reales, de tal forma que a cada elemento del conjunto le corresponde una y sólo una variable aleatoria. Esto se escribe como $\{ Z(t), t \in T \}$; donde T es el conjunto índice, y $Z(t)$ es la variable aleatoria correspondiente al elemento t de T (Guerrero, 1991a, Wei, 1990).

Las partículas de una solución con movimiento browniano, son un claro ejemplo de un proceso estocástico (ver Peitgen, Jürgens y Saupe, 1992. *The Chaos Game: How Randomness Creates Deterministic Shapes*. Capítulo 7 de "Fractals for the Classroom"). Una y sólo una partícula puede ocupar un lugar específico en el

"universo " de la solución a un tiempo "t" dado, es el azar quien define qué partícula ocupa ese lugar y en ese preciso momento. Las posibles partículas que pueden ocupar ese lugar del universo considerado, y en ese momento, constituyen una variable aleatoria y el conjunto índice de números reales al que están asociadas es el tiempo.

El paso de iones y de partículas a través de membranas biológicas, mediante mecanismos de difusión y flujos debidos a gradientes de concentración químicos y/o gradientes electroquímicos también son fenómenos que puede analizarse en términos de fluctuaciones probabilísticas y estocásticas (Berg, 1993: "Random walks in Biology").

Los procesos estocásticos pueden ser discretos o continuos. Si T es un conjunto finito o infinito, pero numerable, el proceso es discreto. Si T es un intervalo de números reales abierto o cerrado, el proceso estocástico es continuo.

Ahora bien, hemos dicho que un proceso estocástico es lineal, si representa una familia de variables aleatorias asociadas a un conjunto índice de números reales. Si el conjunto de números reales es una escala en el tiempo, y si a cada valor de la escala del tiempo le corresponde una y sólo una variable aleatoria, entonces tenemos una serie de tiempo.

Una serie de tiempo es una secuencia ordenada de observaciones. El ordenamiento usualmente es a través del tiempo, a intervalos fijos, de igual magnitud; pero el orden también puede considerarse en otras dimensiones, como el espacio (Wei, 1990). La profundidad en un registro fósil también puede considerarse como una escala de ordenamiento válida para series de tiempo, si tiene intervalos de tiempo fijos.

Las series de tiempo también pueden ser discretas o continuas. Una serie de tiempo es discreta si las observaciones se hacen en los momentos t_1, t_2, \dots, t_n ; y estos intervalos tienen una longitud de tiempo fija. El proceso estocástico respectivo a esta serie, se denota por

$$\{ Z(t_1), Z(t_2), \dots, Z(t_n) \}$$

"Z" denota a la variable aleatoria del proceso estocástico, es decir, su valor observado (Guerrero, 1991a).

En N valores sucesivos de una serie de tiempo, las observaciones se escriben como $Z_1, Z_2, \dots, Z_t, \dots, Z_n$; y denotan las observaciones hechas en los intervalos equidistantes $t_0+h, t_0+2h, \dots, t_0+th, \dots, t_0+Nh$, en donde t_0 es algún punto en el tiempo que hace las veces de origen, y h es la longitud del intervalo de tiempo que separa a dos observaciones contiguas (Guerrero, 1991a).

Un proceso estocástico tiene varias posibles realizaciones (¡las partículas de la solución se pueden acomodar de muchas maneras!). La realización de una serie de tiempo depende de un fuerte elemento probabilístico. El valor observado Z_t es uno en una familia de posibles valores Z . Sin embargo, el comportamiento de todas las variables aleatorias Z involucradas en el proceso estocástico, puede caracterizarse a través de la función de densidad $f(Z)$. Así, dos variables aleatorias Z_1 y Z_2 , quedan descritas en términos probabilísticos por su función de densidad conjunta $f(Z_1, Z_2)$. En general N variables aleatorias, es decir un proceso estocástico, pueden escribirse mediante la función de densidad conjunta $f(Z_1, Z_2, \dots, Z_n)$. (Wei, 1990).

De esta forma, una serie de tiempo es una muestra de las posibles maneras en que puede desarrollarse un proceso estocástico lineal. La serie de tiempo en estudio es una sucesión cronológica de observaciones del proceso estocástico, es un conjunto de mediciones metódicas que se registran a intervalos de tiempo fijos.

1.3. Procesos estocásticos biológicos.

El registro de un proceso estocástico se observa como la trayectoria de una partícula que se mueve al azar en el espacio S , siendo Z_t su posición en el instante t . Los valores que puede tomar Z_t , en el espacio S al tiempo t , son llamados sus estados, y los cambios en el valor de Z_t , reciben el nombre de transiciones entre sus estados (Coleman, 1976). Muchos fenómenos biológicos son

procesos estocásticos y pueden ser entendidos como sistemas cuya realización en el tiempo se desarrolla como fluctuaciones al azar. Los modelos estocásticos son aplicables a cualquier sistema que comprenda variabilidad al azar en el transcurso del tiempo.

Variabilidad al azar en el transcurso del tiempo es una característica de muchos sistemas biológicos. Por tanto, dependiendo de la característica del fenómeno que se considere como variable de interés, los procesos estocásticos en Biología se encuentran en áreas muy variadas. Cabe mencionar que los fenómenos biológicos son resultado de un gran número de variables; por lo cual, cuando se elige una "variable de interés" para estudiar a un fenómeno biológico, las demás variables se consideran constantes.

En Ecología de comunidades la variable de interés puede ser el cambio en el número de especies de una comunidad en relación al tiempo.

En estudios epidemiológicos, se estudia la propagación de una enfermedad mortal en una comunidad mediante modelos estocásticos, donde se clasifican en los estados o valores de la variable Z_i al número de individuos que se encuentran en una categoría dada al tiempo considerado: sanos inmunes, sanos propensos, portadores, infectados y fallecidos. De un tiempo considerado, al siguiente, las transiciones entre las clases un estado al otro indican la velocidad con que se propaga la enfermedad en la comunidad. Los epidemiólogos usan la teoría de los procesos estocásticos para buscar las formas de influir en la rapidez de transición de la enfermedad entre las clases consideradas.

En Ecología de poblaciones los modelos estocásticos son muy importantes, ya que analizan las variaciones en la abundancia de los organismos, es decir, las variaciones en el tamaño poblacional (Begon, Harper y Townsend 1986). Por ejemplo, en estudios con poblaciones de insectos se analiza el número de individuos que se encuentran en un estadio al tiempo t . Aquí en cada valor de la variable Z_i se clasifican a los insectos por el número de individuos que se encuentran en un estadio determinado: las categorías pueden ser: huevecillos, pupas, larvas I, larvas II ..., adultos. La transición entre el

estado de la variable Z_i al tiempo t y sus siguientes estados explica el crecimiento y desarrollo de la población.

En estudios con poblaciones vegetales es posible considerar un modelo similar, si las categorías consideradas para cada estado de la variable Z_i son por ejemplo: semillas, plántulas, adultos en floración, etc.

En estudios poblacionales, las variaciones en la abundancia poblacional de los organismos son características ya clásicamente estudiadas como procesos estocásticos que varían en relación al tiempo. El caso de las fluctuaciones del tamaño poblacional de los linces canadienses es ya un ejemplo clásico porque se trata de un modelo muy bonito donde se tiene el registro del número de linces en Canadá desde 1870 hasta las poblaciones actuales. Para ejemplificar la aplicación del análisis estadístico de las series de tiempo univariadas a un proceso estocástico biológico, en el capítulo 5 se presenta el análisis estadístico de esta serie de datos de los linces cazados en Canadá desde 1870 hasta 1913.

En estudios de Fisiología, los relojes biológicos y los ciclos circadianos también pueden ser analizados como series de tiempo, si la característica en estudio toma distintos valores numéricos continuos a intervalos de tiempo fijos. Sin embargo, en investigaciones sobre potenciales cerebrales evocados y ciclos cardíacos, no es recomendable este tipo de análisis porque se trata de fenómenos con una marcada periodicidad. En este tipo de estudios la característica de interés repite valores en cada periodo, por tanto, no se trata de procesos estocásticos. Para modelar el comportamiento de este tipo de fenómenos se utiliza el "Análisis de Fourier". (ver Wei, 1990, capítulo 10.)

A nivel de dinámica molecular sería muy interesante aplicar esta técnica estadística. En estudios de Bioquímica, y de Biología Celular donde interese modelar el comportamiento de poblaciones celulares u otro tipo de "poblaciones" en relación al tiempo. Por ejemplo, en investigaciones sobre hipertensión arterial esencial, es muy importante conocer las variaciones en la abundancia de las poblaciones de adrenoceptores beta bajo distintas condiciones; el análisis estadístico de las series de tiempo univariadas podría ayudar

a modelar y analizar el comportamiento de este tipo de poblaciones de receptores hormonales, ya que en este caso, el tamaño de las poblaciones de adrenoceptores beta presenta marcadas fluctuaciones en relación al tiempo (Núñez Z. M. 1984. Tesis de Licenciatura: Método de Aislamiento de leucocitos polimorfonucleares en el estudio de receptores beta adrenérgicos, UNAM).

1.4. Procesos estocásticos en Evolución:

En la Evolución de las especies, las alteraciones morfológicas de los organismos (o su persistencia), como resultado de su adecuación al medio, son también procesos estocásticos, ya que se trata de cambios azarosos, variaciones al azar de una característica en relación al tiempo. Si se tiene un registro de datos numéricos, mas que de datos categóricos, de la característica de interés, a intervalos de tiempo fijos, entonces es posible modelar los datos y explicar el fenómeno evolutivo de esa característica en especial, mediante el análisis estadístico de las series de tiempo univariadas. Como ya ha sido mencionado, esta es la propuesta que se investiga en la presente tesis.

- Modelo estocástico determinista de la Evolución.-

¿Es posible explicar a la Evolución de las especies mediante un modelo estocástico-determinista? ¿Cuáles son las fuerzas del cambio y cuales son las fuerzas de mantenimiento de la Evolución?

Un experimento es una pregunta que se hace a la naturaleza. El procedimiento científico empieza por seleccionar las variables de observación. Con el método científico primero se observa, se escucha a la naturaleza y después, mediante la experimentación se realiza una "observación razonada", cuyo resultado se confronta con la pregunta formulada en el hipótesis de trabajo (Rosenblueth,1971).

La ciencia sin matemáticas es una ciencia pobre y limitada. Las matemáticas en la ciencia no dictan normas a la naturaleza, sino que deben ajustarse a ella (Rosenblueth, 1971).

El análisis de datos experimentales que han sido observados en diferentes puntos en el tiempo, conduce a problemas únicos y nuevos de modelaje e inferencia estadística. El método sistemático por el que se llega a responder matemática y estadísticamente a preguntas planteadas sobre bases de datos con este tipo de correlaciones en el tiempo, es el análisis de series de tiempo (Shumway, 1988). Una serie de tiempo en su forma más simple, no es más que un conjunto de datos $Z(t)$ que se observan en intervalos fijos del tiempo t (Diggle, 1990). Este análisis estadístico modela al fenómeno en estudio y permite realizar una "observación razonada" modela las observaciones y realiza inferencia estadística sobre la muestra de datos considerada.

Ahora bien, ¿Si analizamos bases de datos numéricos continuos del registro fósil mediante el análisis de series de tiempo univariadas, puede llevarnos la metodología de las series de tiempo a un modelo matemático que nos permita discernir los componentes deterministas de los componentes estocásticos en un proceso estocástico evolutivo?. ¿Cuáles son los componentes estocásticos y cuáles son los componentes deterministas de la Evolución?. ¿Es reduccionista tratar de entender en estos términos a la Evolución?.

- El fenómeno evolutivo como un proceso estocástico-determinista: Las fuerzas de la Evolución Biológica: Aspectos deterministas y aspectos estocásticos de la evolución biológica.-

1.4.1 Componentes deterministas de la evolución biológica: La selección natural y las fuerzas conservadoras o de mantenimiento: La "memoria" del proceso evolutivo.

La teoría de la evolución moderna considera que una serie de fuerzas actúan sobre la variación genética de las poblaciones (Wright 1932, Dobzhansky 1937,1970, Eguiarte 1988 y Eguiarte y Souza 1993.) Estas fuerzas son: la selección natural, como un componente determinista que elimina parte de la variación dejando los genotipos más adaptados; la deriva génica, que elimina la variación genética de manera azarosa y cuya importancia depende del número "efectivo" de individuos que deja descendientes de una generación a otra; la mutación, que genera la variación genética; el flujo genético, que

introduce nueva variación proveniente de otras poblaciones, la recombinación genética entre cromosomas homólogos, y, por último, la endogamia, que se produce cuando en la población se llevan a cabo más apareamientos entre parientes de los esperados al azar, y modifica la proporción de individuos de un genotipo dado. La importancia de relativa de estas fuerzas, junto con la genética de un sistema dado, determina qué modelo de evolución opera en una población dada.

1.La selección natural.-

La selección natural es considerada como un proceso universal, (Smith, 1982; Ridley, 1993.) y la causa primordial de la evolución y la adaptación de las especies a diferentes medios de vida (Núñez y Cordero, op. cit.).

La teoría de la selección natural fue propuesta por Darwin y Wallace para explicar la evolución adaptativa de las especies (Darwin y Wallace, 1858, Darwin 1859). El 11 de Enero de 1844, Charles R.Darwin le confió a su joven amigo Joseph Hooker que en la cuestión de la evolución de las especies, ".....al fin comienzo a ver algo de luz. A pesar de que al principio mis opiniones eran contrarias a las de ahora, estoy convencido casi por completo que las especies no son inmutables. Decir esto es casi como confesar un asesinato (...) Creo que he descubierto (¡y mire usted qué presuntuoso soy!) la naturaleza de los mecanismos que permiten que las especies se adapten en forma tan exquisita a distintas condiciones" (Darwin Ch. citado por Lazcano,1994).

Actualmente, sabemos que la evolución de las especies puede ser resultado de otros procesos, como deriva génica, recombinación cromosómica, migraciones, mutaciones y su interacción con la selección natural (Wright, 1981,1982, citado en Núñez y Cordero 1993). En la misma forma, la adaptación de las especies, podría tener como origen, procesos estocásticos u otros fenómenos (Lewontin 1978, Gould y Lewontin 1979, citados en Núñez y Cordero op.cit.). La selección natural puede mantener o mejorar la adaptación de las especies, pero no originarla (Núñez y Cordero op.cit.).

Modelos de evolución adaptativa.

Básicamente los modelos adaptacionistas del proceso evolutivo se pueden reducir a 3 (Eguiarte, 1993)

1. Modelo Fisheriano
2. Modelo del Equilibrio cambiante ("shifting balance")
3. Modelo de la Reina Roja de Van Valen (1973).

1. Modelo Fisheriano (Fisher 1958): este modelo sugiere que la selección natural es una fuerza muy poderosa u que las distintas poblaciones de una especie siempre llegan a las mismas adaptaciones en un ambiente similar. Considera que existen pocas (generalmente sólo una) "solución" (adaptación) a un problema dado, y que la deriva génica juega un papel poco importante en la evolución (Eguiarte y Souza, 1993). La gráfica que ilustra a este proceso es una topografía o paisaje adaptativo donde se presenta la adecuación de dos frecuencias alélicas "a" y "b". El valor máximo está en un solo punto, "ab", donde está la máxima adecuación promedio "w".

2. Modelo del equilibrio cambiante o "Shifting balance" de S. Wright (1932): considera que la evolución es el resultado, tanto de la selección natural, como de la deriva génica y que las interacciones genéticas producen una "topografía adaptativa" compleja, con distintas soluciones de diferente adecuación promedio (picos adaptativos). La gráfica que ilustra este modelo es un paisaje topográfico con dos o más "picos adaptativos" que corresponden a los valores más altos de la adecuación "w".

3. Modelo de la Reina Roja de Van Valen (1973): este modelo sugiere que los mismos organismos están modificando la forma de la topografía adaptativa, por lo que nunca se llega a un punto óptimo y que, por lo tanto, las poblaciones tienen que estar cambiando constantemente. Este modelo fue nombrado así en honor de la Reina Roja de Alicia a través del espejo de Lewis Carroll, por Van Valen, ya que es un personaje que tiene que correr todo el tiempo para mantenerse siempre en el mismo lugar. Así, según este modelo, aunque la selección natural constantemente está eliminando variación, las poblaciones no mejoran su adaptación en términos

absolutos, debido a que la topografía adaptativa ha cambiado y las poblaciones están otra vez poco adaptadas (Eguarte y Souza, op. cit.).

Según este modelo, las variaciones del medio ambiente, se solucionan con "plasticidad" de la población gracias a su variabilidad. Aquí surge una pregunta interesante: ¿y de dónde surge tanta variabilidad en la población? ¿Las mutaciones, recombinación génica, las migraciones son explicaciones suficientes?

2. La memoria del proceso evolutivo: EL GENOMA.-

Como ya hemos dicho, los organismos actuales somos los descendientes de organismos pasados que fueron buenos solucionadores de problemas (Soberón, 1987), y de cuyo material genético "rebarajado" y "corregido" somos los representantes actuales. La vida, es la expresión de un proceso continuo que se desarrollan en el tiempo. Aunque la vida de un individuo tenga un inicio, en el cual se constituye precisamente como individuo, y su muerte sea un fin "individual", su proceso vital es parte de un continuo, porque su genoma le ha sido heredado desde sus mas remotos ancestros. Las moléculas de sus ácidos nucleicos se han transmitido de una generación a otra, por más de 3500 millones de años, modificándose y ramificándose, desde las primeras manifestaciones de la vida hasta llegar a su estado actual (Pifero, 1987) y, en su caso, serán transmitidas a sus descendientes. Todos los organismos vivos estamos "hermanados" en mayor o menor medida por nuestro genoma.

1.4.2 Componentes estocásticos del fenómeno evolutivo: las fuerzas generadoras del cambio: mutaciones, deriva génica o deriva genética al azar, recombinación cromosómica y migraciones ó flujo génico.

Estocástico significa azaroso. Los componentes estocásticos del fenómeno evolutivo son aquellos factores que provocan cambios al azar, sin dirección, en las especies. El azar en la evolución de las especies opera a muchos niveles (Dobzhansky 1980). Sin embargo, el

cambio evolutivo se produce en las poblaciones, no en los individuos. El individuo nace, crece y finalmente muere. Desde el punto de vista evolutivo, el individuo es efímero; sólo la población es continua; la continuidad deriva del mecanismo de la herencia biológica. Los organismos individuales pueden cambiar a lo largo de su vida, pero su constitución genética permanece constante. Por el contrario, la constitución genética de una población puede cambiar de una generación a otra, y así sucede por lo general. La evolución consiste en cambios en la constitución genética de las poblaciones. (Ayala, 1980).

El cambio genético en las poblaciones se produce a través de los procesos de mutación, deriva génica al azar, migración o flujo génico y selección natural (Ayala, 1980).

Mutaciones:

La mutación es la fuente última de toda la variación genética (Ayala, 1980). A nivel del acervo de genes ("gene pool"), las mutaciones producen variaciones aleatorias respecto al ambiente, por lo que no pueden dirigir por sí mismas la evolución hacia nuevas formas adaptativas al medio. Las mutaciones, junto con la deriva génica, son el componente estocástico del proceso evolutivo, a nivel molecular.

Deriva génica o deriva genética al azar:

La deriva genética al azar es el cambio en las frecuencias de genes de una generación a otra, debida a los "errores o accidentes de muestreo" de una generación a otra. Estos "errores del muestreo" se deben al azar, son sucesos estocásticos porque por puro azar la composición de genes es diferente en las distintas generaciones de una población. Este "error de muestreo" provoca que exista una gama de alternativas en las frecuencias génicas de una generación a la siguiente; son variantes de un mismo acervo de genes, pero ligeramente diferentes por este error de muestreo. Por esto también es una fuente de variabilidad a nivel molecular. La composición genética de una generación a otra, y de una población a otra no es una copia idéntica de la generación anterior, es una variante de la anterior, es una muestra nueva, ligeramente diferente a la muestra anterior, pero tomada de la misma población de genes disponibles en ese momento. Cuanto más pequeña es la población, mayor es el

efecto de la deriva génica al azar sobre las frecuencias de genes (Ayala, 1980).

La deriva génica o deriva genética es por tanto, un factor que produce cambios, variabilidad, desviaciones aleatorias en la composición génica de la población. Así, nuevas combinaciones de genes entran en juego de una generación a la siguiente, con las exigencias adaptativas del medio marcadas por la selección natural.

Recombinación genética:

Formación de una nueva combinación genética resultante del cambio recíproco entre cromosomas homólogos (Sarukhán, 1988).

Migraciones o flujo genético:

La migración o flujo génico de una población a otra de la misma especie puede aumentar la variación genética en las poblaciones locales (Ayala, 1980).

La selección natural no es un componente estocástico del fenómeno evolutivo, hemos dicho que es un componente determinista porque determina cuales son las formas o variantes más exitosas de una población. La selección natural puede definirse como la reproducción diferencial de variantes genéticas alternativas. Los organismos que poseen variantes génicas útiles para sus portadores porque en ese momento resultan las mejores respuestas adaptativas al medio ambiente, es probable que dejen más prole que los organismos con variantes genéticas alternativas, menos adaptativas (Ayala, 1980).

El potencial evolutivo de una población está determinado por el grado de variación genética que existe en la población. La selección natural, al igual que la deriva génica la azar, solamente puede tener lugar en poblaciones que posean variabilidad genética.

1.5 Modelos actuales sobre la Evolución Biológica.

Históricamente ha habido dos escuelas, dos enfoques o modelos que explican como se dió el proceso evolutivo:

1.5.1 Gradualismo filético: bajo la sombra de Darwin.

La escuela tradicional del Gradualismo filético defiende el enfoque Darwinista tradicional que explica la evolución mediante un modelo determinista en el cual la Selección Natural es la fuerza que determina las frecuencias génicas en una población, y por tanto las características poblacionales más frecuentes que se presenten en dicha población en determinado momento. En este enfoque, las circunstancias ambientales generan un mecanismo de presión que favorece a ciertas características y actúa en contra de otras características también presentes en la población, lo que se traduce como un cambio en las frecuencias génicas de la población.

El ejemplo que Charles Darwin no conoció y que le faltó para ilustrar el mecanismo de acción de la Selección Natural, es el caso del melanismo industrial de las polillas inglesas "Biston betularia" y la contaminación generada por la Revolución Industrial Británica (Bishop 1972, citado en Núñez y Cordero, 1993.). La revolución industrial trajo consigo contaminantes que se depositaron en las cortezas de los árboles; de manera que las polillas claras ahora perdían su mimetismo al posarse en los árboles, resultaban más evidentes a sus depredadores; mientras que las polillas oscuras resultaron favorecidas por el nuevo cambio ambiental y ganaron mimetismo al posarse en las cortezas, ahora oscurecidas por el hollín provocado por la contaminación ambiental. Este cambio ambiental, favoreció a las polillas oscuras. La presión de selección natural aumentó sobre las ahora evidentes polillas claras, (¡sus pájaros depredadores las veían fácilmente contra un fondo oscuro!), disminuyó sobre las polillas oscuras y con el tiempo las frecuencias fenotípicas y genotípicas de la población local, se modificaron.

El enfoque gradualista explica que la evolución de las especies es el resultado de la acumulación gradual de cambios en los organismos. En este punto es importante mencionar que actualmente

hay mucha discusión sobre el nivel al que actúa la Selección Natural, y sobre las unidades que se seleccionan. Se discute sobre megaevolución, la macroevolución y la microevolución. En cuanto a las unidades que se seleccionan se propone a la especie como unidad de selección, a las comunidades, a las poblaciones, a los individuos e incluso se ha considerado al gen como unidad de selección natural, en la teoría del "gen egoísta" de Dawkins (Dawkins 1977; citado en Soberón, 1986).

Bajo la perspectiva Darwiniana, la paleontología explica el origen de nuevos taxa con las siguientes aseveraciones que ilustran claramente los postulados del gradualismo filético: (Eldredge, 1985)

1. Las nuevas especies surgen de la transformación de una población ancestral en sus descendientes modificados.
2. La transformación es lenta y "apenas notoria" (Inconspicua).
3. La transformación involucra a un gran número, si no a todos los miembros de la población ancestral.
4. La transformación ocurre sobre todo o una gran parte del rango geográfico de la especie ancestral.

Estas aseveraciones implican varias consecuencias, dos de las cuales parecen ser especialmente importantes para los paleontólogos:

1. Idealmente, el registro fósil para el origen de una nueva especie debería consistir en una larga secuencia de cambios continuos, de formas intermedias insensiblemente graduadas que unieran a los ancestros con sus descendientes
2. Bajo el postulado de una secuencia de gradualismo filético, las rupturas morfológicas del registro fósil, se deben a imperfecciones en el registro geológico.

La rareza de las series transitorias en el registro fósil permanece como un "punto flaco" del gradualismo filético. Bajo esta perspectiva, siglo y medio después, los reclamos antievolutionistas de Cuvier todavía no han sido satisfactoriamente acallados: "Si la evolución es verdadera, deberían existir miles, millones de formas transitorias constituyendo una cadena irrompible" (Anónimo, 1967.

Did Man Get Here by Evolution or by Creation? -panfleto de un testigo de Jehová citado en Eldredge N. 1985).

1.5.2 Teoría de los Equilibrios Punteados:

En el estudio de la evolución de las especies, el paradigma central es sobre el modo en que se ha dado el cambio. ¿Cómo se ha dado el cambio? ¿El cambio ha sido gradual? o por el contrario ¿las modificaciones se van acumulando durante mucho tiempo, hasta que "de pronto" se da un salto de un estado de estasis inicial, hasta un nuevo estado de "estasis" o equilibrio?.

La posición de la escuela de los gradualistas, cuyo representante más famoso es Charles Darwin, fue refutada en la década de los años 70s, por dos paleontólogos, los investigadores: Stephen Jay Gould y Niles Eldredge, quienes dieron una nueva interpretación a los datos disponibles en el registro fósil, al postular su teoría de los "Equilibrios Punteados".

Gould y Eldredge, afirman que es necesario dejar de buscar "eslabones perdidos", y que se debe interpretar al registro como es. Es decir, no es que el registro fósil sea imperfecto, sino que la evolución de las especies se ha dado "por brincos" y por tanto no hay puntos intermedios entre una característica y la aparición de una característica nueva (Gould y Eldredge, 1977, 1986; Eldredge, 1985).

Como ya se mencionó, el interés central de esta tesis, es investigar si el análisis estadístico de las series de tiempo univariadas es aplicable a procesos estocásticos evolutivos, y si es posible argumentar matemáticamente a favor de un cambio gradual, o por el contrario, el modelo matemático explicaría un comportamiento por "equilibrios punteados". Concretamente, en esta tesis se aplica el análisis estadístico de las series de tiempo univariadas a los datos numéricos del registro fósil publicados en 1977 por Gould y Eldredge en su artículo "Punctuated Equilibria: the tempo and mode of evolution reconsidered". Gould y Eldredge, se basaron en este tipo de registros para fundamentar su teoría de los Equilibrios Punteados.

Es importante mencionar, que en el registro fósil no es fácil encontrar datos numéricos de una característica a intervalos de tiempo constantes. En el registro fósil en general se encuentran datos categóricos, pero es muy difícil encontrar una serie de datos continuos. La escuela de los Gradualistas, defiende su postura argumentando "imperfecciones" del registro fósil. Donde no hay una clara evidencia del paso gradual de una característica a otra, se ha hablado de "eslabones perdidos", es decir "paréntesis" en donde al parecer no hay información porque se ha perdido. Puede ser que esas partes de estructura de los organismos donde se apreciaba el cambio gradual de una característica a otra no se pudieron fosilizar por tratarse de tejidos blandos; o bien porque simplemente no se dieron las condiciones ambientales necesarias para la fosilización de esas estructuras. De hecho, son muy pocos los casos en el registro fósil donde se tiene el registro del cambio gradual de una característica en un tiempo largo.

En contraste con los postulados del gradualismo filético sobre las tendencias y predicciones de la especiación alopátrica, La Teoría de los equilibrios puntuados propone que respecto a la especiación alopátrica (Eldredge, 1985):

1. Las nuevas especies surgen por la división de linajes.
2. Una nueva especie surge y se desarrolla rápidamente.
3. Una pequeña sub-población de la forma ancestral da lugar a la nueva especie.
4. Las nuevas especies se originan en lugares muy pequeños de la extensión geográfica ancestral, en un área aislada en la periferia del rango geográfico total de la especie ancestral.

Estas cuatro aseveraciones implican dos importantes consecuencias:

1. En cualquier sección local que contenga a la especie ancestral, el registro fósil de los descendientes, puede consistir en un **rompimiento morfológico marcado, drástico** entre las dos formas. Este rompimiento marca la migración de los descendientes al área periféricamente aislada en donde se desarrollaron a partir de su forma ancestral. Cambios morfológicos en su antecesor, aún cambios direccionales en el tiempo, no deben mostrar relación con la

morfología de los descendientes, ya que estos cambios en los descendientes surgen como respuesta a las condiciones locales de su área geográficamente aislada. Dado que la especiación ocurre rápidamente en pequeñas poblaciones ocupando regiones geográficas pequeñas alejadas del centro de abundancia ancestral, raramente encontraremos el evento "actual" en el registro fósil (Eldredge, 1985).

2. Muchos rompimientos del registro fósil son reales; ellos expresan el modo en que ocurrió la evolución, no son fragmentos de un registro fósil incompleto. La marcada ruptura en una columna local, registra fielmente lo que ocurrió en esa área a través del tiempo. La tradición institucional milenaria dicta al paleontólogo una reacción de "avestruz" y le hace ver cualquier anomalía como una "imperfección del registro fósil". Sin embargo, son reflejos verdaderos de la historia de la vida. Este testimonio, este registro fósil es mucho mejor (o de menos más rico en el óptimo de los casos) de lo que la tradición dicta (Eldredge, 1985).

1.5.3 ¿El paradigma actual sobre la evolución de las especies?

El Darwinismo es la piedra angular de la Biología Evolutiva, sin embargo, no es legítimo igualar a la Biología Evolutiva con el Darwinismo. Esto llevaría a confundir el fenómeno con la explicación (Lazcano, 1994). La Biología Evolutiva no es el patrimonio de una sola corriente o escuela, actualmente ésta ha trascendido no sólo los postulados de Darwin, sino también las premisas de los neodarwinistas, y el programa de trabajo de la nueva síntesis. Actualmente la Biología Evolutiva considera además del gradualismo filético, a los neutralistas de Kimura, a los saltacionistas, a los estructuralistas y a quienes reconocen el significado evolutivo de los fenómenos simbióticos. No es que el neodarwinismo esté equivocado, más bien, es que está incompleto (Lazcano, 1994).

¿Y cómo completarlo?

En primer lugar, reconociendo los límites de sus aciertos y la legitimidad de algunos de sus prejuicios. En segundo lugar,

admitiendo que el tránsito hacia una descripción más rica y completa de los procesos evolutivos ocurrirá cuando reconozcamos la existencia de una enorme variedad de procesos evolutivos que operan a distinta escala, y con diferente rapidez. Más allá de fobias y afectos, debemos reconocer el valor de Stephen Jay Gould, George Miklos, Niles Eldredge Lynn Margulis y muchos más, al desafiar el diálogo circular y endogámico que hacía de la selección natural el único mecanismo evolutivo en un proceso descrito de manera inevitable como lento y gradual. (Lazcano, ¿Y si Darwin no hubiera muerto?. La Jornada Semanal, Mayo 8 de 1994).

Y en tercer lugar, que es la propuesta del presente trabajo, formalizando matemáticamente las polémicas contemporáneas, en la medida en que los datos lo permitan, y diseñando investigaciones con protocolos que consideren herramientas estadísticas apropiadas para tratar datos en el tiempo, es decir tratar a los datos en evolución como lo que son, series de datos en el tiempo, que puedan dar respuestas con el formalismo de los modelos estadístico-matemáticos.

CAPITULO 2.

ANALISIS ESTADISTICO DE SERIES DE TIEMPO UNIVARIADAS: CONCEPTOS INTRODUCTORIOS Y ANTECEDENTES.

2.1 Series de tiempo.

Una serie de tiempo es una sucesión cronológica de datos u observaciones generadas por un proceso estocástico cuyo índice se toma en relación al tiempo. Una serie de tiempo es un registro metódico de mediciones u observaciones numéricas de una variable o característica de interés, efectuadas a intervalos de tiempo fijos (Guerrero, 1991a).

2.2 Objetivos de un análisis de series de tiempo.

Un modesto objetivo de cualquier análisis de series de tiempo es brindar una concisa descripción de una serie histórica. La descripción puede consistir en unos cuantos estadísticos, y probablemente incluir una o más representaciones gráficas de los datos. Esto es debido a la complejidad de una serie de tiempo, que se opone a una muestra tomada al azar de una distribución. Más que un sencillo número, se requiere de una función que sumarice cada una de sus características esenciales. Un sencillo ejemplo es la descripción del valor de la media, como una función del tiempo, $\mu(t)$, mas que una sencilla cantidad de μ (Diggle, 1990).

Una meta más ambiciosa es pronosticar los valores futuros de una serie. De hecho la mayoría de los análisis de series de tiempo se han realizado con este propósito, con obvia relevancia en pronósticos de ofertas y otras aplicaciones económicas (Box y Jenkins 1970, citados en Diggle, 1990).

Un problema que surge de manera natural en medicina y en otras muchas disciplinas, es el monitoreo de una serie de tiempo para detectar cambios en la conducta de la serie conforme ocurren (Smith y West, 1983. citados en Diggle, 1990).

Un papel más pasivo, pero no menos importante del análisis de series de tiempo, es establecer una dependencia en serie al hacer inferencias sobre los parámetros estructurales básicos. La forma de los datos en series de tiempo es muy importante para el desarrollo de procedimientos inferenciales válidos y eficientes.

Finalmente, se distinguen dos tipos de modelación de datos de series de tiempo. Los modelos pueden realizarse con miras a una comprensión científica posterior de los mecanismos subyacentes que generaron los datos, o de manera más pragmática como una herramienta auxiliar para lograr objetivos previamente establecidos (Diggle, 1990 op. cit.).

2.3 Análisis estadístico de series de tiempo.

La metodología del análisis estadístico de series de tiempo univariadas es aplicable a series de tiempo de cualquier área del conocimiento, donde la información es recopilada en forma de datos numéricos con fines de descripción, análisis, planeación y/o toma de decisiones (Guerrero, 1991a; Wei, 1990).

La estrategia para realizar el análisis de series de tiempo que aquí se presenta, es el sugerido por Box y Jenkins en su libro publicado en 1970. *A partir de aquí toda la exposición de la metodología, se basa en la concepción teórica del Dr. Victor Manuel Guerrero (Guerrero, 1991a)

La construcción de los modelos de la familia ARIMA, ("Autorregresive, Integrated, Moving Average Models") surge tanto de consideraciones meramente estadísticas, a fin de tener en cuenta la aleatoriedad de los fenómenos, como de modelos dinámicos puramente deterministas, que han aparecido dentro del área de la matemática discreta y a los cuales se conoce como ecuaciones en diferencias.

La explicación desde el punto de vista estadístico, de cómo se pueden generar series de tiempo a partir de procesos estocásticos muy generales se presenta de manera detallada en el siguiente capítulo. En contraste, ahora se explicará el aspecto determinista de las series de tiempo, es decir, se explicará cuál es el proceso generador de la serie.

A continuación se presenta una breve introducción a las ecuaciones en diferencias, para entender las herramientas matemáticas en que se basa la metodología de Box y Jenkins para el análisis de las series de tiempo.

2.4 Ecuaciones en diferencias para procesos deterministas.

En esta sección se presentan los conceptos básicos de ecuaciones diferenciales y ecuaciones en diferencias, necesarios para entender la metodología general que se sigue para resolver ecuaciones en diferencias de primero, segundo y de orden "p" empleadas por Box y Jenkins en el análisis de series de tiempo.

Ecuaciones diferenciales.-

Las ecuaciones diferenciales describen funciones entre la variable dependiente (y), en relación a la variable independiente (x), cuando ésta es una variable continua, es decir puede tomar cualquier valor sobre el eje de las abscisas. Entonces, cuando se considera que la variable independiente (x), es el tiempo y éste es continuo, el comportamiento de la variable $Z(t)$, a través del tiempo queda determinado por sus derivadas:

$$\frac{dZ}{dt} = ky \quad \longrightarrow \quad \frac{dZ}{dt}, \frac{d^2Z}{dt^2}, \dots, \frac{d^k Z}{dt^k}, \dots \quad \dots \quad (2.1)$$

Las ecuaciones diferenciales pueden ser de 1er. orden, de 2o. orden o de orden "p", dependiendo del valor de la derivada más alta:

$$1\text{er. orden: } \frac{dz}{dt} = ky \dots\dots\dots(2.2)$$

$$2\text{o. orden: } \frac{d^2y}{dt^2} + b \frac{dy}{dt} + cy = 0 \dots\dots\dots(2.3)$$

$$y'' + by' + cy \dots\dots\dots(2.4)$$

orden "p": cuando $p \geq 2$.

Las ecuaciones diferenciales describen movimientos oscilatorios cuando el

$$y'' + ky = 0 \dots\dots\dots(2.5)$$

(García Reimbert, 1991).

La solución de las ecuaciones diferenciales está dada por la ecuación característica del proceso (2.20), como se explica más adelante.

$$x = \frac{-a_1 \pm \sqrt{a_1^2 + 4a_2}}{2a_2} \dots\dots\dots*(2.20)$$

* El procedimiento para la solución de las ecuaciones diferenciales puede consultarse en cualquier libro de texto de cálculo diferencial (Boyce y DiPrima, 1977, por ejemplo.).

Ecuaciones en diferencias.-

En los casos de procesos estocásticos que nos interesa estudiar, la variable numérica -tamaño poblacional, o tamaños promedios de tórax- varía en relación al tiempo. Como en nuestros casos la escala de tiempo tiene intervalos de tiempo fijos, no continuos, se trata de ecuaciones en diferencias, ya que las ecuaciones en diferencias son el equivalente discreto de las ecuaciones diferenciales que involucran variables en función del tiempo.

Al observar el comportamiento de la variable Z_t con cierto esparcimiento a través del tiempo, ésta depende "tiempos discretos" (en el cual t sólo toma valores enteros....-2,-1,0,1,2,...). Si se consideran intervalos de tiempo constantes, entonces el comportamiento de la variable Z_t está dictado por sus diferencias ∇ , es decir, por los incrementos ó variaciones de Z_t al pasar de un intervalo al siguiente:

$$\Delta Z_t = Z_{t+1} - Z_t \dots \dots \dots (2.6)$$

Este "operador incremento" es una notación más usual para ecuaciones en diferencia.

Entonces, el equivalente discreto de

$$\frac{dZ}{dt}, \frac{d^2Z}{dt^2}, \dots, \frac{d^k Z}{dt^k} \dots \dots \dots (2.7)$$

es

$$\nabla Z_t, \nabla^2 Z_t, \dots, \nabla^k Z_t \dots \dots \dots (2.8)$$

sin embargo, si t sólo toma valores enteros contiguos, se sabe que $t=1$; por consiguiente Z_t , viene a ser el equivalente de $\frac{dZ}{dt}$ cuando t toma los valores..., $-2,-1,0,1,2,\dots$.

Debido a la relación $\nabla Z_t = \Delta Z_{t-1}$ que liga estrechamente a los operadores incremento y diferencia ($\Delta y \nabla$), se considera que no existe riesgo al emplear ∇Z_t , y a la vez se mantiene el concepto de variable retrasada.

En analogía con las ecuaciones diferenciales, las ecuaciones en diferencia pueden ser lineales o no lineales, si ninguno de sus términos Z aparece elevado a alguna potencia distinta de uno, y de diversos ordenes, según el número de retrasos o diferencias, por ejemplo, las ecuaciones en diferencia se clasifican "de primer orden", "de segundo orden", o "de orden p ", según si intervienen diferencias de orden uno, y por lo tanto a lo mas un solo retraso en Z_t , o bien una ecuación en diferencia de "orden $p > 1$ " en donde pueden intervenir hasta p retrasos, o sea $Z_t, Z_{t-1}, \dots, Z_{t-p}$.

En este trabajo consideraremos ecuaciones en diferencia lineales; es decir polinomios de retraso de diferentes grados (1, 2, "p").

Dado que nuestras variables en estudio son observaciones numéricas con cierto esparciamiento en tiempos discretos, tenemos en realidad procesos estocásticos lineales, es decir, series de tiempo. Ahora bien, en series de tiempo, lo que interesa fundamentalmente conocer es:

1. ¿Cuál es la solución de la ecuación en diferencia que se estudia?.
2. ¿Cuáles son las condiciones para que el proceso representado mediante una ecuación en diferencia, llega a alcanzar eventualmente un punto de equilibrio? Esta pregunta es relevante para el estudio de series de tiempo, porque hay una estrecha relación entre procesos deterministas (cuyo comportamineto está dictado por ecuaciones en diferencia), y series de tiempo que

admiten la representación autorregresiva; de hecho, el concepto de equilibrio eventual para procesos deterministas, está ligado con el concepto de estacionariedad para series de tiempo.

*Ecuaciones en diferencia de primer orden.

Como ejemplo de una de las ecuaciones en diferencia más simples, considérese a la ecuación:

$$Z_t = a_0 + a_1 Z_{t-1} \quad \text{donde } t = \dots, -2, -1, 0, 1, 2, \dots \quad (2.9)$$

Las ecuaciones en diferencia se resuelven por métodos iterativos, en los que como su nombre lo indica, se parte de un valor inicial Z_0 y, por aplicación sucesiva (por iteración), se encuentran los valores para Z_1, Z_2 etcétera. Mediante este proceso, es posible inferir el comportamiento de Z_t en general.

La ecuación de primer orden, cuando se conoce el valor inicial de Z_0 , se resuelve así mediante el método iterativo: inicialmente se tiene:

$$Z_1 = a_0 + a_1 Z_0 \quad \dots \quad (2.10)$$

$$Z_2 = a_0 + a_1 Z_1 = a_0(1 + a_1) + a_1^2 Z_0 \quad \dots \quad (2.11)$$

$$Z_3 = a_0 + a_1 Z_2 = a_0(1 + a_1 + a_1^2) + a_1^3 Z_0 \quad \dots \quad (2.12)$$

de donde, en general se llega a que

$$Z_t = a_0 \sum_{j=0}^{t-1} a_1^j + a_1^t Z_0, \quad \text{para } t \geq 1 \quad \dots \quad (2.13)$$

de aquí si $a_1 \neq 1$, se tienen que

$$\sum_{j=0}^{t-1} a_1^j = \frac{1-a_1^t}{1-a_1} \dots\dots\dots(2.14)$$

***Ecuaciones en diferencia de segundo orden**

A continuación se expone de manera breve cómo se soluciona una ecuación en diferencia de segundo orden, y cuáles son las condiciones para que el correspondiente proceso alcance un equilibrio eventual.

Sea la ecuación en diferencia de segundo orden:

$$(1 - a_1 B - a_2 B^2)Z_t = a_0, \\ t = \dots, -2, -1, 0, 1, 2, \dots; a_2 \neq 0 \\ \dots\dots\dots(2.15)$$

para que esta ecuación sea precisamente de orden 2, requiere que a_2 sea distinto de cero. La solución a esta ecuación viene dada por:

$$Z_t = \frac{a_0}{1 - a_1 - a_2} + s_1 g_1^t + s_2 g_2^t \dots\dots\dots(2.16)$$

donde s_1 y s_2 son constantes que pueden ser determinadas mediante dos condiciones iniciales o colaterales. Los valores g_1 y g_2 están relacionados con los coeficientes a_1 y a_2 de la ecuación general de segundo orden, mediante las ecuaciones

$$a_1 = g_1 + g_2 \quad \text{y} \quad a_2 = -g_1 g_2 \dots\dots\dots(2.17)$$

que surgen de la siguiente factorización:

$$1 - a_1 B - a_2 B^2 = (1 - g_1 B)(1 - g_2 B) \quad \dots\dots\dots (2.18)$$

Es importante determinar los valores de g_1 y g_2 para valores dados de a_1 y a_2 , lo cual se logra con el siguiente argumento:

nótese que la ecuación $(1 - g_1 x)(1 - g_2 x) = 0$ tiene como raíces (o ceros) a los valores de $x_1 = g_1^{-1}$ y $x_2 = g_2^{-1}$; por tanto, dados el polinomio $1 - a_1 B - a_2 B^2$ y la factorización anteriormente mencionada, las raíces g_1^{-1} y g_2^{-1} se encuentran al resolver la ecuación característica del proceso:

$$1 - a_1 x - a_2 x^2 = 0 \quad \dots\dots\dots (2.19)$$

para dos valores de x . Esto es, los valores g_1 y g_2 se obtienen como recíprocos de las raíces dadas por

$$x = \frac{-a_1 \pm \sqrt{a_1^2 + 4a_2}}{2a_2} \quad \dots\dots\dots (2.20)$$

ecuación, esta última, que puede resolverse una vez que se conocen a_1 y a_2 .

Nota: el tamaño de la raíz depende del grado del polinomio; si el polinomio es de segundo orden, las raíces características serán raíces cuadradas; si se trata de un polinomio de orden "p", las raíces características serán de "tamaño "p".

*Ecuaciones en diferencia de orden "p".

Para resolver ecuaciones en diferencia de orden $p \geq 2$, se considerará la ecuación

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p) Z_t = a_0, \quad t = \dots, -2, -1, 0, 1, 2, \dots$$

$$a_p = 0 \dots \dots \dots (2.21)$$

por el Teorema Fundamental del Algebra, se sabe que el polinomio de retraso involucrado se puede escribir como

$$(1 - g_1 B)(1 - g_2 B) \dots (1 - g_p B) = G(B) \dots \dots \dots (2.22)$$

de tal forma que las raíces de la ecuación característica

$$G(x) = 0 \quad \dots \dots \dots (2.23)$$

Para que un proceso descrito por ecuaciones en diferencia de primero y segundo orden, alcance su punto de equilibrio, es necesario que el recíproco del módulo de cada una de las raíces de la ecuación característica sea menor que la unidad (círculo unitario). Esto también es cierto para el caso general (2.21), y su resultado lo describe el Teorema de Shur:

los módulos de las raíces de la ecuación

$$g^p - a_1 g^{p-1} - a_2 g^{p-2} - \dots - a_{p-1} g - a_p = 0 \quad \dots (2.24)$$

serán todos menores que la unidad, si y sólo si los p determinantes que se muestran a continuación son positivos.

$$D_i = \begin{vmatrix} -1 & a_p \\ a_p & -1 \end{vmatrix}$$

$$D_1 = \begin{vmatrix} -1 & 0 & a_p & a_p - 1 \\ a_p - 1 & -1 & 0 & a_p \\ a_p & 0 & -1 & a_p - 1 \\ a_p - 1 & a_p & 0 & -1 \end{vmatrix} \dots$$

$$D_p = \begin{vmatrix} -1 & 0 & \dots & 0 & a_p & a_{p-1} & \dots & a_1 \\ a_1 & -1 & \dots & 0 & 0 & a_p & \dots & a_2 \\ \dots & \dots \\ a_{p-1} & a_{p-2} & \dots & -1 & 0 & 0 & \dots & a_p \\ a_p & 0 & \dots & 0 & -1 & a_1 & \dots & a_{p-1} \\ a_{p-1} & a_p & \dots & 0 & 0 & -1 & \dots & a_{p-2} \\ \dots & \dots \\ a_1 & a_2 & \dots & a_p & 0 & 0 & \dots & 1 \end{vmatrix}$$

.....(2.25)

El grado del polinomio me dá el número de soluciones, y corresponden al número de veces que la curva cruza al eje de las abscisas o el número de veces que cruza a la recta con tendencia.

Para verificar si una ecuación en diferencia de orden p representa a un proceso convergente, se deben calcular los p determinantes (2.25), si todos estos son positivos, podrá concluirse que el proceso tenderá a equilibrarse eventualmente; por el contrario, bastará con que alguno de los determinantes no sea positivo para concluir que el proceso no se estabilizará jamás. (¿Hacia el CAOS?).

Adviértase que para una ecuación en diferencias de primer orden $\nabla z_t = 0$, ($t=1,2,\dots$), se obtiene como solución general una constante, o sea un polinomio de grado cero, y se requiere una condición inicial para encontrar la solución específica. Para la ecuación en diferencias de segundo orden $\nabla^2 z_t = 0$, ($t= 2,3,\dots$), la solución general es una línea recta, es decir, un polinomio de grado 1, y se requieren dos condiciones iniciales para obtener la solución específica. El caso general de la ecuación en diferencias de orden d , $\nabla^d z_t = 0$, tendrá como solución general a un polinomio de grado $d-1$ y requerirá de d condiciones iniciales para especificar completamente la solución, esto es:

Ecuación en diferencias iniciales	solución general	No. de condiciones
$\nabla z_t = 0$,	$z_t = \beta_0$	1
$\nabla^2 z_t = 0$,	$z_t = \beta_0 + \beta_1 t$	2
...
$\nabla^d z_t = 0$	$z_t = \beta_0 + \beta_1 t + \dots + \beta_{d-1} t^{d-1}$	$d \geq 1$

Cabe destacar que las condiciones del problema en estudio, son las importantes para definir finalmente el polinomio que describe a la serie en análisis.

2.5 Modelos de series de tiempo.

El análisis estadístico de datos en series de tiempo consiste en construir un modelo matemático que represente y explique el comportamiento de los datos observados y permita hacer pronósticos. Más que lograr "el mejor ajuste" de modelos preconcebidos a los datos con que se cuenta, se trata de elegir al modelo que los propios datos sugieran. El modelo matemático se construye determinando los valores "p", "d" y "q", ya que dichos valores permiten identificar al posible modelo dentro de la familia de modelos ARIMA (Guerrero, 1991).

La familia de modelos ARIMA ("Autoregressive, Integrated, Moving Average Models") comprende los modelos de promedios móviles (AM), los modelos autorregresivos (AR), los modelos autorregresivos y de promedios móviles (ARMA), y los modelos autorregresivos integrados y de promedios móviles (ARIMA). La familia de modelos ARIMA es bastante general, de manera que prácticamente cualquier serie de tiempo observada puede representarse con alguno de los modelos ARIMA, pero, cabe enfatizar que la elección del modelo debe ser sugerida por los propios datos.

La estrategia para construir modelos ARIMA se presenta en el siguiente capítulo.

CAPITULO 3.

LA METODOLOGIA ESTADISTICA DEL ANALISIS DE SERIES DE TIEMPO UNIVARIADAS: METODOS DE ANALISIS Y CONSTRUCCIÓN DE MODELOS DE SERIES DE TIEMPO.

El Análisis estadístico de una serie de datos numéricos, se realiza en dos etapas. La primera etapa consiste en un enfoque descriptivo, un análisis descriptivo de los datos, y posteriormente la segunda etapa consiste en un enfoque inferencial mediante la construcción del modelo ARIMA.

En este capítulo estudiaremos la primera etapa para analizar los datos de la serie de tiempo en estudio, que consiste en visualizar la información que se tiene. Este enfoque descriptivo, resume y describe de forma concisa la información que se tiene mediante gráficas o a través de unas cuantas medidas descriptivas (Guerrero 1991a).

Hay tres métodos de análisis descriptivos de los datos:

1. Análisis gráfico.
2. Análisis estadístico.
3. Análisis de la estructura de autocorrelación de la serie.

La segunda etapa del análisis de una serie de tiempo, es la construcción propiamente del modelo. Con los datos descritos y visualizados mediante los métodos de análisis mencionados, se procede a construir el modelo. En el capítulo 4 estudiaremos la segunda etapa del análisis estadístico de series de tiempo

la etapa: Enfoque descriptivo:

Métodos de análisis descriptivo de una serie de tiempo.

1. Análisis gráfico de una serie.

Dentro de los elementos descriptivos de una serie de tiempo, las gráficas tienen un lugar prioritario. Al construir la gráfica del proceso en estudio, se verifica visualmente la congruencia de los datos. En el presente estudio, se trabaja con observaciones derivadas de procesos estocásticos cuyo conjunto índice se toma en relación al tiempo. Dichas observaciones se hacen a intervalos con longitud de tiempo fija, por lo que se habla de series de tiempo discretas.

Las series de tiempo discretas se clasifican de la siguiente manera:

Estacionarias...

El concepto de estacionariedad de una serie de tiempo está ligado al concepto de equilibrio estocástico. Es interesante entender cuál es el proceso generador de la serie de tiempo.

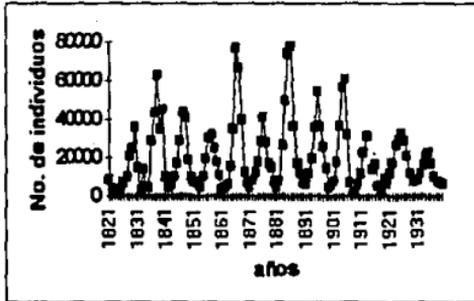
A diferencia de los procesos estocásticos, los procesos **deterministas** son predecibles, su desarrollo está determinado, definido. Es posible conocer desde un principio el destino final del proceso determinista porque "dejan" poca oportunidad a que el resultado final se modifique por el azar. Al dejar caer un objeto, una manzana por ejemplo, se precipitará al suelo por efecto de la ley de la gravedad. Es bastante improbable que dadas las condiciones iniciales, la manzana vuele, si simplemente se le deja caer. El destino final del proceso se puede determinar, su comportamiento obedece a leyes específicas.

El comportamiento de los procesos deterministas, está definido por sistemas de ecuaciones diferenciales. Con ecuaciones en diferencias se pueden representar procesos que eventualmente alcanzan un punto de equilibrio, dadas ciertas condiciones, como ya hemos visto en el capítulo anterior.

El concepto de equilibrio eventual para procesos deterministas está ligado al concepto de estacionariedad o equilibrio estocástico de las series de tiempo.

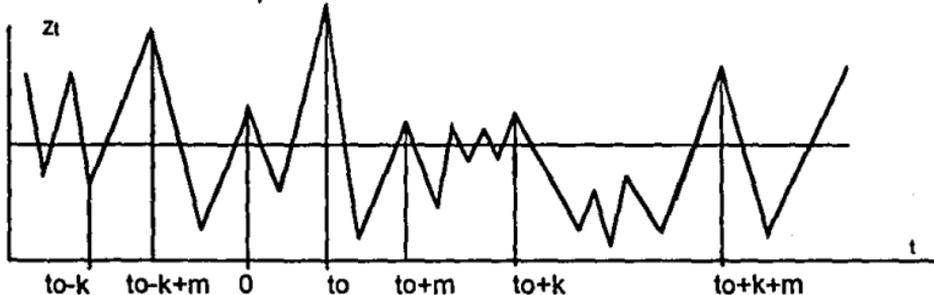
La realización de un proceso estocástico estacionario se ve así:

Z_t



Las fluctuaciones aleatorias del proceso siempre ocurren alrededor de un valor promedio: el punto de equilibrio ó EQUILIBRIO ESTOCÁSTICO. Este valor promedio es la media o "el nivel" del proceso, y no depende del tiempo, aunque el proceso se aleje de la media durante un cierto período, t eventualmente regresa a una vecindad de la misma.

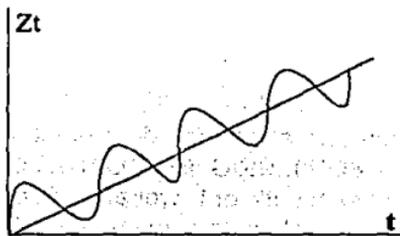
una serie de tiempo estacionaria se ve así:



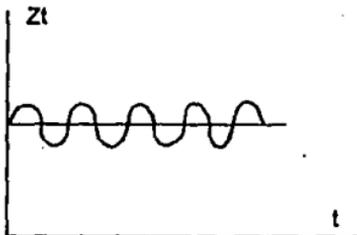
En este caso, los valores de la variable Z_t también oscilan alrededor de un nivel o valor promedio μ (media), con una cierta varianza γ_0 que define los límites de las vecindades en las que oscila la serie.

Estacionales...

Las series de tiempo y los procesos se llaman estacionales, cuando presentan un comportamiento cíclico, es decir, que se repite por estaciones. Las series cíclicas o estacionales pueden presentar o no una tendencia:



Serie estacional con tendencia.



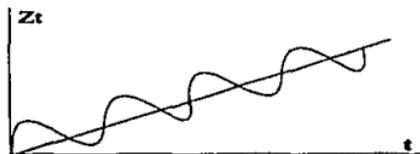
Serie estacional (cíclica), sin tendencia.

Con tendencia...

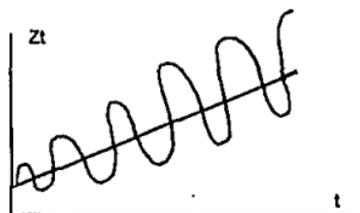
Se habla de una serie de tiempo con tendencia o de series de tiempo no-estacionarias, cuando "el nivel" de la serie se modifica constantemente, es decir, el proceso tiene una "media ponderada" w . En estas series, hay una no-estacionariedad homogénea.

En los procesos no estacionarios el proceso puede tener media constante, pero la varianza y las autocovarianzas dependen del tiempo (Guerrero 1991a)

Series de tiempo con tendencia:



con componentes aditivos

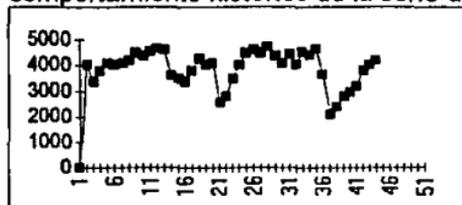


con componentes multiplicativos

La tendencia en las series de tiempo se puede suprimir aplicándole a los datos originales el operador diferencia " ∇ " un número apropiado de veces, como se explicará posteriormente.

Con intervenciones...

Las intervenciones son eventos exógenos que alteran el comportamiento histórico de la serie de tiempo.



Los eventos exógenos modifican el nivel de la serie y en ocasiones no sólo su tendencia, sino su desarrollo histórico.

2. Análisis estadístico de una serie.

Para poder caracterizar completamente a un proceso estocástico, habría que conocer la función de densidad conjunta de todas las variables aleatorias involucradas. Esto es poco práctico y muy complicado de lograr con una serie de tiempo. Sin embargo, los primeros momentos de las variables aleatorias resumen bastante bien su distribución. Los momentos de primero y de segundo orden, es decir la media, la varianza y la covarianza de las variables involucradas en el proceso estocástico resumen, y describen de manera concisa la distribución de la serie (Degroot 1988).

Para Z_t , su media μ_t es

$$E(Z_t) = \mu_t \quad \dots\dots\dots(3.1)$$

Si al proceso $\{Z_t\}$ lo expresamos en función de $\{a_t\}$ mediante la relación lineal

$$Z_t = \mu + a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \dots \quad \dots\dots\dots(3.2)$$

$$Z_t = \mu + \psi(B)a_t \quad \dots\dots\dots(3.3)$$

entonces se tiene

$$\mu_t = \mu + E(a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \dots) \quad \dots\dots\dots(3.4)$$

Si a la varianza del proceso Z_t la denotamos por γ_0

$$\gamma_0 = E[(Z_t - \mu)]^2 \quad \dots\dots\dots(3.5)$$

$$\gamma_0 = E[(a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \dots)]^2 \quad \dots\dots\dots(3.6)$$

$$\gamma_0 = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2 \quad \dots\dots\dots(3.7)$$

como la varianza de una suma es la suma de las varianzas, siempre y cuando no estén correlacionadas, se tienen que

$$E(\psi_i a_{t-i} \cdot \psi_j a_{t-j}) = \psi_i E(a_{t-i}) \cdot \psi_j E(a_{t-j}) = 0 \quad \dots\dots\dots(3.8)$$

para toda $i = j$

La covarianza del proceso Z_t se denota por

$$\gamma_k = \text{Cov}(Z_t, Z_{t+k}), \quad \dots\dots\dots(3.9)$$

pero dado que la serie de tiempo es una muestra del proceso estocástico, es necesario considerar a todas las variables aleatorias involucradas en el proceso, esto es:

$$Z_1, Z_2, \dots, Z_n. \quad \dots\dots\dots(3.10)$$

entonces la covarianza :

$$\gamma_k = E[(Z_t - \mu)(Z_{t-k} - \mu)] \quad \dots\dots\dots(3.11)$$

A los procesos cuyos momentos de primero y segundo orden no dependen del tiempo, se les denomina procesos estacionarios de segundo orden. Y los procesos estocásticos son estrictamente estacionarios, si la función de densidad para un conjunto arbitrario de variables $(Z_t, Z_{t+1}, \dots, Z_{t+m})$ es invariante respecto a desplazamientos en el tiempo.

En la práctica es común suponer que la distribución asociada con las series de tiempo es normal, y se considera que es suficiente conocer la media y la función de autocovarianza $\{\gamma_k\}$ para caracterizar

completamente a una serie de tiempo estacionaria. Sin embargo, es preferible trabajar con las autocorrelaciones $\rho_0, \rho_1, \rho_2, \dots$, definidas a través de

$$\rho_k = \frac{E[(Z_t - m)(Z_{t+k} - m)]}{E(Z_t - m)^2} = \frac{\gamma_k}{\gamma_0}, k = 0, \pm 1, \pm 2, \dots \quad \text{.....(3.12)}$$

las cuales generan a la función de autocorrelación $\{\rho_k\}$

La autocorrelación, es la relación que se presenta entre dos estados de cualquier proceso, en los que ocurre una transición al azar. En nuestra serie de lince por ejemplo, la autocorrelación es la relación que guarda el tamaño de la población de "éste año", y el tamaño que tenía la población el año pasado. En este caso, la transición entre dos estados es la variación del tamaño de la población de un año al otro, y hay una relación entre el número de pieles que se consiguieron cada año ("tamaño de la población"), porque el tamaño de la población en un momento determinado, depende del tamaño que tenía la población el año anterior, y será determinante en el tamaño de la población al año siguiente.

Ahora bien, la función de autocorrelación sirve para especificar junto con la μ y la σ^2 , a un proceso estacionario. Dado que una serie de tiempo es una muestra de las posibles realizaciones del proceso estocástico, la estimación de la función de autocorrelación $\{\gamma_k\}$ del proceso a partir de la serie de tiempo, se logra suponiendo que el proceso estacionario en estudio posee ciertas propiedades ergódicas que permiten la equivalencia entre los valores esperados y los promedios muestrales obtenidos de una realización suficientemente larga del proceso. (Guerrero 1991a). De la propiedad ergódica se deriva que la media del proceso puede estimarse como la media muestral de la serie observada, y con base en esta media muestral de la serie y con los datos se obtiene la varianza muestral y las autocovarianzas

El cociente $r_1 = \frac{c_1}{c_2}$ (3.13)

nos dá el primer valor de la función de autocorrelación muestral $\{ \rho_k \}$

$$r_2 = \frac{c_1}{c_2} \text{(3.14)}$$

es el segundo valor de la función de autocorrelación muestral $\{ k \}$.

Este procedimiento se realiza para todas las observaciones de la serie de tiempo, y los valores obtenidos de toda la función de autocorrelación de la muestra, se presentan en una gráfica llamada "EL AUTOCORRELOGRAMA ".

Nota:

El Teorema de la media ergódica define que entre más se alejan los valores de la variable del valor inicial de "x", más tienden a tomar el valor de una media promedio en el tiempo, y a no estar correlacionados. Ergodicidad implica una conducta no caótica.

Un proceso no ergódico es aquel cuyos parámetros no pueden ser estimados precisamente, a partir de una sola serie muestral (Diggle 1990). En un proceso estocástico ergódico, si se tuvieran todas sus posibles realizaciones, se tendría toda una tendencia armónica y así sería posible hacer inferencias sobre la varianza de los parámetros que definen al proceso.

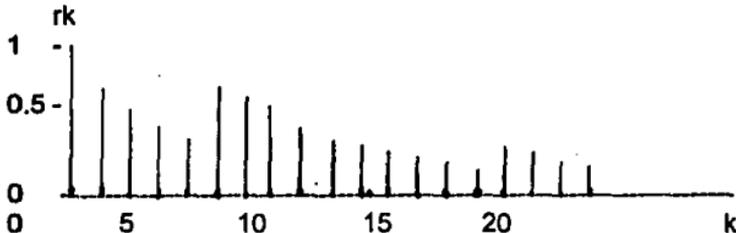
3. Análisis de la estructura de autocorrelación de la serie

El autocorrelograma.

En el análisis de series de tiempo, las observaciones que se tienen NO provienen de variables aleatorias independientes, en contraste con prácticamente cualquier análisis estadístico donde con el conocimiento de las funciones de densidad individuales se puede obtener la función de densidad conjunta.

En series de tiempo se supone que existe toda una estructura de correlación entre las observaciones, cuyos valores se pueden estimar mediante la función de autocorrelación de la muestra.

Como ya se mencionó, el autocorrelograma es la gráfica de los valores obtenidos de la función de autocorrelación de la muestra:



Función de autocorrelación muestral para una serie de tiempo.

El autocorrelograma habla del desarrollo del proceso en términos de como se relaciona cada una de las observaciones de la muestra, con la varianza muestral σ^2 .

Cada modelo de la familia ARIMA, tiene autocorrelogramas típicos, de manera que el análisis del gráfico permite identificar a la serie de datos observados dentro de alguno de los grupos de los modelos ARIMA.

CAPITULO 4

LA METODOLOGIA ESTADISTICA DEL ANALISIS DE SERIES DE TIEMPO UNIVARIADAS: METODOS DE ANALISIS Y CONSTRUCCIÓN DE MODELOS DE SERIES DE TIEMPO.

Segunda etapa:

Construcción del modelo ARIMA para la serie de tiempo.

La segunda etapa del análisis estadístico de las series de tiempo, consiste en construir el modelo matemático que describa las observaciones y permita hacer inferencias, es decir pronósticos sobre el fenómeno o proceso en estudio.

La estrategia para la construcción de modelos ARIMA a partir de las observaciones de la serie de tiempo muestra, se resumen en el cuadro de la página siguiente: →

Cuadro 1: METODOLOGIA

La estrategia de construcción de modelos para series de tiempo consta de cuatro etapas fundamentales (Box y Jenkins 1970), (Guerrero 1991a): →

cuadro 1:

serie de tiempo
(datos)



1 IDENTIFICACION DEL MODELO

De acuerdo con lo que los datos indiquen, se identifica al modelo dentro de las clases de modelos: AM, AR, ARMA ó ARIMA, mediante la determinación de los valores "p", "d" y "q".



2. ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

En esta etapa es donde se construye propiamente el modelo. Una vez conocidos los ordenes de los polinomios ("p" para los autorregresivos AR; orden "q" para los de promedios móviles, y "p", "q" para los ARMA), así como el grado de diferenciación "d" para los ARIMA, se postula que el modelo:



$$(B) T(Z_t) = \alpha + (B) a_t$$

resulta el adecuado para representar a la serie $\{T(Z_t)\}$, (si la serie se transformó). La asignación de los valores a $1, \dots, p$ y $1, \dots, q$, se hace por el método de máxima verosimilitud, maximizando la función de verosimilitud de los parámetros para obtener la representación



$$(B) T(Z_t) = \alpha + (B) a_t$$

mas apropiada de la serie $\{W_t\}$ en estudio.



3. VERIFICACIÓN DE SUPUESTOS

¿Es un buen modelo? ¿Se cumplen los supuestos básicos implícitos en el modelo? ¿Ajusta y describe bien los datos?

Análisis de los Residuos
Pruebas de Bondad del Ajuste



* Si no se cumplen los supuestos, se determinan las modificaciones necesarias, se repiten las etapas anteriores hasta que la verificación indique resultados aceptables.



CORREGIR



4. USO DEL MODELO

Para los fines para los que el investigador lo construyó: Pronóstico, control, simulación o explicación del fenómeno en estudio.

Polinomios de retraso.

Todos los modelos de la familia ARIMA para representar los procesos estocásticos, se basan en el uso operadores de retraso en forma de polinomios. El operador de retraso más sencillo es el operador B ("backward") definido por la relación :

$$BZ_t = Z_{t-1} \quad \text{para toda } t \quad \dots\dots\dots(4.2)$$

así, se puede ver el proceso Z_t en momentos anteriores ...

$$B^2Z_t = B(BZ_t) = Z_{t-2} \quad \dots\dots\dots(4.3)$$

$$B^3Z_t = B(B^2Z_t) = Z_{t-3} \quad \dots\dots\dots(4.4)$$

$$B^kZ_t = B(B^{k-1}Z_t) = Z_{t-k} \quad \dots\dots\dots(4.5)$$

y en general: $B^kZ_t = Z_{t-k}$,

$$\text{para } k=0,1,2,\dots \text{ y toda } t \quad \dots\dots\dots(4.6)$$

es decir, al "multiplicar" a B^k por Z_t se obtiene la variable retrasada k periodos, y debido a que $B^0Z_t = Z_t$, estrictamente debería escribirse $B^0 = I$, donde I = operador identidad que deja intacta a la variable.

4.1. Identificación del modelo ARIMA.

La familia de modelos ARIMA comprende cuatro tipos de modelos: Autorregresivos (AR), de promedios móviles (AM), una combinación de éstos dos, los autorregresivos de promedios móviles (ARMA), y los autorregresivos integrados y de promedios móviles (ARIMA). Sin embargo, los modelos no son válidos para series no estacionarias, por tanto si se tiene una serie no estacionaria, es necesario hacerla estacionaria como se verá más adelante.

4.1.1.- Modelos autorregresivos (AR)

La de este tipo de modelos es expresar al valor numérico del proceso Z_t , al tiempo t , como una variable dependiente, pero a diferencia del modelo de regresión, la variable dependiente, no "depende" de los valores de la variable independiente, sino de sus propios valores observados en períodos anteriores a " t ", ponderados por coeficientes autorregresivos $\phi_1, \phi_2, \dots, \phi_p$.

De esta manera, se expresa al proceso estocástico mediante la relación lineal

$$(Z_t - \mu)\phi B = \{a_t\} \quad \dots\dots\dots(4.7)$$

donde

μ = denota la media de la serie (ó el "nivel" del proceso)
 $(Z_t - \mu)$ = son las desviaciones del proceso Z_t con respecto a la media

ϕ = son los parámetros o coeficientes autorregresivos.

(B) = el operador de retraso

$\{a_t\}$ = el proceso aún como "ruido"

Así, los modelos autorregresivos se definen como

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad Z_t - \mu = a_t \quad \dots\dots\dots(4.8)$$

Modelos autorregresivos de primer orden (AR1)

Los modelos autorregresivos se denominan "de primer orden", "de segundo orden", ó "de orden p" ("AR"1, "AR2, o "AR"p), dependiendo del número de parámetros ϕ .

Los procesos autorregresivos podrán ser estacionarios o no estacionarios, es decir alcanzarán una situación de equilibrio o no, dependiendo de los valores que tomen las raíces de la ecuación característica

$$\phi(x) = 0. \quad \dots\dots\dots(4.9)$$

De hecho, para que Z_t sea estacionario, los valores de ϕ deben estar dentro del círculo unitario.



Las raíces de la ecuación característica sirven para indicar el comportamiento dinámico de la función de autocorrelación.

Si al resolver la ecuación característica (2.20) (*en este caso a X_1 Y A X_2 les llamamos ϕ), se encuentra que los valores de ϕ están dentro del círculo unidad:

$$-1 < \phi < 1 \quad \dots\dots\dots(4.10)$$

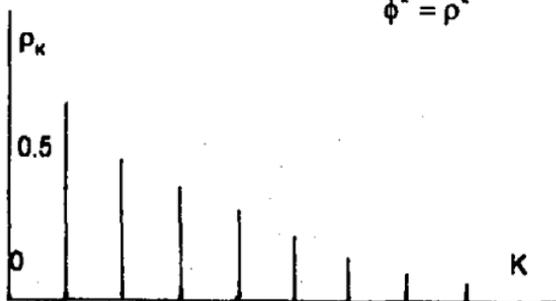
entonces el proceso es estacionario.

En los procesos autorregresivos de primer orden, la función de autocorrelación ACF decrece exponencialmente, porque

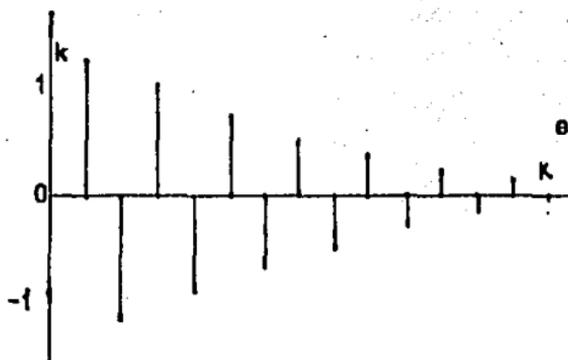
$$\phi^k = \rho^k \quad \dots\dots\dots(4.11)$$

si $\phi > 0$

$\phi(+)$ → ACF decrece exponencialmente



pero si $\phi < 0...$



$\phi(-) \rightarrow$ ACF decrece
exponencialmente, alternando
signos

Modelos AR de segundo orden

El proceso AR2 está definido mediante:

$$(1 - \phi_1 B - \phi_2 B^2) \bar{Z}_t = a_t \quad \dots\dots\dots(4.12)$$

*Nótese que el grado del polinomio y el orden de la ecuación característica son iguales.

Para que este proceso sea estacionario, debe cumplirse que la suma de las raíces de la ecuación característica

$$1 - bX - aX^2 = 0 \Rightarrow$$

$$1 - \phi_1 X - \phi_2 X^2 = 0 \quad \dots\dots\dots(4.13)$$

se encuentren fuera del círculo unitario.

La estimación de los parámetros ϕ_1 y ϕ_2 , se obtiene de las ecuaciones de Yule-Walter:

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \quad \dots\dots\dots(4.14)$$

El ACF de los modelos AR2, en general tiene un comportamiento similar al ACF de los modelos AR1, porque todas las autocorrelaciones decaen exponencialmente a cero si ϕ es positiva, y decrecen exponencialmente alternando signos si ϕ es negativa.

Modelos autorregresivos de orden "p". (ARp)

Este proceso se describe mediante la ecuación

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \phi_2 \bar{z}_{t-2} + \dots + \phi_p \bar{z}_{t-p} + a_t \quad \dots\dots\dots(4.15)$$

En este caso, el polinomio característico es de grado "p", y el valor x_n está relacionado con los x_{n-p} valores anteriores.

Para que este proceso sea estacionario, es necesario que todas las raíces de la ecuación característica

$$1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0 \quad \dots\dots\dots(4.16)$$

se encuentren fuera del círculo unitario.

Para estimar el número de los "p" parámetros ϕ involucrados en el modelo, se aplica el principio de parsimonia, que consiste en ir eliminando parámetros lo más que se pueda, sobretodo si la información que se tiene del fenómeno es la misma.

En general el autocomelograma de los modelos autoregresivos, presenta una caída exponencial de los valores de las autocorrelaciones, así que si se observa una laaaarga caída exponencial o un comportamiento sinusoidal, es decir, alternando signos, que no se corta. El autocomelograma identifica a un modelo AR. El orden del modelo se determina mediante el valor "p", pero en general el orden de un proceso AR(p), no se puede detectar con el solo uso del autocomelograma ACF de la muestra, se

requiere de otro instrumento estadístico: la Función de Autocorrelación Parcial FACP, la cual adquiere características determinadas dependiendo del orden del proceso y de los parámetros involucrados.

Estas características se resúmen en el cuadro 1 de la página 68

4.1.2. Modelos de promedios móviles MA ("moving average").

Existen procesos que se representan mejor en términos de choques aleatorios, que en términos de autorregresiones. A este tipo de procesos se les denomina como "procesos de promedios móviles" y se les denota por MA.

Se trata un proceso estocástico $\{Z_t\}$, cuyos valores pueden ser dependientes unos de otros, como una suma finita ponderada de choques aleatorios independientes $\{a_t\}$, es decir

$$Z_t - \mu = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \dots\dots\dots(4.17)$$

De manera compacta se denotan como

$$Z_t - \mu = \theta(B) a_t \dots\dots\dots(4.18)$$

en donde

Z_t = es el valor del proceso estocástico al tiempo t .

μ = es la media de la serie, es el "nivel" de la serie. Su valor es constante

θ = es el promedio móvil

$\theta_1, \theta_2, \dots, \theta_q$ = son las ponderaciones, (los parámetros de los promedios móviles) asociadas con los choques aleatorios en los periodos $t-1, t-2, \dots, t-q$, respectivamente.

(B) = es el operador de retraso.

$\{a_t\}$ = es una sucesión de variables aleatorias del proceso, es una suma finita ponderada de choques aleatorios independientes. Es el "ruido blanco" del proceso.

σ_a^2 = es la varianza del ruido blanco

Todo proceso MA es estacionario porque se trata de un proceso lineal general, cuya suma de parámetros $|\theta_i| = \sum_{i=1}^q$ es finita, pues considera a un número finito de sumandos (si θ_i fuese infinito, implicaría que el proceso Z_t es infinito).

Los procesos MA, son procesos que se encuentran en equilibrio, las fluctuaciones alrededor del punto de equilibrio, $\{Z_t\}$, son causadas por choques aleatorios, esto es, choques asociados con eventos inesperados. Tales choques no se asimilan de manera instantánea, sino que pueden seguir causando efectos aún después de transcurrido un cierto número de periodos y además, la intensidad del choque se refleja en el valor de su ponderación θ_i (Guerrero V. 1991).

Se trata de una serie de observaciones no correlacionadas.

Modelos de promedios móviles de orden uno (MA1)

Este proceso es el más simple, y se expresa como

$$\tilde{Z}_t = (1 - \theta B)a_t \quad \dots\dots\dots(4.19)$$

su media vale 0: $E(\tilde{Z}_t) = 0, \mu = 0 \quad \dots\dots\dots(4.20)$

y su varianza es $\gamma_0 = \text{Var}(\tilde{Z}_t) = \sigma^2 a(1 + \theta^2) \quad \dots\dots\dots(4.21)$

las covarianzas están dadas por

$$\gamma_k = E[(a_t - \theta a_{t-1} - 1)(a_{t-k} - \theta a_{t-k-1})] \quad \dots\dots\dots(4.22)$$

donde...

$$\gamma_k = \begin{cases} -\theta\sigma^2 a & \text{si } k=1 \\ 0 & \text{si } k \geq 2 \end{cases}$$

.....(4.23)

de ahí que la función de autocorrelación de MA1 sea

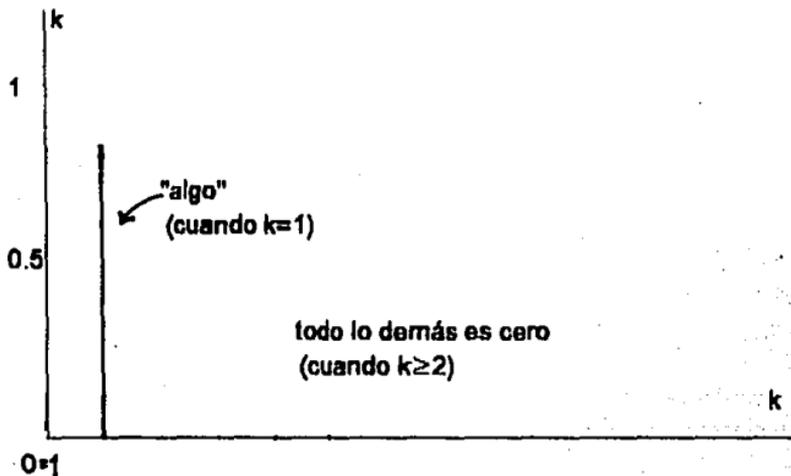
$$\rho_k = \begin{cases} \frac{\theta}{1+\theta^2} & \text{si } k = 1 \\ 0 & \text{si } k \geq 2 \end{cases}$$

si $k \geq 2$ es decir 0 si $k=2,3,4,\dots$ etc.

.....(4.24)

El hecho de que las autocorrelaciones para retrasos mayores que un período sean cero, indica que el proceso MA 1 "no recuerda" más allá de lo ocurrido en el período anterior.

El autocorrelograma de un proceso móvil de orden 1, es un clásico...



Modelos de promedios móviles de orden 2 (MA2)

Su modelo general es

$$\bar{Z}_t = (1 - \theta_1 B - \theta_2 B^2) a_t \quad \dots\dots\dots(4.25)$$

En la gráfica de su función de autocorrelación, se ve que la función se recorre:



Modelos de promedios móviles de orden "q" (MAq)

De manera general,

$$\bar{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad \dots\dots(4.26)$$

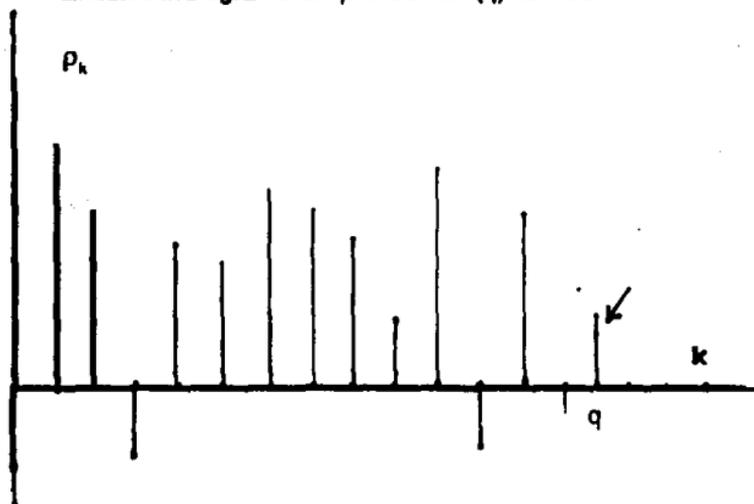
con $\bar{Z}_t = Z_t - \mu$, en donde μ es el nivel del proceso, $\theta_1, \theta_2, \dots, \theta_q$ son los parámetros de promedios móviles del proceso, $\{a_t\}$ es un proceso de ruido blanco.

Como se trata de un proceso estacionario, ni la media, ni la varianza, ni las covarianzas del proceso MA(q) dependen del tiempo.

La función de autocorrelación del proceso MA(q) es

$$\rho_k = \begin{cases} \neq 0 & \text{para toda } k = 1, 2, 3, \dots, q \\ = 0 & \text{para toda } k > q \end{cases}$$

El autocorrelograma del proceso MA(q) se ve así:



Por esto se dice que cuando un autocorrelograma presenta un corte, ése corte es el valor de q , y si a partir de ése valor q , todo vale 0, entonces tienes un modelo AM (ó MA) simple, sin estacionalidades y sin tendencia.

4.1.3.-Modelos autorregresivos y de promedios móviles (ARMA)

Estos modelos consisten en una generalización de los modelos autorregresivos AR y de promedios móviles MA, que consiste en combinar ambas clases de modelos (ARMA).

Su ecuación general es:

$$\phi(B)\tilde{Z}_t = \theta(B)a_t \quad \dots(4.27)$$

en donde

$\phi(B)$ y $\theta(B)$ son polinomios de retraso de orden p y q respectivamente.

$\phi(B)$ es el operador autorregresivo

$\theta(B)$ es el operador de promedio móvil

$\{a_t\}$ es un proceso de ruido blanco

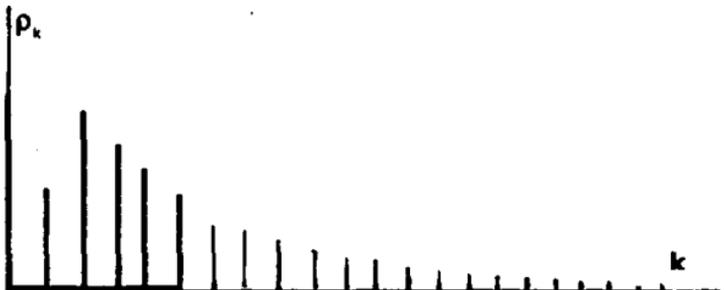
\tilde{Z}_t es la serie de desviaciones de la variable Z_t , respecto a su nivel μ .

Modelos ARMA (1,1)

Es el modelo ARMA más sencillo. Está definido por

$$(1 - \phi B)\tilde{Z}_t = (1 - \theta B)a_t \quad \dots\dots\dots(4.28)$$

En general, en los modelos ARMA (1,1), el comportamiento exponencial se dá a partir del 2o. valor de la función de autocorrelación, por lo que su autocorrelograma se ve así:



Modelos ARMA (p,q)

El caso general de los procesos ARMA (p,q), tiene la ecuación:

$$\phi(B)\bar{Z}_t = \theta(B)a_t \quad \dots(4.29)$$

En donde los polinomios $\phi(B)$ y $\theta(B)$ son de orden p y q respectivamente, es decir,

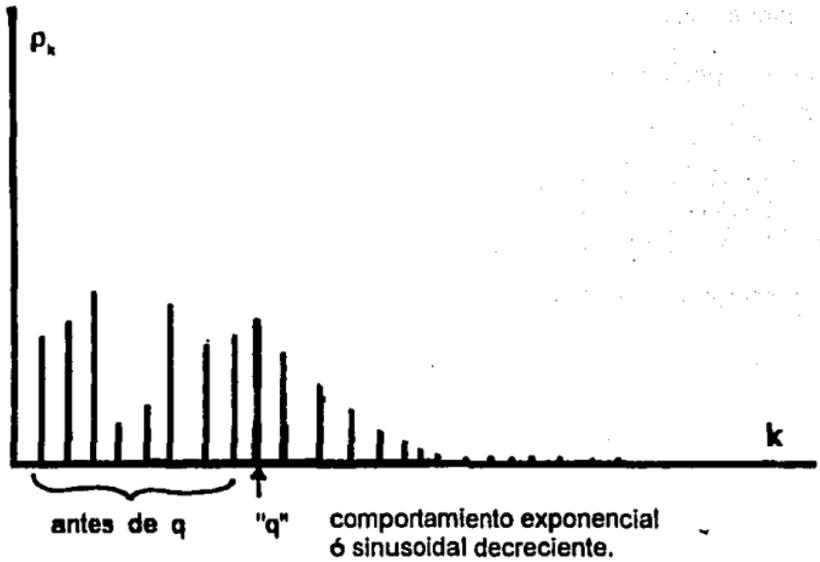
$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} - \dots - \phi_p \bar{Z}_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad \dots\dots\dots(4.30)$$

Para que este proceso sea estacionario, es necesario que las raíces del polinomio característico $\phi(X) = 0$, estén fuera del círculo unitario, y para que sea invertible la condición es que si las raíces de la ecuación $\phi(X) = 0$, se encuentren también fuera del círculo unitario

Un proceso es invertible, si se puede escribir razonablemente en términos de su pasado, es decir, viendo su historia, y que con ésta información, se puedan hacer pronósticos. Como ya se dijo, los procesos ARMA son invertibles, si las raíces de la ecuación $\phi(X) = 0$, se encuentran fuera del círculo unitario.

En los modelos ARMA(p,q), la gráfica de la función de autocorrelación, presenta un patrón diferente antes de encontrar el valor "q", y a partir de dicho valor "q", presentan un comportamiento exponencial ó sinusoidal decreciente. Esto es debido a la parte AM y a la parte AR de los modelos ARMA, como se puede apreciar en la gráfica de un autocorrelograma ARMA típico:

Autocorrelograma del modelo ARMA típico.



4.1.4.- Modelos autorregresivos e integrados de promedios móviles (ARIMA).

En este caso tenemos procesos que **no son estacionarios**. La serie de tiempo no es estacionaria cuando su comportamiento Z_t , depende del tiempo, es decir, conforme el tiempo pasa su tendencia se modifica.

Si la serie no es estacionaria, los modelos que se le ajusten no son válidos. Entonces, la serie Z_t en sí, no es analizable.

si Z_1, Z_2, \dots, Z_n es una serie no estacionaria

Es necesario hacerle diferencias, ∇ y el análisis se realiza en la serie de diferencias resultante:

$$\text{si } X_t = \nabla^d Z_t \quad \text{para toda } t \quad \dots\dots\dots(4.31)$$

entonces X_1, X_2, \dots, X_n esta serie si es estacionaria

$$X_t = (1-B)^d Z_t \quad \dots\dots\dots(4.32)$$

$$n = N-d$$

con cada diferencia se pierde una observación de la serie de datos originales

Entonces, cuando se tienen series no estacionarias, es decir, series con tendencia, se hacen estacionarias aplicándoles el operador diferencia ∇^d .

Así, se tiene un modelo ARMA para la serie de las X_s , pero en realidad se tiene un modelo ARIMA para la serie de datos original Z_t :

$$\phi_p(B)X_t = \theta_q(B)a_t \quad \text{modelo ARMA para las } X_s \quad \dots(4.33)$$

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \quad \text{modelo ARIMA } (p,d,q) \quad \dots$$

para la serie original Z_t

...(4.34)

Entonces.

1. Se hacen transformaciones para estabilizar la varianza.

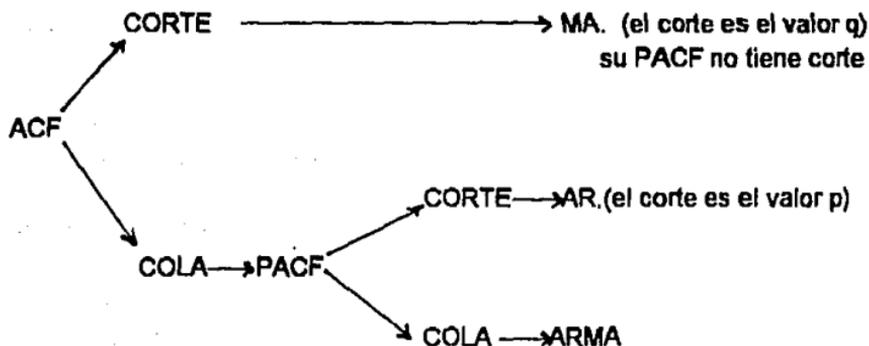
2. Se hacen diferencias de logaritmos para quitar tendencias.
(Desarrollo de Taylor, apuntes de clase).

Finalmente, es necesario mencionar que la sobrediferenciación acarrea problemas en la interpretación de los resultados y además se pierden observaciones innecesariamente.

Cuadro 2:

En suma, para identificar al modelo, hay que fijarse primero en el autocorrelograma ...¿tiene corte o tiene una cola?

Después se va a la función de autocorrelación parcial (PACF)



Es conveniente recordar el comportamiento de la ACF y de la PACF para una serie estacionaria $W_t = T(Z_t)$. En palabras del Dr. Guerrero:

Cuadro 3:
Comportamientos típicos de la FAC y la FACP (ACF Y PACF)

Proceso	FAC	FACP
AR (p)	Convergencia a cero, con comportamiento dictado por la ecuación para (tienen cola o son sinusoidales)	Solamente las primeras p autocorrelaciones parciales son distintas de cero. (su FACP corta en p)
MA(q)	Sólo las primeras q autocorrelaciones son distintas de cero.	Sucesión infinita convergente a cero.
ARMA (p,q)	Comportamiento irregular de las primeras q autocorrelaciones y después convergen a cero de acuerdo con	Sucesión infinita convergente a cero.

En el anexo 1 que se presenta al final de este trabajo, hay mas información sobre la identificación de Modelos ARMA de acuerdo al comportamiento de la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP).

4.2 Estimación de los parámetros del modelo.

En esta etapa, se presupone que ya se han encontrado los parámetros del modelo, y de ser éste el adecuado, lo que hace falta es conocer los valores de los parámetros para que dicho modelo represente apropiadamente la serie en consideración. Es decir, una vez conocidos los órdenes de los polinomios autorregresivo y de promedios móviles p y q , así como el grado de diferenciación d , (porque para estimar a los parámetros del modelo, hay que tomar a la serie pero estacionaria), se postula que el modelo

$$\phi(B)\nabla^d T(Z_t) = \theta_0 + \theta(B)a_t \quad \dots\dots\dots(4.35)$$

resulta adecuado para representar a la serie $\{T(Z_t)\}$ y se requiere entonces asignar valores a $\phi_1, \dots, \phi_p, \theta_0, \theta_1, \dots, \theta_q$. Un método objetivo y estadísticamente apropiado, es el método de "máxima verosimilitud", que se puede consultar en el anexo al final de este trabajo.

Redundancia de parámetros.-

Un aspecto que es necesario mencionar porque puede ser causa de problemas durante la estimación de los parámetros, es la "redundancia de parámetros". Esto es, hay factores aproximadamente comunes en los polinomios autorregresivos y de promedios móviles; por ejemplo:

$$(1 - 0.3B - 0.28B^2)W_t = (1 + 0.4B)a_t \quad \dots\dots(4.36)$$

puede escribirse de manera equivalente como:

$$(1 - 0.7B)W_t = a_t \quad \dots\dots(4.37)$$

que es una representación más parsimoniosa del proceso, y permite además estimar con mayor precisión al parámetro autorregresivo,

cosa que no permite el primer modelo, porque cualquier cambio en la parte autorregresiva puede compensarse con un cambio en la parte de promedios móviles y esto provocaría inestabilidad en los parámetros estimados. Las verificaciones de los supuestos de parsimonia y de estabilidad de los parámetros del modelo, se exponen mas adelante en el siguiente inciso de las pruebas de bondad del ajuste.

En las series estacionarias que se analizan en este trabajo, se ejemplifica esta situación y se discute posteriormente la elección mas parsimoniosa del modelo.

En suma, el resultado de la etapa de estimación es un modelo, con sus parámetros estimados, es decir, ahora se conocen los valores de los parámetros del modelo. Entonces, los resultados de esta etapa son:

- 1.- Parámetros estimados, es decir, los valores de: $\hat{\phi}_i$ y $\hat{\theta}_i$
- 2.- La varianzas $\hat{\sigma}_{\hat{\phi}_i}$ y $\hat{\sigma}_{\hat{\theta}_i}$: Básicamente se estima al 1.96, para ver si es significativo o no.
- 3.-La varianza del ruido, $\hat{\sigma}_a^2$. Es la varianza de lo que queda sin explicar por el modelo, es decir, es la varianza de los residuales.

Conviene recordar, que a mis datos los quiero explicar con:

$$\text{DATOS} = 1 \text{ modelo} + \text{Ruido.}$$

Ahora bien, del "ruido" {a} espero que

- *- su media sea cero.
- *- que los "i" elementos no estén correlacionados.
- *- que la varianza del ruido {a} sea pequeña.

Algunos paquetes, dan estos resultados en términos de " R^2 ":
 R^2 es:

$$R^2 = \frac{SCR}{SCT} \quad \text{.....(4.38)}$$

donde:

SCR = la suma de cuadrados de la regresión
 (es la información que sí explica el modelo)

SCT = es la suma de cuadrados totales,

pero, la suma de cuadrados totales SCT:

SCT= Suma de cuadrados del error + suma de cuadrados de la regresión:

$$SCT = SCE + SCR$$

por lo que

$$SCR = SCT - SCE$$

por lo tanto, si

$$R^2 = \frac{SCR}{SCT}; = \frac{SCT - SCE}{SCT} = \frac{SCT}{SCT} - \frac{SCE}{SCT} = \frac{SCE}{SCT} - 1 \quad \text{.....(4.38)}$$

así

$$R^2 = 1 - \frac{SCE}{SCT} \quad \text{.....(4.39)}$$

Ahora, dijimos que $SCT = SCE + SCR$. Sin embargo, hay casos en que la suma de cuadrados de la regresión puede no significar nada, y entonces la suma de cuadrados totales es igual a la suma de cuadrados del error: $SCT = SCE$,

si

$$SCTotales = \sum (z_t - \hat{z}_t)^2 \quad \dots(4.40)$$

y $SCT = SCE$

entonces

$$SCE = \sum (z_t - \hat{z}_t)^2 \quad \dots(4.41)$$

La suma de cuadrados del error SCE = La suma de los ruidos.

$$\sum (z_t - \hat{z}_t)^2 = \sum a_t^2 \quad \dots(4.42)$$

Distribución normal del ruido blanco.-

Otro aspecto que resulta importante mencionar respecto a la estimación máximo verosímil, es el supuesto de distribución normal para el ruido blanco $\{a_t\}$.

Recordemos que a mis datos los quiero explicar mediante un modelo + ruido:

$$DATOS = 1 \text{ modelo} + \text{Ruido.}$$

El ruido blanco, es la parte de la información contenida en los datos, que el modelo no alcanza a explicar. El ruido blanco es la parte "residual" de las observaciones que no puede ser explicada por el modelo, y se conoce justamente como "los residuales o los residuos".

El análisis y la evaluación de esta parte de las observaciones que el modelo no alcanza a explicar, es una etapa decisiva en la verificación del modelo ajustado. El análisis de dichos residuos considera la evaluación y corroboración de varios supuestos que fundamentan al modelo y que se estudian un poco más adelante.

Por lo pronto, es necesario mencionar que el supuesto de normalidad de los residuales es muy importante en la verificación del modelo. Este supuesto puede validarse mediante la aplicación de una transformación "normalizante", como la transformación potencia, que transformaron Box y Cox en 1964 para evitar la discontinuidad en $\lambda = 0$ y que se convirtió en

$$Z_t^\lambda = \begin{cases} \frac{(Z_t^\lambda - 1)}{\lambda} & \text{si } \lambda \neq 0 \\ \log(Z_t) & \lambda = 0 \end{cases} \quad \dots(4.43)$$

con $Z_t > 0$.

El índice de esta transformación λ , puede estimarse de manera simultánea con los demás parámetros $\phi, \theta, \bar{\theta}$, mediante el método de máxima verosimilitud (Ansley, Spivey y Wroblewski, 1977, citados en Guerrero V.M. 1991).

En realidad al usar esta transformación (3.62), se trata de validar al modelo completo, ya que el parámetro " λ " se elige con el criterio de que el modelo ARIMA expresado al inicio de esta sección, sea válido, así como el supuesto de normalidad de los errores o residuales (con varianza constante).

Terminamos así todo el proceso de estimación de los parámetros (II.2), justamente con un modelo AR, AM, ARMA O ARIMA, a cuyos parámetros les hemos asignado un valor. El modelo está ajustado a los datos. Muy bien, pero...¿y el modelo ajustado sirve o no sirve?. Esto hay que verificarlo. El modelo estimado, con o sin transformación, debe someterse a la verificación de supuestos, que se expone en el siguiente inciso II.3.

4.3 Verificación de supuestos: Análisis de residuales y pruebas de Bondad del ajuste.

Se ha mencionado ya que todo modelo es una mera representación simplificada de la realidad. Bajo esta idea, todo modelo es erróneo. Entonces, habrá que elegir entre varios modelos a aquél que presente menos fallas, o fallas menos importantes. Las fallas de los modelos se miden como violaciones a los supuestos que fundamentan al modelo. Por este motivo, es necesario poner en tela de juicio a todos los posibles modelos para detectar sus fallas. En nuestro caso, "la tela de juicio" son las pruebas de la bondad del ajuste; es decir, para comprobar si el modelo sirve o no sirve, se verifica si cumple con los supuestos del Análisis de residuales, así como con los supuestos de parsimonia, admisibilidad y estabilidad.

4.3.1.- Análisis de residuales:

Análisis de residuales es una de las formas mas claras y simples para detectar violaciones a los supuestos. "Residual" es aquella parte de las observaciones que no puede ser explicada por el modelo. Los "residuos" de la regresión, tienen que ser un ruido blanco, dicho de otra forma, los residuales para el modelo

$$\phi(B)W_t = \theta(B)a_t, \quad \dots(4.44)$$

se definen mediante

$$\begin{aligned} \hat{a}_t &= [\hat{\theta}(B)]^{-1} \phi(B)W_t \\ &= \hat{\pi}(B)W_t \end{aligned} \quad \dots(4.45)$$

de esta relación se sigue que

$$W_t = \hat{a}_t + \hat{\pi}_1 W_{t-1} + \hat{\pi}_2 W_{t-2} + \dots \quad \dots(4.46)$$

ESTADÍSTICA
ACTIVIDAD

ahora, la estimación de W_t que se puede hacer con base en el polinomio estimado $\hat{\pi}(B)$ y en las observaciones hasta el tiempo $t-1$, por lo que viene a ser

$$\hat{W}_t = \hat{\pi}_1 W_{t-1} + \hat{\pi}_2 W_{t-2} + \dots \quad \dots\dots(4.47)$$

por tanto, de estas dos últimas expresiones se obtiene que

$$\hat{a}_t = W_t - \hat{W}_t \quad \dots\dots\dots(4.48)$$

es decir, los residuales, como son la parte de las observaciones que no explica el modelo, miden la discrepancia entre los valores observados y los valores estimados por el modelo.

Cuando el tamaño de muestra es grande, los errores aleatorios y los residuales (que también son variables aleatorias), son esencialmente iguales; por esta razón, al analizar los residuales observados $\{a_t\}$ se analiza básicamente lo que debería ser una realización del proceso de ruido blanco $\{a_t\}$. Los residuos de regresión tienen que ser un ruido blanco precisamente. Si los residuales tienen un patrón de comportamiento con valores en la media, aumento o disminución en la varianza, o si tienen autocorrelaciones, éso está mal y el modelo no me sirve para nada.

Entonces, los supuestos acerca del proceso $\{a_t\}$ pueden verificarse y probablemente corregirse, de la siguiente manera.

***Supuesto 1.- $\{a_t\}$ tiene media cero.**

Es decir, los residuales o "ruido blanco", deben tener media igual a cero. Si la media de los residuales no es cero, el ajuste no está bien.

Verificación:

Calcular la media aritmética $m(\hat{a})$ y la desviación estándar muestra de los residuales $\hat{\sigma}_a$.

$$m(\hat{a}) = \sum_{t=1}^N \frac{\hat{a}_t}{(N-d-p)} \quad \dots(4.49)$$

$$\hat{\sigma}_a = \sqrt{\sum_{t=1}^N \frac{[\hat{a}_t - m(\hat{a})]^2}{(N-d-p-q)}} \quad \dots(4.50)$$

con $t'=d+p+1$, para construir el cociente $\frac{\sqrt{(N-d-p)(m(\hat{a}))}}{\hat{\sigma}_a}$

Si el valor absoluto de dicho cociente es menor que dos, se dirá que no hay evidencia de que la media del proceso de ruido blanco sea distinta de cero y por lo mismo no se rechaza el supuesto; por el

contrario, si $\left| \frac{\sqrt{(N-d-p)(m(\hat{a}))}}{\hat{\sigma}_a} \right|$ es ≥ 2 , entonces la media de los residuales es significativamente distinta de cero, lo cual implica que el supuesto se ha violado.

Corrección si se viola el supuesto 1.-:

El que la media de los residuales sea significativamente distinta de cero, implica que existe una parte determinista o semideterminista en $\{a_t\}$ que no ha sido considerada por el modelo, por lo tanto, se podría requerir la inclusión de un tendencia determinista θ_0 en el modelo, que sea estimada conjuntamente con el resto de los parámetros. El valor inicial para θ_0 vendría dado por $m(\hat{a})$ y es la parte determinista no tenida en cuenta por el modelo. Ahora bien,

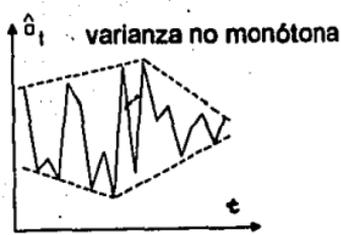
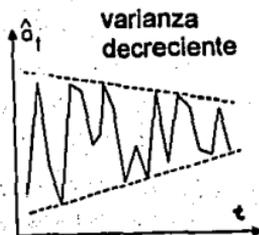
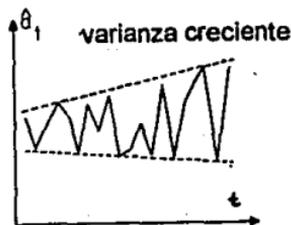
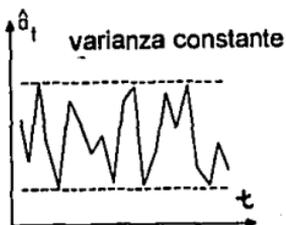
antes de incluir al nuevo parámetro θ_0 en el modelo, conviene considerar la posibilidad de que un término autorregresivo o una diferencia más, sean necesarios en el modelo, en lugar de θ_0 . La diferencia o el parámetro autorregresivo pueden representar a una posible tendencia adaptativa presente aún en los residuales, mientras que θ_0 representaría a una tendencia estrictamente determinista.

***Supuesto 2.-{at} tiene varianza constante.**

Es decir, la varianza de los residuales o ruido blanco, no debe ir aumentando ni disminuyendo, debe ser constante. Si la varianza no es constante, hay que hacerle transformaciones a los datos.

Verificación:

Hacer una gráfica de residuales contra el tiempo para observar, visualmente, si la varianza parece o no ser constante. Algunos de los comportamientos típicos de la varianza residual, se muestran a continuación. La verificación visual de la varianza residual se puede permitir, porque las violaciones muy notorias a este supuesto son las que realmente llegan a causar problemas.



Corrección si se viola el supuesto 2.-:

Si la varianza parece seguir un patrón de crecimiento o de decrecimiento como los de las figuras mostradas anteriormente, es posible aplicar una transformación potencia para estabilizar la varianza de la serie. La transformación potencia o transformación "normalizante" ya ha sido mencionada al final del inciso II.2, en la ecuación (4.43).

***Supuesto 3.- Las variables aleatorias $\{a_t\}$ son mutuamente independientes.**

Como ya ha sido mencionado, los residuos del modelo tienen que ser justamente un ruido blanco, si tienen un patrón, implica que tienen autocorrelación y eso está mal, los residuos del modelo no deben estar correlacionados. ¡Ni covarianza ni autocorrelación en los residuales!. Para verificar ésto, se pide un autocorrelograma de los residuales. Si hay autocorrelación entre los residuales, el modelo no sirve.

Verificación:

Independencia implica no autocorrelación, por tanto, se debe requerir que $\rho_k(\hat{a}) = 0$ para toda $k \neq 0$. Esto es que las variables aleatorias no están autocorrelacionadas. Esto se puede verificar, calculando en primer lugar la Función de Autocorrelación muestra de

los residuales $\{\hat{r}_k(\hat{a})\}$, que en el supuesto de que su media es cero, está dada por:

$$r_k(\hat{a}) = \frac{\sum_{t=1}^{N-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^N \hat{a}_t^2} \quad k=1,2,\dots \quad \dots(4.51)$$

con $t' = d+p+1$, y posteriormente:

i) estimar la desviación estándar de $r_k(\hat{a})$ como

$$\sqrt{\text{Var}[r_k(\hat{a})]} = \frac{1}{\sqrt{(N-d-p)}} \quad \dots(4.52)$$

para determinar la significación estadística individual de las

autocorrelaciones de los residuales: si el valor absoluto de $|r_k(\hat{a})| \geq$

$\frac{2}{\sqrt{(N-d-p)}}$ se dirá que la autocorrelación k-ésima es significativamente distinta de cero. Para autocorrelaciones correspondientes a retrasos pequeños ($k \leq 3$), conviene realizar una prueba conjunta de la significación de las primeras K autocorrelaciones.

ii) Calcúlese el estadístico Q de Box y Pierce (1970) para realizar una prueba de significación conjunta (Prueba "Portmanteau") de K autocorrelaciones simultáneamente, con el estadístico

$$Q = (N-d-p) \sum_{k=1}^K r_k^2(\hat{a}) \quad \dots(4.53)$$

el cual, si K es grande ($k > 20$), sigue aproximadamente una distribución Ji cuadrada con $K-p-q$ grados de libertad; de aquí que el valor de Q calculado debe ser comparado con valores de tablas de Ji cuadrada, con los correspondientes grados de libertad, para efectuar la prueba de significación. En ocasiones, la aproximación lograda con

Q no es del todo apropiada. Por esta razón, se utiliza el estadístico Q' de Ljung t Box (1978):

$$Q' = (N-d-p)(N-d-p+2) \sum_{k=1}^K \frac{r_k^2(\hat{a})}{(N-p-d-k)} \dots (4.54)$$

que resulta ser más adecuado para realizar la prueba

Corrección si se viola el supuesto 3.-:

Si la verificación indica que las autocorrelaciones no son las correspondientes a un ruido blanco, se puede suponer que corresponden a un cierto proceso ARMA; por tanto, es necesario

grafilar algunos valores de $\{r_k(\hat{a})\}$, y tratar de identificar algún proceso ARMA para los residuales. Esto sugeriría modificaciones al modelo originalmente identificado para $\{W_t\}$, ya que faltaron componentes del tipo ARMA en el modelo originalmente identificado. Esto es, si el proceso real es

$$\phi(B)W_t = \theta(B)a_t, \text{ con } E(W_t) = 0 \text{ y } \{a_t\} \text{ ruido blanco } \dots (4.55)$$

pero el modelo identificado erróneamente fué

$$\phi'(B)W_t = \theta'(B)b_t, \text{ donde } \{b_t\} \text{ no es ruido blanco} \quad (4.56)$$

entonces $\{b_t\}$ seguirá el comportamiento dictado por el siguiente proceso ARMA

$$\theta'(B)\phi(B)b_t = \phi'(B)\theta(B)a_t \quad \dots (4.57)$$

Si éste es el caso, se estima un nuevo modelo para el proceso original W_t considerando los componentes ARMA del ruido blanco $\{b_t\}$.

***Supuesto 4.-at tiene una distribución normal, para toda t**

EL ruido blanco debe tener un histograma con una distribución normal, lo que habla de su papel probabilístico. El histograma de los residuos puede delatar observaciones atípicas, desviaciones atípicas. Las desviaciones atípicas pueden provenir de un error, pero también pueden ser indicativo de una intervención, es decir, una alteración exógena, ajena al comportamiento histórico de la serie en estudio.

Verificación:

(i) Para una distribución normal, aproximadamente 95% de las observaciones deben localizarse dentro de un intervalo de dos desviaciones estándar por arriba y por debajo de la media; y, si se cumple que la media de los residuales sea cero, se esperaría que a lo

más de un total de $\frac{(N - d - p)}{20}$ observaciones se localizaran fuera del intervalo $(-2\hat{\sigma}_a, 2\hat{\sigma}_a)$

Este supuesto se puede verificar también visualmente con la misma gráfica que se recomendó emplear en la verificación del supuesto de varianza constante, ya que las desviaciones atípicas salen del límite del intervalo $(-2\hat{\sigma}_a, 2\hat{\sigma}_a)$ considerado en este ejemplo.

(ii) Un histograma de los residuales visualizaría la forma de su distribución, con esto se podría detectar en particular, si la distribución contiene asimetrías. El histograma de los residuos, también nos permite detectar desviaciones atípicas, que basadas en la normalidad pueden provenir de un simple error, o bien ser indicativos de una intervención, como se explica mas adelante.

Corrección en caso de violación a este supuesto:

El supuesto de normalidad se debe cumplir para los errores aleatorios $\{a_t\}$, pero no tiene por que ser satisfecho exactamente por los residuales $\{a_t\}$; por esta razón, es de esperar que pequeñas violaciones a este supuesto no causen problemas en absoluto. Ahora, si las violaciones son muy notorias, se puede aplicar una transformación "normalizante", anteriormente mencionada.

***Supuesto 5.-** Implicítamente, se ha supuesto que no existen observaciones aberrantes (posiblemente ajenas a la serie en estudio).

Verificación:

Nuevamente, la gráfica de residuales contra el tiempo permitirá visualizar si existe este tipo de observaciones anómalas. Por ejemplo, si un residual se encuentra fuera del intervalo $(-3\hat{\sigma}_a, 3\hat{\sigma}_a)$ implica que o bien sucedió un evento cuya probabilidad de ocurrencia era de aproximadamente 0.2 %, que sería muy, muy extraño, o el residual en cuestión corresponde a una observación que no fué generada por el mismo proceso generador del resto de la serie. Como una regla empírica de trabajo, podría considerarse como "sospechosas" a las observaciones cuyos residuales estén fuera del intervalo $(-3\hat{\sigma}_a, 3\hat{\sigma}_a)$.

Corrección en caso de violación al supuesto 5.-:

Toda observación, de no tratarse de un error al copiar los datos, puede contener información muy valiosa para los fines del estudio. No hay que descartar la posibilidad de que una "observación sospechosa" señale la ocurrencia de un evento altamente improbable, y si se desecha la observación o se "ajusta" de alguna manera, se perjudique el análisis. Es necesario investigar las causas de que tal o cual observación sea sospechosa. Si se descarta la posibilidad de que se trate de un simple error de copiado de datos, la "observación

sospechosa" podría deberse en realidad, a un cambio estructural en el comportamiento del fenómeno, causado quizá por una **INTERVENCION EXOGENA** a la serie en estudio. Esto ameritaría un análisis específico de la intervención, que se comentará en el capítulo 5.

En suma en un análisis de residuales se pide que todo lo que no explique el modelo, sus "residuos", se comporten realmente como "ruido blanco", es decir, se pide que:

- la media de los residuales sea cero.
- la varianza de los residuales debe ser constante.
- no deben haber covarianza entre los residuales, ni deben estar correlacionados entre ellos.
- los residuales tengan una distribución normal.

4.3.2.- Otras pruebas para verificar la Bondad del ajuste:

***Supuesto 6.-:El modelo considerado es parsimonioso.**

Verificación:

La parsimonia implica que no es posible reducir el número de parámetros involucrados en el modelo, porque todos son necesarios para explicar el comportamiento del fenómeno, y no pueden ser considerados como iguales a cero. Para comprobar que el parámetro en cuestión no puede ser considerado como igual a cero, se construyen intervalos al 95% de confianza del tipo:

$$(\hat{\theta} - 2\sqrt{\text{Vâr}(\hat{\theta})}, \hat{\theta} + 2\sqrt{\text{Vâr}(\hat{\theta})}) \quad \dots(4.58)$$

para cada uno de los parámetros involucrados en el modelo y se observa si el valor cero se encuentra dentro del intervalo.

Corrección en caso de violación a el supuesto 6.-

Si el cero es un valor razonable para el parámetro en turno, entonces, este parámetro deberá cancelarse y se volverá a estimar el modelo sin él. Pueden llegar a existir situaciones en las que el conocimiento del fenómeno indique que el modelo debe considerar a un parámetro particular, aunque la verificación indique que su valor es cero.

Es importante mencionar que la sobreparametrización del modelo provoca inestabilidad en el modelo, como se verá posteriormente.

*** Supuesto 7.- El modelo es admisible.**

Verificación:

La verificación para el caso de modelos AR(1), AR(2), MA(1), MA(2) O ARMA (1,1), puede efectuarse corroborando que los parámetros estimados se encuentran dentro de las regiones admisibles correspondientes que se presentan en las siguientes figuras. Para modelos de órdenes superiores o más complicados, se debe consultar la obra del Dr. Guerrero mencionada en la bibliografía.

***Supuesto 8.- El modelo es estable en los parámetros.**

Verificación:

La principal causa de inestabilidad en un modelo, es la redundancia de parámetros. En ocasiones es posible compensar cambios en un parámetro del modelo mediante cambios en otro parámetro del modelo, sin que esto altere la suma de cuadrados. En particular, se debe estar alerta en cuanto a correlaciones altas (positivas o negativas) entre los parámetros estimados, ya que éstas, son indicadoras de posibles causas de inestabilidad. Se requiere por tanto, calcular las correlaciones entre parejas de parámetros estimados.

Corrección a violaciones de este supuesto:

Una correlación alta se puede corregir cancelando alguno de los parámetros involucrados cuya correlación es alta. Sin embargo en

ocasiones la inclusión de los dos parámetros es necesaria para obtener una representación adecuada de la serie. En estos casos, se acepta el hecho de que existan correlaciones altas.

4.3.3 Tablas de resultados

Presentación de resultados de la estimación de modelos.

Dado que se trata de seleccionar a un modelo dentro de varias alternativas, a continuación se presenta un formato para presentar los resultados numéricos necesarios para realizar la verificación de modelos, y facilitar de esta manera la elección del mejor modelo.

El cuadro consta de diez columnas:

1. Período de observación de la serie, y el número de observaciones de que consta.
2. Descripción explícita del modelo. Aquí se muestran todos los parámetros que intervienen en el modelo.
3. Valores estimados para cada uno de los parámetros.
4. Intervalos de confianza al 95% de cada uno de los parámetros involucrados en el modelo.
5. Correlaciones entre los parámetros estimados que sean mayores, en valor absoluto, que 0.5.

ANÁLISIS DE RESIDUALES:

6. La media residual $m(\hat{a})$ y el cociente $\frac{\sqrt{(N-d-p)(m(\hat{a}))}}{\hat{\sigma}_a}$

$$m(\hat{a}) = \sum_{t=t}^N \frac{\hat{a}_t}{(N-d-p)}$$

7. La desviación estándar residual:8.

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

8. El estadístico Q de Vox y Pierce o Q' de Ljung y Vox, con sus correspondientes grados de libertad.

9. Las autocorrelaciones de los residuales $r_k(\hat{a})$, que se encuentran

fuera de $\left(\frac{-2}{\sqrt{N-d-p}}, \frac{2}{\sqrt{N-d-p}} \right)$.

10. Los residuales "grandes", es decir, que están fuera de los intervalos $(-2\hat{\sigma}_a, 2\hat{\sigma}_a)$ o de $(-3\hat{\sigma}_a, 3\hat{\sigma}_a)$, escritos en la forma $\hat{a}_t = C\hat{\sigma}_a$, donde C es el número de veces que \hat{a}_t es mayor que $\hat{\sigma}_a$.

Los resultados para cada serie analizada se resumen en tablas similares a la siguiente:

Resumen de Resultados de la Estimación de Modelos ARIMA

Serie: $Z_t = \text{Linces}$

Transformación:

-----Análisis de Residuales-----

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Periodo y número de observaciones	Modelo	Parámetros estimados	Intervalos del 95% de Confianza	Correlación entre parámetros >0.5 o <0.5	$\hat{\sigma}_a$	$\hat{\sigma}_a$	Q.g.l.	r_k	Residuales grandes
N=									
N=									

4.4. Uso del modelo: pronósticos.

En el cuadro 1 de la página 56, observamos que la última parte de la estrategia para construir modelos ARIMA es el uso del modelo.

Los modelos ARIMA para series de tiempo, se construyen para diferentes finalidades. Para explicar el fenómeno en estudio, para hacer simulaciones, como una herramienta de control, o bien para hacer pronósticos.

La teoría que sustenta la obtención de pronósticos se estudia de manera general en el siguiente capítulo, donde también se presentan los pronósticos en condiciones óptimas, para series estacionarias; así como dos casos especiales de pronósticos, para series de tiempo con tendencia, y para series de tiempo con intervenciones.

CAPITULO 5:

PRONOSTICO CON MODELOS ARIMA.

5.1. Pronósticos con modelos ARIMA: generalidades:

Hemos dicho que uno de los objetivos más ambiciosos al aplicar un análisis de series de tiempo a un conjunto de datos, es poder pronosticar los valores futuros de la serie en estudio.

Pronosticar con series de tiempo, no es saber exactamente cuando va a ocurrir un evento, sino dar intervalos de lo que va a pasar. El criterio estadístico que dan las herramientas estadísticas de series de tiempo, es para tener un idea de lo que va a ocurrir. No da números. El estadístico en un grupo de trabajo, no debe hacer operaciones y aplicar técnicas nada más; debe ser capaz de incidir en la toma de decisiones (Martínez C. apuntes de clase).

La metodología de series de tiempo permite hacer pronósticos a corto plazo. A largo plazo lo que hace es dar tendencias.

5.1.1. Tres formas diferentes de expresar al modelo ARIMA:

Al modelo

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t \quad \dots(5.1.1)$$

lo podemos expresar de diferentes formas:

1a. forma:

Si multiplico al polinomio $\phi(B)(1-B)^d$, del modelo (5.1) encuentro que

$$\phi(B)Z_t = \theta(B)a_t \quad \dots(5.1.2)$$

$$(1 - \phi_1 B - \dots - \phi_{p+d} B^{p+d}) Z_t = \theta(B)a_t \quad \dots(5.1.3)$$

la primera forma de escribir al modelo (5.1) es

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_{p-d} Z_{t-p-d} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad \dots(5.1.4)$$

Esta es la forma que se utiliza para construir el pronóstico, porque sólo depende de las observaciones que tengo.

2a. forma de expresar el modelo:

$$\theta^{-1}(B)\phi(B)(1-B)^d Z_t = a_t, \quad \dots(5.1.5)$$

Mediante esta forma puedo ver el pasado de una observación presente. Si introducimos a Π ,

$$\Pi(B)Z_t = a_t, \quad \dots(5.1.6)$$

$$(1 - \Pi_1 B - \Pi_2 B^2 - \dots)Z_t = a_t, \quad \dots(5.1.7)$$

ahora tenemos Π ponderaciones para ver cada una de la ponderaciones pasadas de Z_t ,

$$Z_t = \Pi_1 Z_{t-1} + \Pi_2 Z_{t-2} + \dots + a_t, \quad \dots(5.1.8)$$

3a. forma de expresar al modelo: para construir la varianza de las ponderaciones Ψ

$$Z_t = \varphi^{-1}(B)\theta(B)a_t, \quad \dots(5.1.9)$$

$$Z_t = \varphi(B)a_t, \quad \dots(5.1.10)$$

$$Z_t = (1 - \varphi_1 B - \varphi_2 B^2 - \dots)a_t, \quad \dots(5.1.11)$$

$$Z_t = a_t - \varphi_1 a_{t-1} - \varphi_2 a_{t-2} - \dots, \quad \dots(5.1.12)$$

Para obtener el pronóstico, se utiliza cualquiera de las tres formas.

Notación:

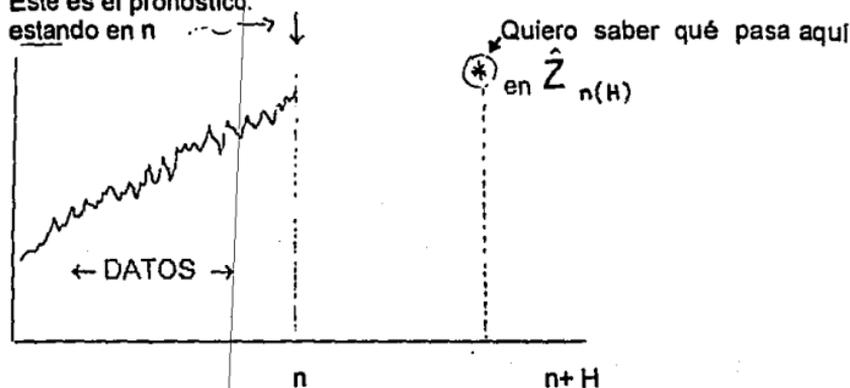
$$Z_1, \dots, Z_n \quad \dots(5.13)$$

lo que nos interesa conocer es qué pasa después de n:

$$Z_{n+H} \quad ? \quad H \geq 1 \quad \text{en el tiempo } n+H \quad \dots(5.14)$$

H = horizonte de pronóstico.

Este es el pronóstico.
estando en n



El pronóstico es el valor de

$$(Z_{n+h})E_t = \hat{Z}_t(h) \quad \dots(5.15)$$

En otras palabras,

$$Z_{n+H} = a_{n+H} + \phi_1 a_{n+H-1} + \dots + \phi_{H-1} a_{n+1} + \phi_H a_n \quad \dots(5.16)$$

$$\hat{Z}_{n+H} = \phi_1 a_{n+1} + \phi_2 a_{n+2} + \dots + \phi_{H-2} a_{n-2} + \phi_{H-1} a_{n-3} + \dots(5.17)$$

estando en n, yo no sé cuál es el pronóstico. El ruido $\{a_t\}$ depende de lo que sucedió antes de n.

La idea es encontrar

$$\hat{\varphi}_H, \hat{\varphi}_{H+1}, \dots \quad \dots(5.1.18)$$

tales que

$$E(Z_{n+H} - \hat{Z}_{n+H})^2 \quad \text{sea mínimo.} \quad \dots(5.1.19)$$

$$E\{a_{n+H} + \varphi_1 a_{n+H+1} + \dots + \varphi_{H-1} a_{n+1} + (\varphi_H - \hat{\varphi}_H) a_n + (\varphi_{H+1} - \hat{\varphi}_{H+1}) a_{n-1} + \dots\}^2$$

.....(5.1.20)

Para que yo pueda encontrar la esperanza de una suma, se necesita que exista la suma. Es decir, que converja.

$$E(Z_{n+H} - \hat{Z}_{n+H})^2 = (1 + \varphi_1^2 + \dots + \varphi_{n-1}^2) \sigma^2 + \sigma^2 \sum_{j=0}^{\infty} (\varphi_{H+j} - \hat{\varphi}_{H+j})^2$$

.....(5.1.21)

5.1.2. Proceso para obtener el pronóstico.

Para obtener el pronóstico, lo que se necesita hacer es minimizar la suma (4.1.21.), y para eso, es necesario que

$$\sum_{j=0}^{\infty} (\varphi_{H+j} - \hat{\varphi}_{H+j})^2 \quad \text{sea igual a cero.} \quad \dots(5.1.22)$$

y para esto necesito que

$$\varphi_{H+j} = \hat{\varphi}_{H+j}^* \quad \dots(5.1.23)$$

Entonces, para hacer el pronóstico se siguen los siguientes pasos:

1. Escribo el modelo

$$\hat{Z}_{n+H} = \varphi_1 a_n + \varphi_2 a_{n-1} + \dots \quad \dots(5.1.24)$$

2. ¿Qué ocurre si tomo el valor esperado de

$$E_n(Z_{n+H}) \quad \text{estando en } n? \Rightarrow 0$$

Pues no tengo ningún valor en H.

Pero de n hacia atrás, el valor esperado es:

$$E_n(Z_{n+H}) = \varphi_1 a_n + \varphi_2 a_{n-1} + \dots \quad \dots(5.1.25)$$

pero

$$\hat{Z}_{n+H} = \varphi_1 a_n + \varphi_2 a_{n-1} + \dots \Rightarrow \varphi_1 a_n + \varphi_2 a_{n-1} + \dots \quad \dots(5.1.26)$$

por lo tanto,

$$\hat{Z}_{n+H} = E_n(Z_{n+H}) \quad \dots(5.1.27)$$

este concepto es muy importante para la construcción del pronóstico.

entonces escribo el modelo para n + H, y ése es el pronóstico:

$$\hat{Z}_{(n+H)} = \hat{Z}_{(H)} \quad \dots(5.1.28)$$

3. Ahora, el error de pronóstico en el tiempo h, es

$$Z_{(n+H)} - \hat{Z}_{(n+H)} = E\{a_{n+H} + \varphi_1 a_{n+H-1} + \dots + \varphi_{n-1} a_{n-1}\} \quad \dots(5.1.29)$$

$$Z_{(n+H)} - \hat{Z}_{(n+H)} = a_{n+H} + \varphi_1 a_{n+H-1} + \dots + \varphi_{n-1} a_{n-1} \quad \dots(5.1.30)$$

el error del pronóstico sólo depende de los ruidos futuros, pero los ruidos futuros son variables aleatorias que no conozco (Martínez Collantes 1993, apuntes de clase)

4. La media del error de pronóstico:

Si tengo el valor esperado, tengo la media del proceso.

$$E\{Z_{n+H} - \hat{Z}_{n+H}\} = 0 \quad \dots(5.1.31)$$

5. La varianza del error de pronóstico:

$$V(Z_{n+H} - \hat{Z}_{n+H}) = \left(1 + \varphi_1^2 + \dots + \varphi_{n-1}^2\right) \sigma^2 \quad \dots(5.1.32)$$

6. Los pronósticos son un período adelante:

$$Z_{(n+1)} - \hat{Z}_{(n+1)} = a_{n+1} \quad \dots(5.1.33)$$

Los errores de pronóstico un período adelante, no son sino los ruidos.

$$e_n(1) = a_{n+1} \quad \dots(5.1.34)$$

7. Los errores del pronóstico no están correlacionados:

$$\rho[e_n(1); e_n^*(1)] = 0 \quad \dots(5.1.35)$$

8. Pero los errores de pronóstico con un mismo origen si están correlacionados

$$\rho[e_n(H); e_n^*(H)] \neq 0 \quad \text{.....(5.1.36)}$$

si H es > 1 .

9. Cuando tengo orígenes diferentes y el mismo horizonte de pronóstico,

$$\rho[e_n(H); e_n^*(H)] \neq 0 \quad \text{.....(5.1.37)}$$

$$\sum_{j=0}^{\infty} (\varphi_{n+j} - \varphi_{n+j}^*)^2 \quad \text{.....(5.1.38)}$$

$$|n - n^*| \leq h \quad \text{.....(5.1.39)}$$

Si yo "sobreprometido", tiendo a que todos los errores "se sobreprometiquen".

(Martínez Collantes 1993. apuntes de clase)

Es mejor hacer pronósticos cambiando de origen, porque de esta manera los errores no están correlacionados.

Si el ruido tiene una distribución normal, los errores también tienen una distribución normal:

$$\text{Si} \quad a_t \cong N; \quad e_t(h) \cong N \quad \text{.....(5.1.40)}$$

5.2. Pronósticos ARIMA en condiciones óptimas: Series estacionarias

Calcular al pronóstico equivale a calcular el valor de $\hat{Z}_t(h)$. El pronóstico es el valor de $\hat{Z}_t(h) = (Z_{t+h})E_t$ (5.2.1)

Hay tres maneras de hacerlo:

- 1.- En términos de los ruidos.
- 2.- En términos de la historia pasada.
- 3.- En términos de ruidos blancos.

La fórmula para actualización de pronósticos es:

$$\hat{Z}_n(h) = \varphi_n a_n + \varphi_{n+1} a_{n-1} + \varphi_{n+2} a_{n-2} \quad \dots(5.2.2)$$

cuando se quiere pronosticar desde $n+1$:

$$\hat{Z}_{n+1}(h-1) = \varphi_{n-1} a_{n+1} + \varphi_n a_n + \varphi_{n+1} a_{n-1} + \varphi_{n+2} a_{n-2} \quad \dots(5.2.3)$$

\Rightarrow

$$\hat{Z}_{n+1}(h-1) = \varphi_{n-1} + a_{n+1} + \hat{Z}_n(h) \quad \dots(5.2.4)$$

en éste caso, todos los errores de pronóstico están correlacionados.

Si suponemos que $\{W_t\}$ es una serie de tiempo estacionaria con media cero, obtenida a partir de una serie original $\{Z_t\}$ (con N observaciones), entonces

$$W_t = \nabla^d T(Z_t), \quad \dots(5.2.5)$$

para algún valor de " d " y para una cierta transformación T , con W_t medida como desviación respecto a su media.

Si además W_t admite la representación

$$W_t = \varphi(B)a_t \quad \dots(5.2.6)$$

para la cual existe un modelo ARMA equivalente, modelo que se desea utilizar en la obtención de pronósticos de la serie.

En particular, si a partir del origen " t ", se desea pronosticar a la observación W_{t+h} , un pronóstico cualquiera de esta observación, que se obtenga como combinación lineal de los valores de la serie $\{W_t\}$, y en consecuencia de los errores $\{a_t\}$, será denotado por $\hat{W}_t(h)$, mientras que el pronóstico óptimo se escribirá como $\hat{W}_t(h)$ (Guerrero 1991a),

5.3. Pronósticos ARIMA para series no estacionarias:

En la práctica, la mayoría de las series que se observan son no estacionarias, ya sea porque presentan tendencias, ciclos, y/o intervenciones.

En consecuencia, es necesario mencionar una generalización de los conceptos vistos para series estacionarias. En general, la no estacionariedad de la serie original $\{Z_t\}$ se cancela aproximadamente al determinar alguna transformación $T(Z_t)$ y aplicarle un número apropiado de diferencias, o sea, al considerar a

$$W_t = \nabla^d T(Z_t), \quad \dots\dots (5.3.1)$$

5.3.1. Series con tendencia.

En el capítulo 3 ya se ha explicado que las series con tendencia son series en las que el nivel μ de a serie se modifica constantemente,

La tendencia en una serie se estabiliza mediante la aplicación del factor diferencia ∇^d .

5.3.2. Series estacionales o cíclicas.

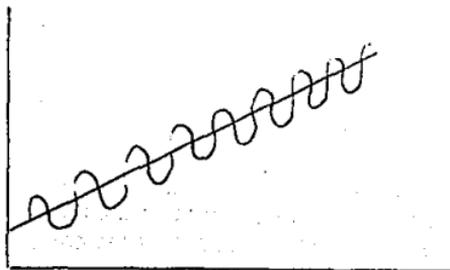
Las series estacionales o cíclicas son series que tienen una parte que se repite cíclicamente, o bien toda la serie tiene un comportamiento por estaciones.

Las gráficas de las series estacionales y de las series con tendencia se ilustran en el capítulo 3 de esta tesis.

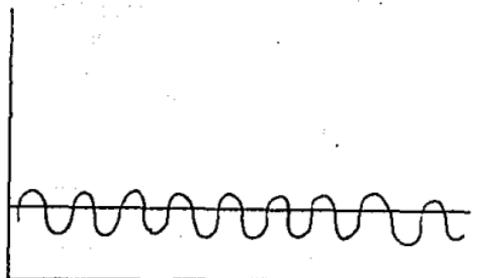
Las series estacionales se deben desestacionalizar para poderlas trabajar. El pronóstico así obtenido es para la serie diferencial, lo mismo que cuando se trabaja la serie sin tendencia. Esto debe tomarse en cuenta

al interpretar los pronósticos obtenidos con el modelo ARIMA de la serie diferencial.

Serie original: cíclica y con tendencia:

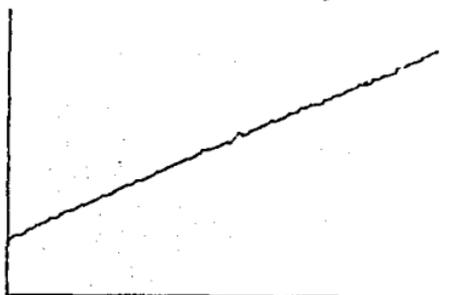


Serie sin tendencia, se pueden apreciar los efectos estacionales netos:



diferencia regular

Serie desestacionalizada, elimino el efecto estacional y evalúo la tendencia:



diferencia estacional

5.4 Pronósticos ARIMA restringidos por cambios en los parámetros del modelo debido a factores exógenos: Análisis de series influenciadas por intervenciones.

Hemos dicho (capítulo 3, pp 49.), que las intervenciones son eventos exógenos que alteran el comportamiento histórico de la serie de tiempo. En ocasiones, al realizar el análisis estadístico de una serie de tiempo, es necesario medir explícitamente la influencia de algún evento anormal, inusual o extraño, que incide sobre la serie de tiempo en estudio, ya sea porque se juzgue de importancia la contribución de dicho evento para explicar el comportamiento de la variable que se estudia, o bien porque se teme que el fenómeno anormal pueda afectar la estimación de los parámetros y los resultados del modelo, en general (Guerrero 1991a).

La presencia de intervenciones en una serie de tiempo se sugiere desde el análisis de la gráfica de los datos originales. Los residuos también pueden insinuar la presencia de una intervención, cuando se observan uno o más residuales mayores que lo esperado, o bien si las gráficas de varianzas de residuales no son monótonas. (ver capítulo 4, pp 82). Un valor muy alto del estadístico de Q, es decir un valor mayor al número de grados de libertad que se tengan, está diciendo que el modelo identificado para representar a la serie no está bien, lo que indica es que hay autocorrelación en los residuos; se está dejando como "basura residual", fuera del modelo, información importante de los datos. La autocorrelación en los residuos se puede verificar también en su autocorrelograma, si hay valores mayores al valor umbral, entonces en esos valores hay que reevaluar al modelo. Si pese a consiguientes ajustes, no se logran modificar las alteraciones en los residuales, esta situación puede indicar la presencia de intervenciones que cambian la estructura histórica de la serie.

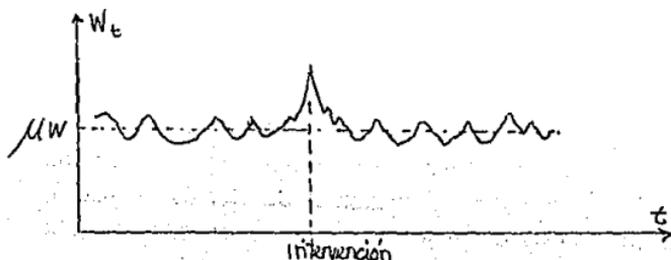
En ocasiones, los efectos de eventos anormales impiden siquiera la identificación de un modelo para representar a la serie. En estos casos, es importante llevar a cabo lo que se conoce como "Análisis de intervención".

El análisis de intervención es una herramienta poderosa que no debe utilizarse solamente para reducir arbitrariamente la magnitud de residuales con valores grandes (Guerrero 1991a). El análisis de intervención fué diseñado para aplicarse en situaciones en las que existe un conocimiento preciso de que un fenómeno tuvo lugar en algún punto (conocido) de la serie de tiempo.

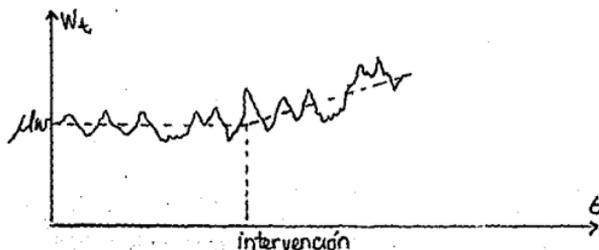
Una intervención puede ser interpretada como la ocurrencia de un evento exógeno al comportamiento histórico de la variable en estudio; tal intervención podría ser un cambio climático drástico, como una glaciación, en el caso de la evolución de una variable biológica.

Los efectos causados directamente por la intervención de fenómenos externos al comportamiento de una serie de tiempo, pueden ser de naturaleza muy variada. En particular, si se considera que una serie de tiempo estacionaria $\{W_t\}$ está constituida por dos partes, una determinista que refleja esencialmente el nivel de la serie (μW) y otra estocástica, que da origen a las fluctuaciones alrededor de dicho nivel, es posible distinguir tres tipos de efectos debido a una intervención:

1.- Efectos que se dejan sentir como una elevación o caída momentánea del nivel, y que desaparecen sin influir en el comportamiento posterior de la serie:



2- Efectos que ejercen una influencia sostenida (no momentáneamente) sobre el nivel de la serie. son efectos que alteran la parte determinista de la serie pero que dejan intacta su parte estocástica.



3.-Efectos que independientemente de influir sobre la parte determinista de la serie, si alteran la estructura de la parte estocástica.



Las intervenciones que causan efectos del tipo 1 y 2, son los que de manera ortodoxa pueden estudiarse mediante el análisis de intervención descrito más adelante. Las intervenciones causantes de efectos del tipo 3, requieren de otro tipo de análisis de intervención, un tanto ad-hoc, para estudiar la serie correspondiente.

5.5 Análisis de intervención:

Un modelo para una serie $\{T(Z_t)\}$ que contenga los efectos de una intervención, puede expresarse como

$$T(Z_t) = \varepsilon_{it} + N_t \quad \dots(5.5.1)$$

donde N_t es un modelo ARIMA estacionario e invertible, porque se puede escribir en términos de su pasado, que representa a la parte estocástica de la serie y que puede expresarse como:

$$\phi(B) \nabla^d N_t = \theta_0 + \theta(B) a_t \quad \dots(5.5.2)$$

con N_t medida como desviaciones respecto a su media si $d=0$, mientras que ε_{it} es una función que permite representar los efectos de la intervención; dicha función debe ser tal que permita apreciar los cambios ocurridos en la serie no sólo en el momento en que ocurrió la intervención, sino los que se dieron posteriormente, atribuibles a ella, es decir, se requiere que ε_{it} sea en realidad un modelo dinámico de la intervención. (Guerrero 1991)

Al modelar series de tiempo con intervenciones, se trata de darles funciones "impulso" o funciones de pulso al modelo. Un impulso es un "empujón" en el tiempo,

$$\begin{aligned} & 1 \text{ si } t = l \\ \text{Pulso } l \text{ al tiempo } t = & \\ & \dots 0 \text{ si } t \neq l \end{aligned} \quad \dots(5.5.3)$$

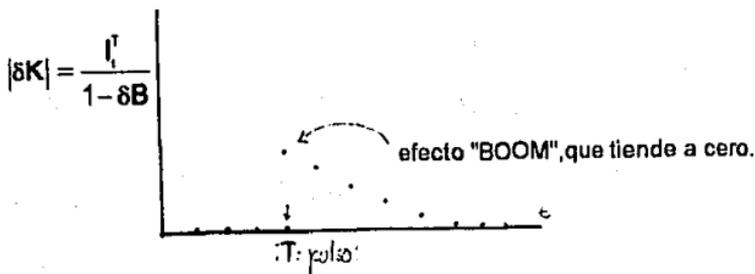
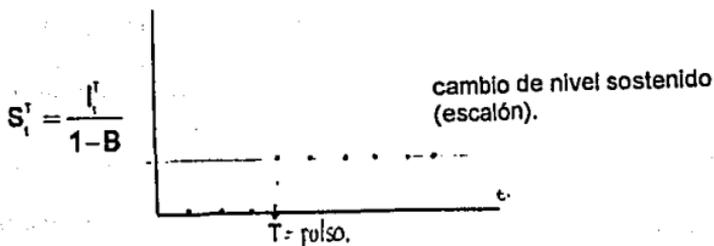
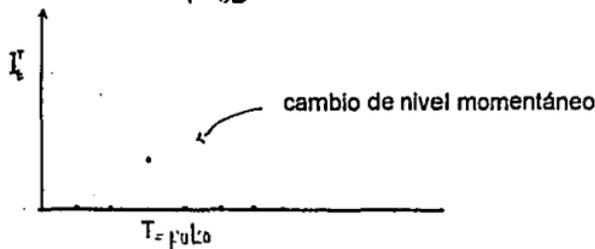
$$(1 + B + B^2 + \dots)_t^l = I_t^l + B I_t^l + B^2 I_t^l + \dots \quad (5.5.4)$$

$$= I_t^l + I_t^{l+1} + I_t^{l+2} + \dots \quad (5.5.5)$$

que provoca un cambio de nivel para el caso de las intervenciones con efecto tipo 1 y 2.

En las $|\delta| < 1$:

$$\frac{I_t^r}{1-\delta B} = (1 + \delta B + \delta^2 B^2 + \dots) \quad (5.5.6)$$



* el efecto "BOOM" es un cambio repentino que se devalúa con el tiempo.

Para el caso de las intervenciones con efecto de tipo 3, que alteran no sólo la estructura estocástica de la serie sino su estructura determinista, es decir, para intervenciones que cambian toda la estructura del modelo, se sugiere partir la serie de tiempo en estudio, ajustar el modelo hasta el momento previo a la intervención, y después ajustar otro modelo hasta el momento previo a la intervención, si el cambio en la secuencia de la serie fue muy drástico. Se trata de hacer un análisis de la serie por partes. (Martínez Collantes 1993, apuntes de clase).

5.6 Pronósticos ARIMA bajo diferentes condiciones:

Por último para concluir este capítulo, cabe mencionar cuatro casos en los que se puede obtener el pronóstico con el modelo ARIMA resultante del análisis de la serie. En primer lugar están los pronósticos obtenidos en "condiciones óptimas"; en este caso, los pronósticos caen dentro del intervalo de confianza al 95% de probabilidad. Para profundizar más en este tema se puede consultar el artículo del Dr. Guerrero "Optimal conditional ARIMA Forecasts" (Guerrero, 1989). La segunda situación, es el caso de los pronósticos que si bien se alejan de dicho intervalo de confianza son aceptables en términos de la variabilidad propia del proceso estocástico en estudio (Guerrero, 1991c. "Restricted ARIMA Forecasts with account for parameter changes"). En tercer lugar, mencionaremos los pronósticos que explican cambios en la estructura determinística del proceso, es decir, cambios en los parámetros del modelo ARIMA (Guerrero, 1991b. "ARIMA Forecasts with restrictions derived from a structural change"). Y finalmente el cuarto caso es el de los pronósticos obtenidos mediante la combinación de la historia de la serie de tiempo, con información preliminar brindada por la propia serie de tiempo (Guerrero 1993. "Combining historical and preliminar information to obtain timely Time Series data").

Las series:

CAPITULO 6: *primera serie.-Aplicación del método de Análisis estadístico de series de tiempo a un ejemplo clásico en Biología:*

El clásico ejemplo de series de tiempo biológicas:

Los lince canadienses.

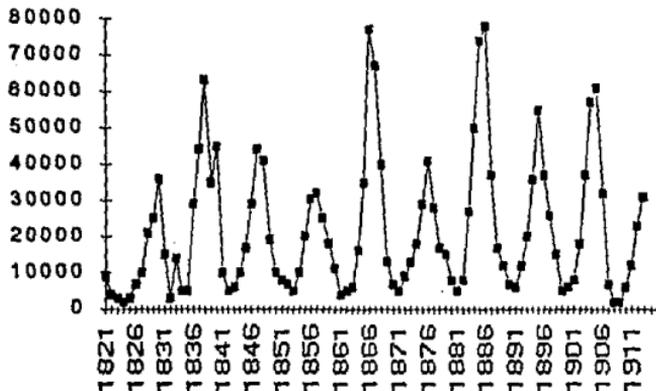
Aplicación del análisis estadístico de las series de tiempo a los datos originales de la compañía exportadora de pieles "Hudson Bay Company" desde 1821 hasta 1913.

Para investigar los alcances y limitaciones del análisis estadístico de series de tiempo aplicado a un proceso estocástico, en este capítulo se analiza la serie de datos de los lince canadienses atrapados en distintas regiones del Canadá por los cazadores y tramperos para la compañía de venta de pieles de la Bahía del Hudson. Esta serie de datos ha recibido muchísima atención tanto por parte de Biólogos, como por analistas de series de tiempo, ya que se trata de un conjunto de datos muy largo y completo. Dado que es un magnífico ejemplo de series de tiempo biológicas, en este capítulo se ejemplifica toda la parte teórica estudiada en los capítulos anteriores y se investigan las posibilidades que brinda el análisis para distinguir los componentes deterministas y estocásticos de la serie.

Desde el punto de vista biológico, es muy interesante observar la regularidad y amplitud con que se repiten ciclos de abundancia poblacional a lo largo de largos períodos. Como es una serie ya clásica por haber sido analizada por muchos ecólogos y analistas de series de tiempo, es un estupendo ejemplo para ilustrar la aplicación del análisis de series de tiempo a un proceso estocástico biológico.

A continuación se presentan los datos originales de la "Hudson's Bay Company", publicados en 1942 por los investigadores Charles Elton y Mary Nicholson (Elton Ch. & Nicholson M. 1942. "The ten year cycle in numbers of the lynx in Canada". J. Anim. Ecol. 11: 2215-2244).

6.1. Gráfica de los cambios en la abundancia anual de los lince canadienses desde 1821 hasta 1913. (archivos de la compañía de la Bahía de Hudson (Elton Ch. & Nicholson M. 1942. "The ten year cycle in numbers of the lynx in Canada". J. Anim. Ecol. 11: 2215-2244).



6.2. Tabla de datos.

AÑOS	No.de lincs
1821	9000
1822	4000
1823	3000
1824	2000
1825	3000
1826	7000
1827	10000
1828	21000
1829	25000
1830	36000
1831	15000
1832	3000
1833	14000
1834	5000
1835	5000
1836	29000
1837	44000
1838	63000
1839	35000
1840	45000
1841	10000
1842	5000
1843	6000
1844	10000
1845	17000
1846	29000
1847	44000
1848	41000
1849	19000
1850	10000
1851	8000
1852	7000
1853	5000
1854	10000

1855	20000
1856	30500
1857	32000
1858	25000
1859	18000
1860	11000
1861	4000
1862	5000
1863	6000
1864	16000
1865	35000
1866	77000
1867	67000
1868	40000
1869	13000
1870	7000
1871	5000
1872	9000
1873	13000
1874	18000
1875	29000
1876	41000
1877	28000
1878	17000
1879	15000
1880	8000
1881	5000
1882	8000
1883	27000
1884	50000
1885	74000
1886	78000
1887	37000
1888	17000
1889	12000
1890	7000
1891	6000
1892	12000
1893	20000
1894	36000

1895	55000
1896	37000
1897	26000
1898	15000
1899	5000
1900	6000
1901	8000
1902	18000
1903	37000
1904	57000
1905	61000
1906	32000
1907	7000
1908	2000
1909	2000
1910	6000
1911	12000
1912	22877
1913	30991

6.3. Análisis de los datos y construcción del modelo ARIMA mediante el uso del paquete PEST: Metodología

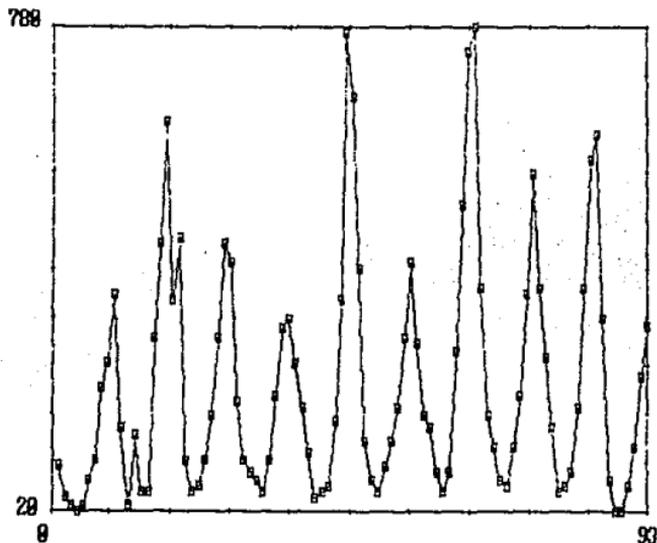
1a etapa: Análisis descriptivo.-

6.3.1 Gráfica de los datos originales presentada por el paquete pest.

Esta gráfica nos muestra las variaciones en el número de lince cazados en Canadá, desde 1821 hasta 1913. Nótese la presencia de ciclos de aproximadamente de 10 años de amplitud, por tanto se trata de una serie estacional (con estaciones). Obsérvese también que la serie es estacionaria, es decir, no presenta tendencias.

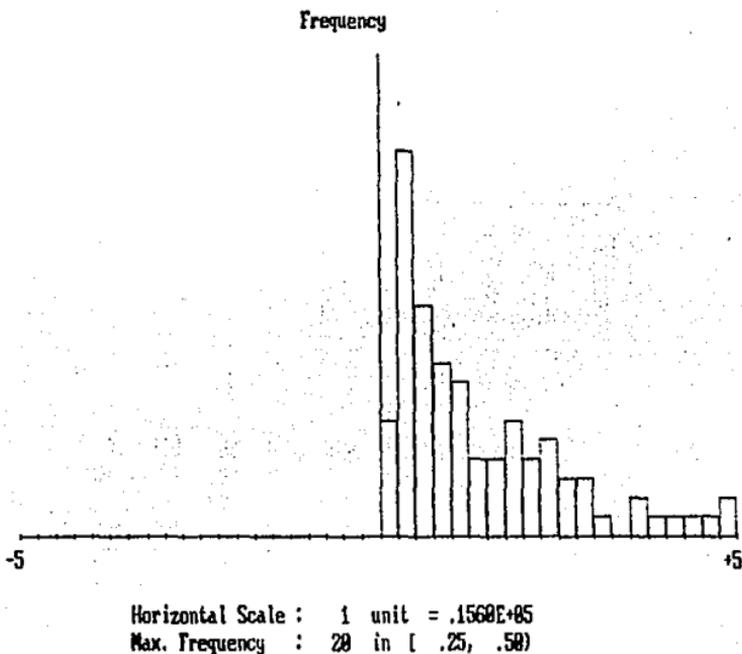
El total de datos es de 93 años, ($n = 93$).

El valor inicial es de 9000 pieles de lince cazados en el año 1821, el valor final es de 30991 pieles de lince cazados en 1913.



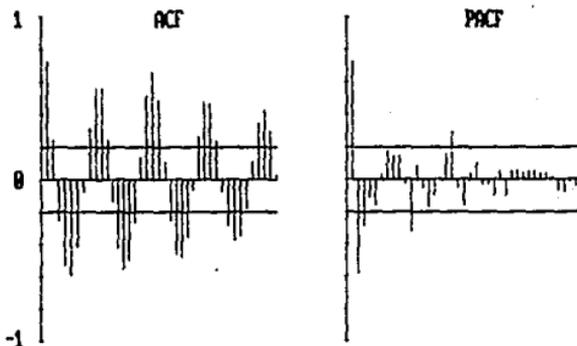
Vertical scale: 1 unit = .100000E+03;
 Max. on vertical scale = .700000E+05; Min. = .200000E+04

**6.3.2. Histograma y estadísticos de los datos originales.
(lincs 1821-1913).**



valor de la media = 21692.13000
desviación estándar = 18603.4000

6.3.3. Autocorrelograma y autocorrelograma parcial de los datos originales.



ACF:	.717	.233	-.242	-.528	-.588	-.416	-.069	.385	.555	.551
	.241	-.135	-.429	-.554	-.498	-.253	.129	.588	.668	.488
	.185	-.243	-.455	-.487	-.349	-.054	.253	.472	.462	.235
	-.067	-.286	-.379	-.344	-.173	.183	.338	.426	.291	.021
PACF:	.717	-.578	-.283	-.104	-.148	.019	.180	.137	.136	-.029
	-.323	.007	-.053	-.161	-.009	.015	.156	.289	-.049	-.157
	.031	.111	-.019	-.029	-.100	.044	-.097	.044	.043	.035
	.049	.051	.036	.038	.018	-.061	-.076	-.013	-.036	-.067

2a. etapa: Construcción del modelo:

6.4. Metodología para la construcción del modelo ARIMA de la serie de tiempo de los lince canadienses.

6.4.1. Identificación del modelo:

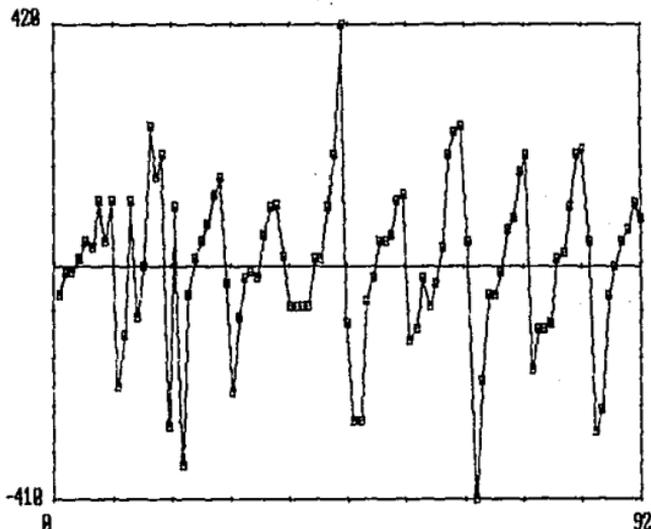
a) Determinar una serie estacionaria en función de los datos originales: Estabilización de la varianza, comparando diferencias. Determinar el valor de "d".

¿Es necesario diferenciar? ¿ARIMA o ARMA?.

1a. diferencia: (se comparan las varianzas de los datos originales y diferenciando una vez los datos originales).

Total de observaciones con la 1a. diferencia = 92

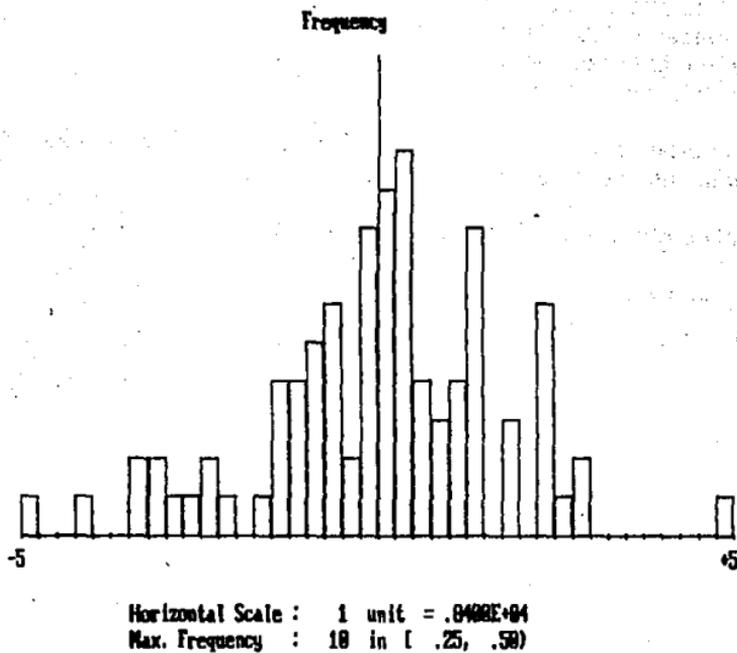
Gráfica de los datos originales diferenciados una vez:



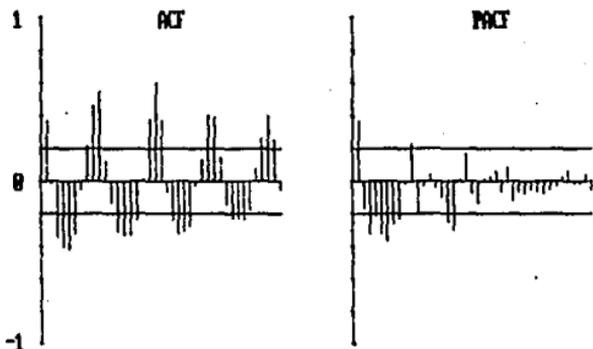
valor de la media = 239.03

desviación estándar = 13969.84000

Histograma de la 1a. diferencia:



Autocorrelograma de la 1a. diferencia:



ACF:	.368	-.018	-.336	-.484	-.411	-.316	-.052	.280	.464	.546
	.118	-.134	-.308	-.324	-.327	-.241	.013	.378	.598	.362
	-.062	-.248	-.322	-.382	-.274	-.023	.134	.397	.394	.145
	-.139	-.227	-.227	-.241	-.182	.069	.253	.399	.243	-.043
PACF:	.368	-.161	-.323	-.217	-.313	-.367	-.255	-.229	-.013	.226
	-.284	-.028	.033	-.034	-.096	-.244	-.298	.008	.171	-.078
	-.134	.015	.027	.062	-.059	.004	-.119	-.058	-.073	-.053
	-.068	-.067	-.058	-.023	.028	.068	-.008	-.014	.033	-.047

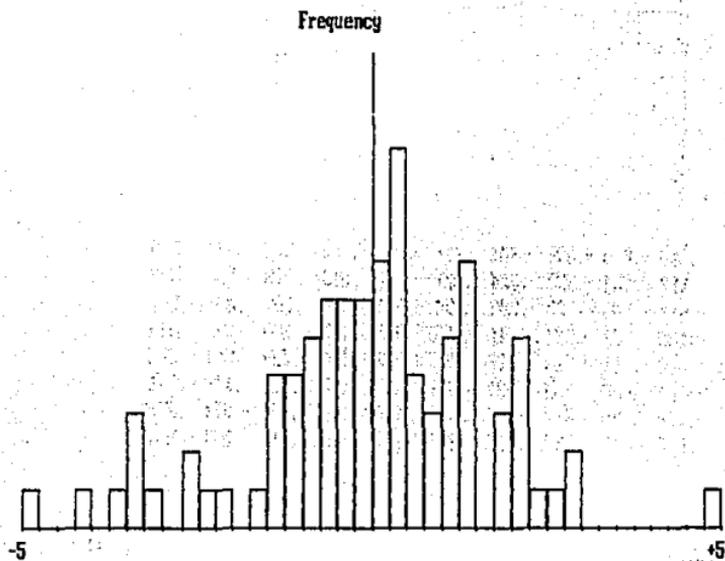
DATA FILE =

,DIFF'D

sigue 6.4.1.

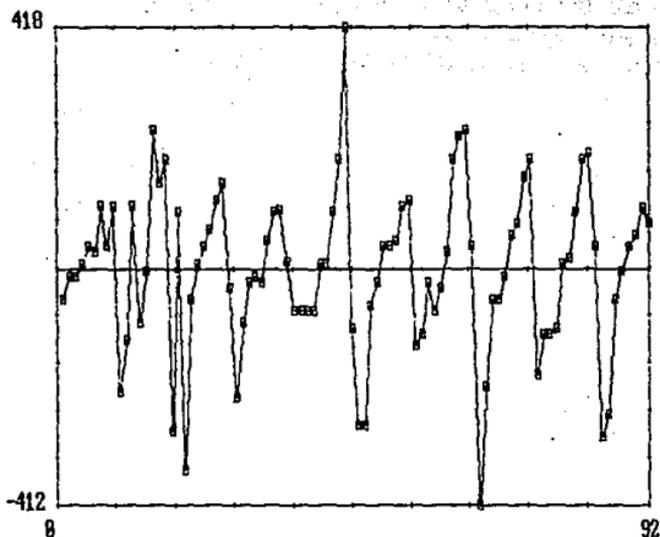
b) Sustraer la media:

Histograma:



Horizontal Scale : 1 unit = .8352E+04
Max. Frequency : 10 in [.25, .50)

datos con una diferencia:



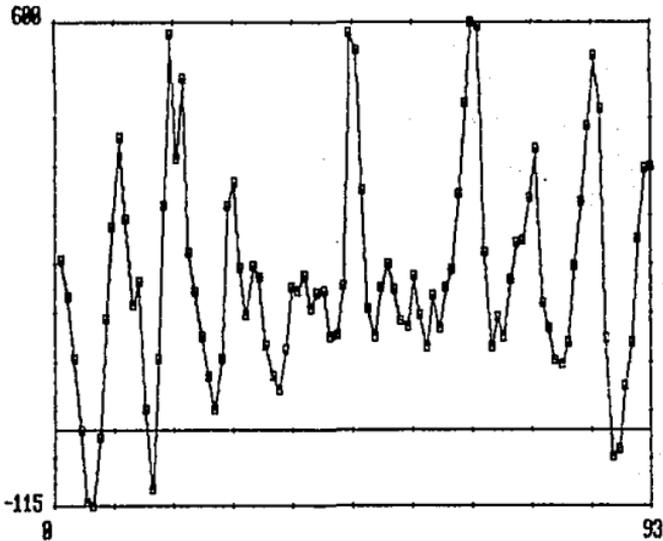
valor de la media = 0,00
desviación estándar = 13969,840

* Nótese que la desviación estándar es menor, por tanto, mejora:
18603.4 > 13969.840.

sin embargo no es conveniente sobreparametrizar; parsimonia.

sigue 6.4.1 identificación del modelo:

c) Análisis de estacionalidad (ciclos) a partir de los datos originales.
(regreso a los datos originales)
gráfica de los datos con análisis de estacionalidad:



valor de la media = 22109.98
desviación estándar = 15569.95

Tabla de los componentes estacionarios:

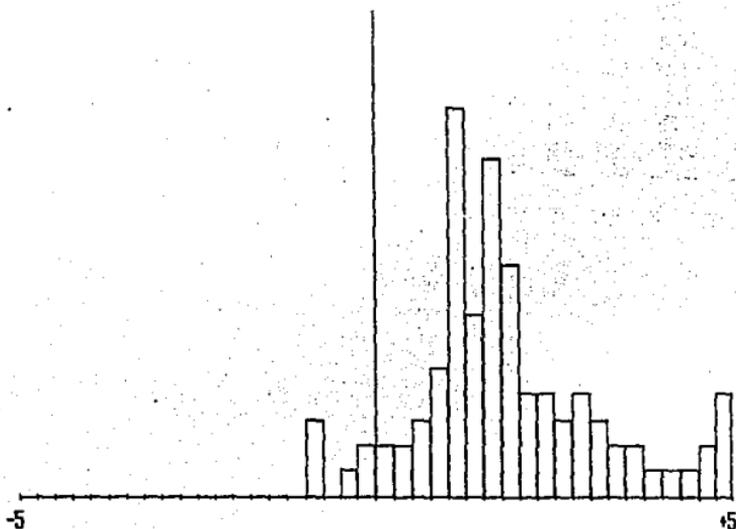
```

PERIOD OF SEASONAL COMPONENT
  (enter period) 1, or 0 to escape): 10
SEASONAL COMPONENT 1 = #####
SEASONAL COMPONENT 2 = #####
SEASONAL COMPONENT 3 = -7574.5160000000
SEASONAL COMPONENT 4 = 1937.9850000000
SEASONAL COMPONENT 5 = 14819.2300000000
SEASONAL COMPONENT 6 = 18456.8400000000
SEASONAL COMPONENT 7 = 11856.7200000000
SEASONAL COMPONENT 8 = 4736.3400000000
SEASONAL COMPONENT 9 = -4599.5160000000
SEASONAL COMPONENT 10 = -6885.7660000000

```

Histograma.

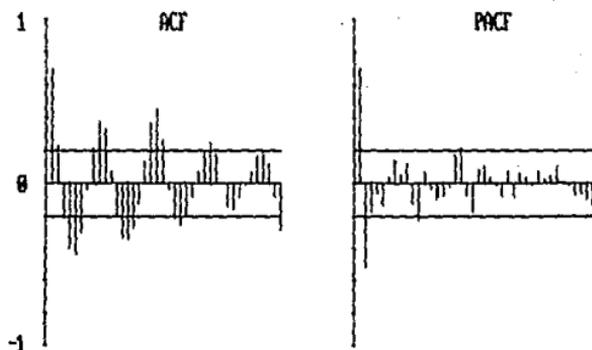
Frequency



Horizontal Scale : 1 unit = .1280E+05

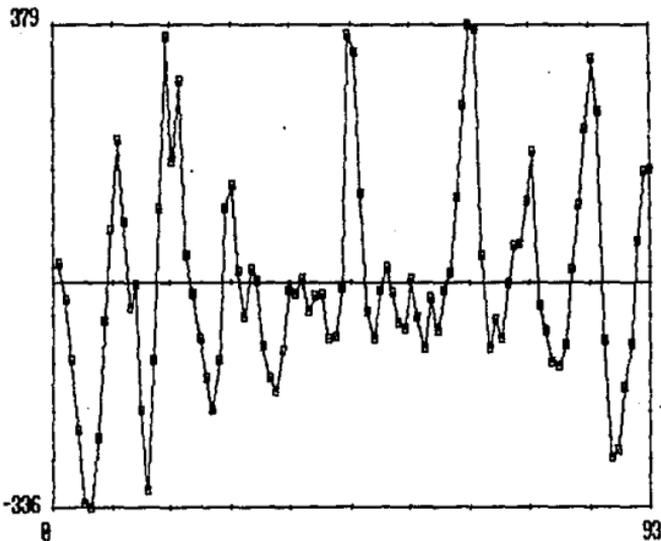
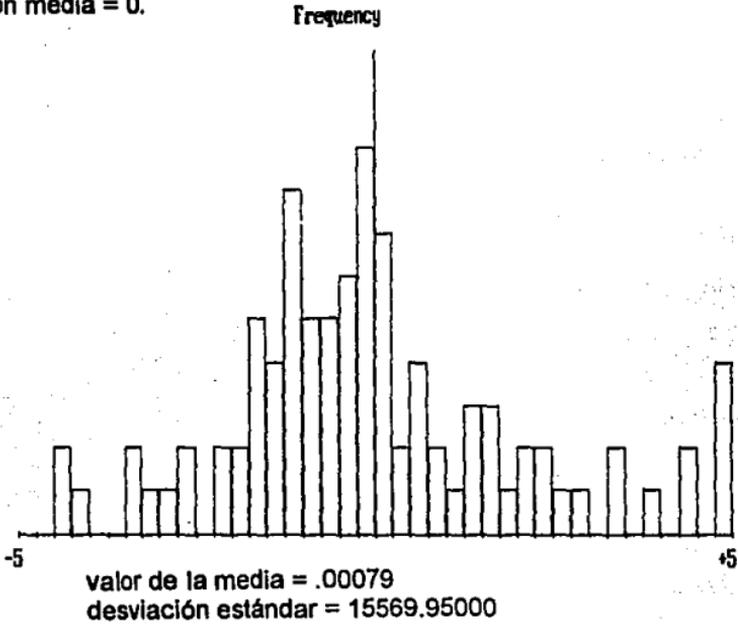
Max. Frequency : 15 in (1.00, 1.25)

sigue análisis de estacionalidad :
autocorrelograma



ACF:	.698	.221	-.194	-.406	-.435	-.299	-.038	.214	.376	.328
	.067	-.191	-.338	-.338	-.265	-.114	.138	.379	.451	.255
	-.031	-.214	-.251	-.198	-.086	.069	.186	.249	.174	-.004
	-.135	-.152	-.084	-.009	.068	.168	.193	.121	-.077	-.281
PACF:	.698	-.519	-.176	-.058	-.132	.029	.143	.046	.128	-.121
	-.228	.074	-.837	-.097	-.078	-.024	.181	.206	-.075	-.173
	.077	.101	.031	-.011	-.077	.070	-.006	.063	.033	-.009
	.065	.025	.052	.107	.012	-.011	-.065	-.065	-.091	-.132

Histograma y gráfica de los datos originales "desestacionalizados" y con media = 0.



- d) Obtener una representación ARMA para la serie estacionaria.
 e) Fijar los valores de "p" y "q": modelo ARIMA (2,0,0)

orden de autorregresión: 2.
 orden de promedios móviles: 0.

Order of autoregression to be fitted (p<27)? 2

Order of moving average to be fitted (q<27)? 0

varianza del ruido blanco

6.4.2. Estimación de los parámetros del modelo: ARIMA (2,0,0)

Tablas de resultados.

valor de phi 1 = 1.060

valor de phi 2 = -0.518940

AR COEFFICIENTS

1.0590428 -0.5189403

RATIO OF COEFFICIENTS TO (1.96*STANDARD ERROR)

6.1883708 -2.9869040

(Press any key to continue)

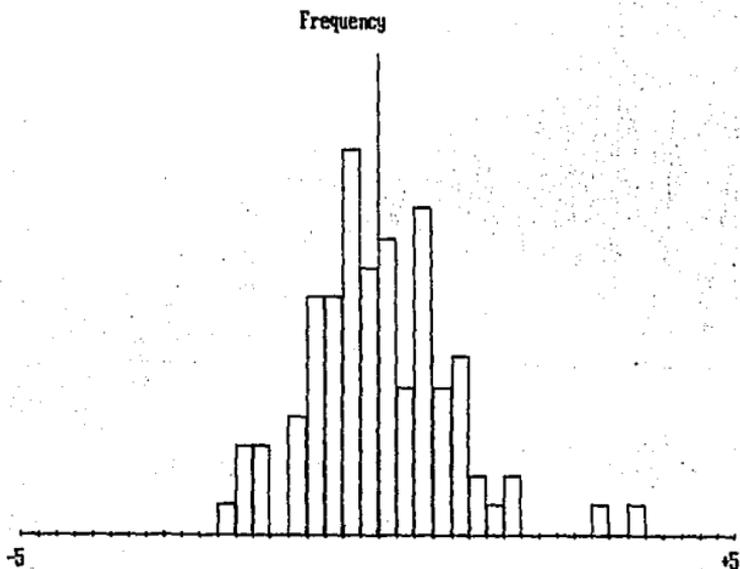
MIN VARIANCE ESTIMATE = (RESID SS)/N

.889903E+08

(YULE-WALKER MIN VARIANCE ESTIMATE = .988977E+08)

-2 ln(LIKELIHOOD) = 1967.49388008

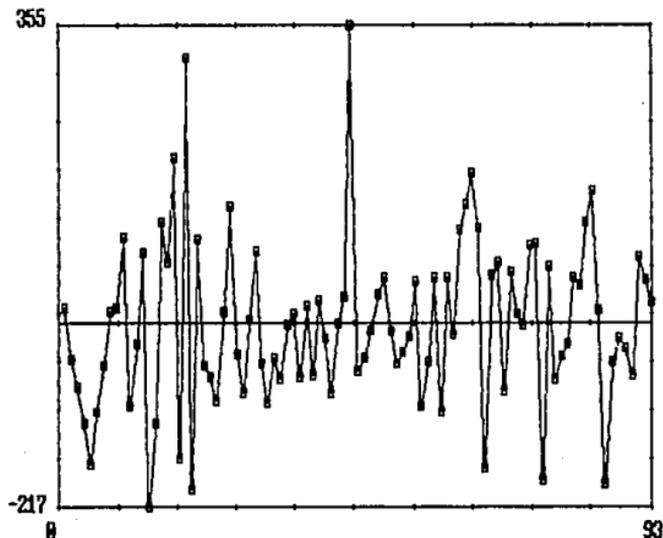
AICC STATISTIC = 1973.76288008

6.4.3. Verificación de supuestos:**a) Análisis de residuales:****Histograma de residuales:**

Horizontal Scale : 1 unit = 1 std. deviation

Max. Frequency : 13 in [-.50, -.25)

sigue análisis de residuales: varianza de residuales: ARIMA (2,0,0).



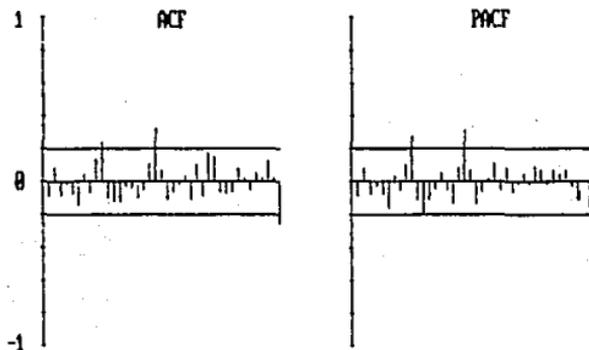
Vertical scale: 1 unit = .100000E+01;
 Max. on vertical scale = .354635E+01; Min. = -.216650E+01

Max. on vertical axis = 3.546349
 Min. on vertical axis = -2.166504

se observan valores "disparados", no se puede decir que sea una varianza monótona.

Mean of data plotted = .001324
 Std. deviation = .999999

autocorrelograma de residuales del modelo ARIMA (2,0,0)



ACF:	-.002	.072	-.002	-.008	-.072	-.141	.037	-.068	.133	.240
	-.097	-.118	-.113	-.023	-.035	-.091	-.042	.100	.322	.057
	-.111	-.065	-.024	.019	-.102	.090	-.001	.164	.141	-.063
	-.074	-.059	.069	.009	-.050	.046	.009	.122	.016	-.243
PACF:	-.002	.066	-.072	-.024	-.066	-.150	.020	-.052	.100	.275
	-.109	-.190	-.101	-.043	.047	-.044	-.132	.071	.305	.059
	-.126	-.056	.009	.102	-.053	.074	-.050	-.007	.034	-.015
	.003	.056	-.012	.054	.037	.060	-.020	-.106	-.002	-.157

En este autocorrelograma se observa que hay dos valores que rebazan el valor del umbral, mientras que en el autocorrelograma parcial se observa que los valores que rebazan el umbral son el 10 = .240, y el 19 = .322.

Esto sugiere que hay que evaluar la parte de los promedios móviles, considerar un modelo ARIMA (2,0,19), es decir, un modelo ARMA (2,19), con theta 10 = .1; y theta 19 = .1.

sigue análisis de residuales del modelo ARIMA (2,0,0)
 b) Pruebas de Bondad del ajuste: Tablas de resultados.

RANDOMNESS TEST STATISTICS (see section 9.4)

PORT'EAU with h= 20:	26.9140	chi sq	df= 18
TURNING POINTS =	52. ANORMAL(68.67	4.83**2)
DIFFERENCE-SIGN =	51. ANORMAL(46.00	2.80**2)
RANK TEST =	2463. ANORMAL(2139.00	451.97**2)

Obsérvese que el valor de la Prueba de Port'EAU o estadístico Q tiene un valor de 26.9140, que es mayor que el valor de $h=20$, y que los 18 grados de libertad.

Optimización de los parámetros del modelo ARIMA (2,0,0)

THE ARMA(2, 0) MODEL IS $X(t) = Z(t)$
 $+ (1.000)X(t- 1) + (-.519)X(t- 2)$

AUTOREGRESSIVE PARAMETERS :

PHI(1) =	1.85984200	PHI(2) =	-.51894838
WHITE NOISE VARIANCE =	.889983E+08		
-2 ln(LIKELIHOOD) =	1967.493000	LAST=	1967.493000
AICC STATISTIC =	1973.762000	LAST=	1973.762000

Ver las raíces de la ecuación característica, las raíces complejas indican ciclos. Las raíces no complejas indican tendencias.

6.4.4. Usos del modelo: pronósticos con el modelo ARIMA (2,0,0)

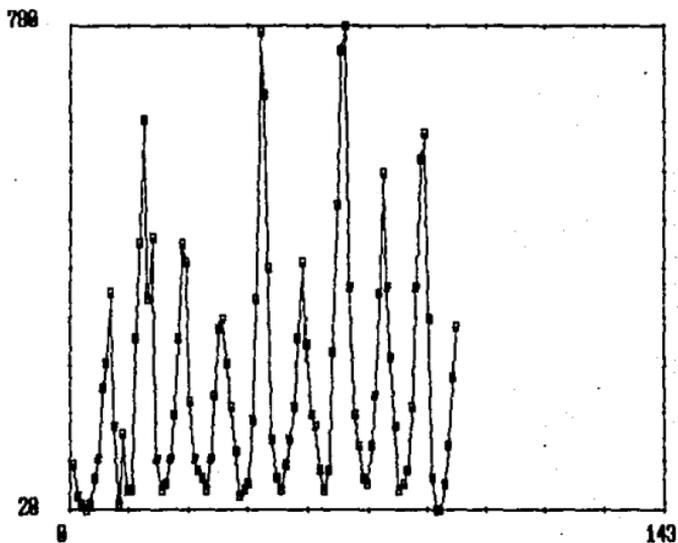
Phi 1 = 1.06; phi 2 = -0.519.

Tabla de valores estimados para el pronóstico:

#	XHAT	SQRT(MSE)	XHAT+MEAN (= .22110E+05)
94	.908324E+04	.943346E+04	.311932E+05
95	.108736E+04	.137459E+05	.231973E+05
96	-.356123E+04	.140812E+05	.165480E+05
97	-.433061E+04	.149857E+05	.177714E+05
98	-.275018E+04	.150465E+05	.193598E+05
99	-.663278E+03	.152728E+05	.214467E+05
100	.724211E+03	.153683E+05	.228342E+05
101	.111175E+04	.153758E+05	.232217E+05
102	.082458E+03	.153804E+05	.229124E+05
103	.273547E+03	.153948E+05	.223835E+05
104	-.126511E+03	.154028E+05	.219835E+05
(Press any key to continue)			
126	.267385E+00	.154857E+05	.221102E+05
127	.234598E+00	.154857E+05	.221102E+05
128	.109088E+00	.154857E+05	.221101E+05
129	-.528677E-02	.154857E+05	.221100E+05
130	-.626243E-01	.154857E+05	.221099E+05
131	-.636284E-01	.154857E+05	.221099E+05
132	-.349378E-01	.154857E+05	.221099E+05
133	-.408919E-02	.154857E+05	.221100E+05
134	.138815E-01	.154857E+05	.221100E+05
135	.167927E-01	.154857E+05	.221100E+05
136	.185940E-01	.154857E+05	.221100E+05
137	.251352E-02	.154857E+05	.221100E+05
138	-.283371E-02	.154857E+05	.221100E+05
139	-.438765E-02	.154857E+05	.221100E+05
140	-.389498E-02	.154857E+05	.221100E+05
141	-.104478E-02	.154857E+05	.221100E+05
142	.498857E-03	.154857E+05	.221100E+05
143	.187884E-02	.154857E+05	.221100E+05

sigue 6.4.4. Usos del modelo: pronósticos con el modelo ARIMA (2,0,0)

Phi 1= 1.06; phi 2 = -0.519.



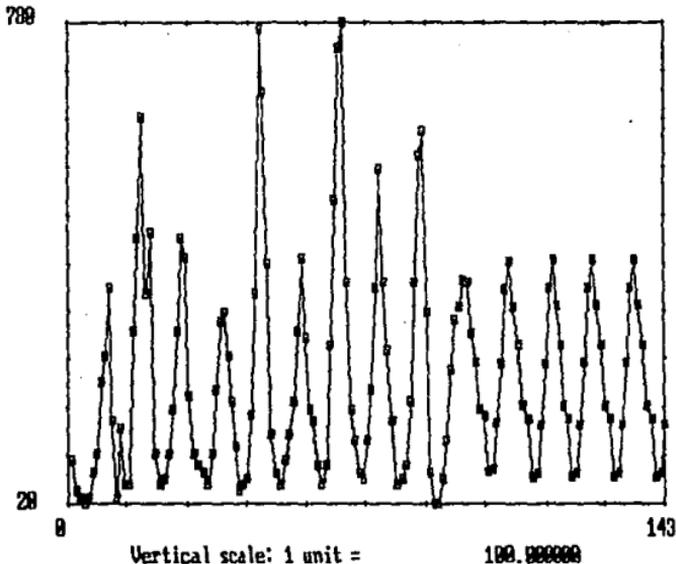
THE ARMA(2, 0) MODEL IS $X(t) = Z(t)$
 $+ (1.068)X(t-1) + (-.519)X(t-2)$

AUTOREGRESSIVE PARAMETERS :

PHI(1) = 1.05904288

PHI(2) =

-.51894838



Max. on vertical axis = 70000.000000

Min. on vertical axis = 2000.000000

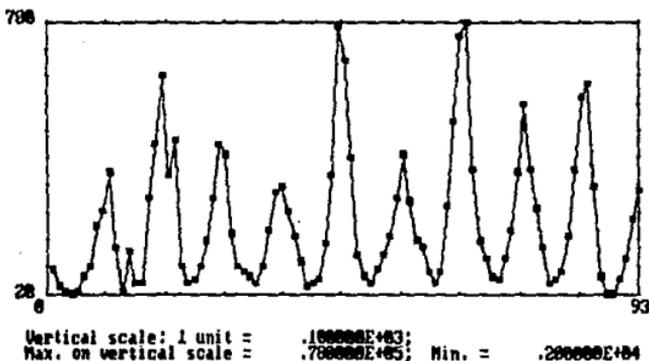
El pronóstico es bueno, pero con este modelo hay problemas que se señalan en los residuales ($Q = 26.914$, $gl = 18$).

Otro modelo: ARMA (2,19)

A partir de este punto, se ajusta un nuevo modelo a los datos originales, por lo cual, se repetirá toda la metodología señalada en el cuadro 1, a partir del punto 6.3.1.: "gráfica de los datos originales".

1a etapa: Análisis descriptivo.-

Gráfica de los datos originales presentada por el paquete pest.



.. Histograma y estadísticos de los datos originales.
(lincs 1821-1913).

NAME OF THE DATAFILE TO BE ANALYZED : lincs2.prn

TOTAL OBSERVATIONS = 93

CHECKING THE FIRST THREE AND LAST DATA POINTS:

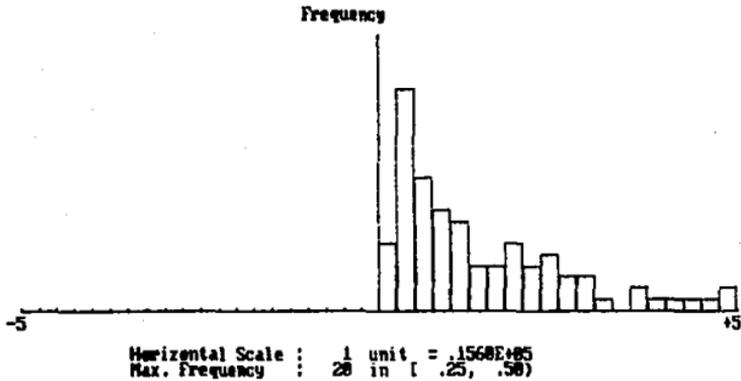
9000.000000

4000.000000

3000.000000

30991.000000

DATA FILE = lincs2.prn



valor de la media = 21692.13000
desviación estándar = 18603.4000

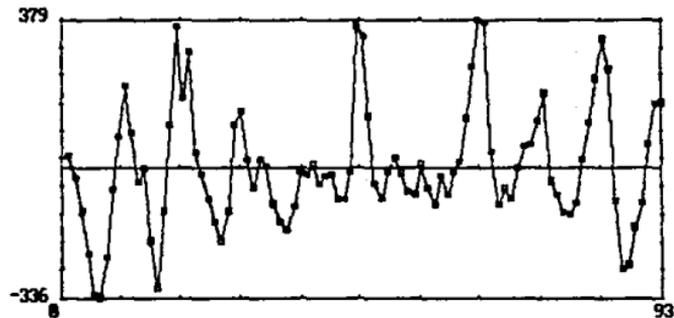
Análisis de estacionalidad (ciclos) a partir de los datos originales.

Tabla de los componentes estacionales.

```

PERIOD OF SEASONAL COMPONENT
  (enter period) 1, or 0 to escape: 10
SEASONAL COMPONENT 1: *****
SEASONAL COMPONENT 2: *****
SEASONAL COMPONENT 3: -7574.516*****
SEASONAL COMPONENT 4: 1837.56*****
SEASONAL COMPONENT 5: 14019.2*****
SEASONAL COMPONENT 6: 18456.*****
SEASONAL COMPONENT 7: 11856.7*****
SEASONAL COMPONENT 8: 4736.5*****
SEASONAL COMPONENT 9: -4599.512*****
SEASONAL COMPONENT 10: -6885.76*****
  
```

gráfica de los datos originales "desestacionalizados"

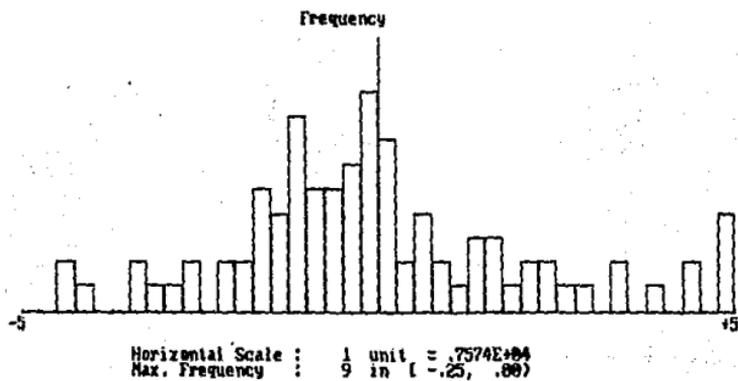


Vertical scale: 1 unit = .188888E+03;
 Max. on vertical scale = .378788E+03; Min. = -.335668E+05

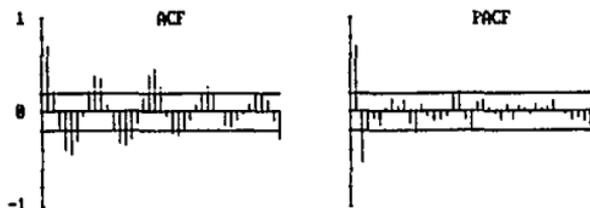
Max. on vertical axis = 37878.798888
 Min. on vertical axis = -33566.828888

valor de la media = 0.011614
 desviación estándar = 155690.95000

Histograma de los datos originales con análisis de estacionalidad ("desestacionalizados"):



sigue análisis de estacionalidad :
 autocorrelograma de los datos sin componente estacional, y con
 media cero. (desvío estándar =15569.95000).



ACF:	.698	.221	-.194	-.486	-.435	-.299	-.838	.214	.376	.328
	-.867	-.191	-.338	-.338	-.265	-.114	.138	.379	.451	.255
	-.831	-.214	-.251	-.198	-.886	-.869	.186	.249	.174	-.884
	-.135	-.152	-.864	-.889	-.868	-.160	.193	.121	-.877	-.281
PACF:	.698	-.519	-.176	-.858	-.132	-.829	.143	-.846	.128	-.121
	-.228	.874	-.837	-.897	-.878	-.824	.181	.286	-.875	-.173
	.877	.181	-.831	-.811	-.877	-.878	-.866	-.863	-.833	-.889
	.865	.825	.852	.187	.812	-.811	-.865	-.855	-.891	-.132

DATA FILE = lincas2.prn

,DE-SERS'D

· Obtener una representación ARMA para la serie estacionaria.
Fijar los valores de "p" y "q"

- Componente autorregresivo .-"p" = 2: (AR = 2); phi1, phi2.
- Componente de promedios móviles.- "q" = 19: (MA = 19);theta 1, theta 2,.....theta19.

Valores de los coeficientes considerados:

Phi (1) = .1
Phi (2) = .1
Theta (1) = 0
Theta (2) = 0
Theta (3) = 0
Theta (4) = 0
Theta (5) = 0
Theta (6) = 0
Theta (7) = 0
Theta (8) = 0
Theta (9) = 0
Theta (10) = .1
Theta (11) = 0
Theta (12) = 0
Theta (13) = 0
Theta (14) = 0
Theta (15) = 0
Theta (16) = 0
Theta (17) = 0
Theta (18) = 0
Theta (19) = .1

Estimación de los parámetros del modelo: ARMA (2,19)
Optimización de los parámetros del modelo ARMA (2,19)

Tablas de resultados.

valor de phi 1= 0.9740

valor de phi 2= -0.3969 *el signo negativo de este coeficiente de la parte autorregresiva del modelo, corresponde a los ciclos de la serie.

valor de theta 10 = 0.1929

valor de theta 19 = 0.4448 estos valores son interesantes y denotan una retrospección en esta parte de los promedios móviles asociada con los 10 y 19 años atrás.

$$+(.974)*X(t- 1) +(-.397)*X(t- 2) +(.193)*Z(t-10) +(.445)*Z(t-19)$$

AUTOREGRESSIVE PARAMETERS :

PHI(1) = .97403910 PHI(2) = -.39697740

MOVING AVERAGE PARAMETERS :

THETA(10)= .1929268 THETA(19)= .44488368

WHITE NOISE VARIANCE = .7029305400

BIC STATISTIC = 1964.362000

-2 ln(LIKELIHOOD) = 1958.172000

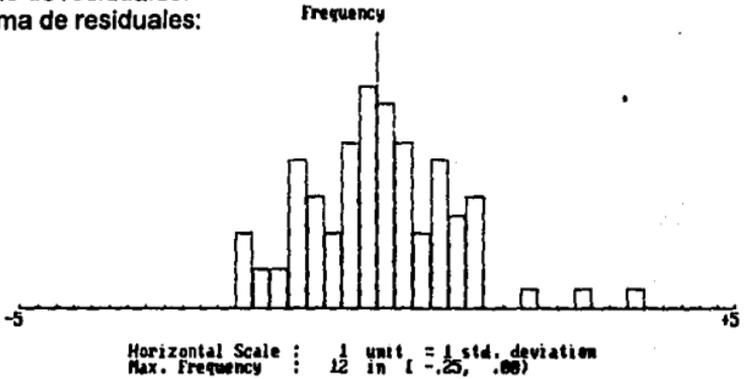
AIC STATISTIC = 1960.062000

FUNCTION CALLS = 60 ;# ITERATIONS = 7;ACCURACY PARAM.= .001010

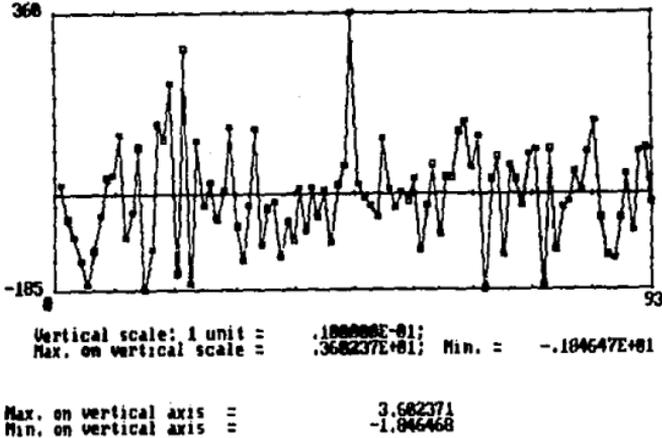
STOPPING VALUE 4 : ITERATION LIMIT EXCEEDED

CONVERGENCE OCCURRED AT STEP 90 OF 93; VALUE OF R(N) = 1.0007430

Verificación de supuestos:
 a) Análisis de residuales:
 Histograma de residuales:



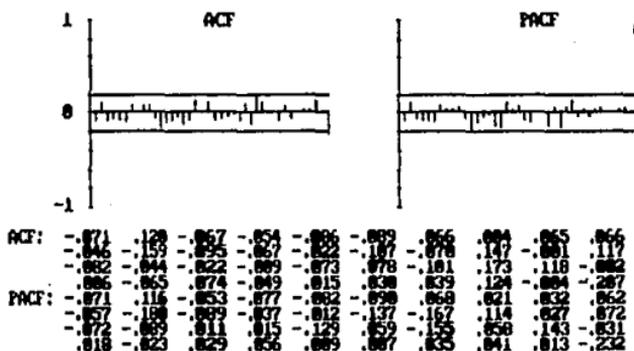
varianza de residuales



media = 0.0018
 desviación estándar = 0.9999

En términos generales se puede decir que la varianza es monótona. Hay que corroborar esta afirmación con las Pruebas de Bondad del Ajuste...

autocorrelograma de residuales del modelo ARMA (2,19)



Observamos un autocorrelograma de residuales muy decente, donde no hay valores que sobresalgan de los límites del umbral. ¡BRAVO!!!.

b) Pruebas de la Bondad del ajuste:

El modelo es :

```

THE ARMA(2,19) MODEL IS  $X(t) = Z(t)$ 
+ (.973)*X(t-1) + (-.399)*X(t-2) + (.192)*Z(t-10) + (.444)*Z(t-19)

AUTOREGRESSIVE PARAMETERS :
PHI(1) = .97307000          PHI(2) = -.39891550

MOVING AVERAGE PARAMETERS :
THETA(10) = .19199360      THETA(19) = .44391460

WHITE NOISE VARIANCE = .783164E+00

CONVERGENCE OCCURRED AT STEP 98 OF 93; VALUE OF R(N) = 1.0007240
The MINV has been set to (RESIDUAL SS)/n = .783164E+00

WHITE NOISE VARIANCE == .782930E+00
BIC STATISTIC          == 1964.342000
-2 ln(LIKELIHOOD)      == 1960.172000
AICC STATISTIC         == 1960.862000

```

Tablas de Pruebas de Bondad del Ajuste:

RANDOMNESS TEST STATISTICS (see section 9.4)

PORT' TEAU with h= 20:	13.7893	chi sq	df= 16	
TURNING POINTS =	60.	ANORMAL(60.67	4.83**2)
DIFFERENCE-SIGN =	50.	ANORMAL(46.00	2.80**2)
RANK TEST =	2363.	ANORMAL(2139.00	451.97**2)

(Press any key to continue)

Nótese que el valor del estadístico de Q (Prueba de Port' teau) es de $Q = 13.789$, con 16 grados de libertad. (¡muy bien!)

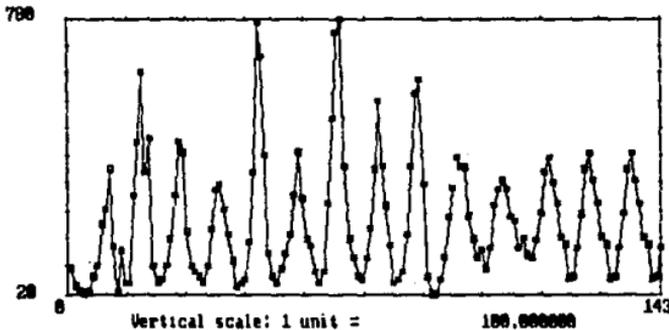
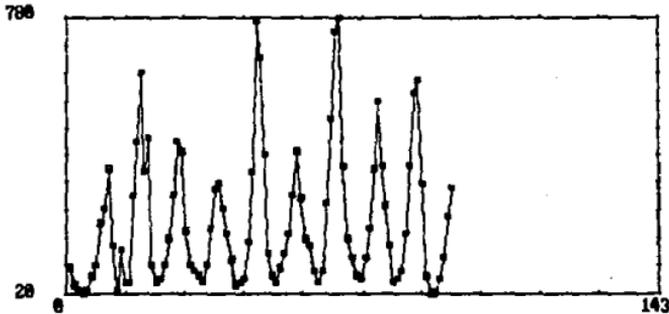
Pronósticos con el modelo ARMA (2,19)

-Tabla de valores estimados para el pronóstico:

#	XHAT	SERT(NSE)	XHAT+MEDIAN (= ,22118E+85)
94	.152147E+85	.838848E+84	.272247E+85
95	.668369E+83	.117824E+85	.227832E+85
96	-.377473E+84	.125718E+85	.183352E+85
97	-.188238E+85	.126301E+85	.128842E+85
98	-.991298E+84	.126464E+85	.121971E+85
99	-.543591E+84	.126353E+85	.164641E+85
100	-.101517E+84	.127221E+85	.218048E+85
101	.282888E+84	.127221E+85	.243811E+85
102	.183821E+84	.127221E+85	.222482E+85
103	.116759E+85	.127221E+85	.378591E+85
104	.649636E+84	.128463E+85	.286863E+85
126	-.741223E+81	.141796E+85	.221826E+85
127	-.253776E+81	.141796E+85	.221874E+85
128	.487433E+80	.141796E+85	.221185E+85
129	.148766E+81	.141797E+85	.221115E+85
130	.125218E+81	.141797E+85	.221112E+85
131	.625487E+80	.141796E+85	.221186E+85
132	.189831E+80	.141796E+85	.221181E+85
133	-.143371E+88	.141797E+85	.221898E+85
134	-.183812E+88	.141797E+85	.221898E+85
135	-.128098E+88	.141797E+85	.221899E+85
136	-.446286E+81	.141797E+85	.221899E+85
137	.478277E+82	.141797E+85	.221188E+85
138	.224721E+81	.141797E+85	.221188E+85
139	.193528E+81	.141797E+85	.221188E+85
140	.184518E+81	.141797E+85	.221188E+85
141	.221889E+82	.141796E+85	.221188E+85
142	-.281834E+82	.141796E+85	.221188E+85
143	-.284578E+82	.141797E+85	.221188E+85

sigue Pronósticos con el modelo ARMA (2,19)

- Gráficas:



The datafile is lincos2.prn ; Total data points= 93
 The data was deseasonalized with period 10

The subtracted mean is 22189.9000

THE ARMA(2,19) MODEL IS $X(t) = Z(t) + (.973)*X(t-1) + (-.399)*X(t-2) + (.192)*Z(t-10) + (.444)*Z(t-19)$

White noise variance = .783164E+08

$+ (.974)*X(t-1) + (-.397)*X(t-2) + (.193)*Z(t-10) + (.445)*Z(t-19)$

AUTOREGRESSIVE PARAMETERS :

PHI(1) = .97483918

PHI(2) = -.39697748

MOVING AVERAGE PARAMETERS :

THETA(10)= .19296260

THETA(19)= .44488360

Resumen de Resultados de la Estimación de Modelos ARIMA

Serie: Z_t = Linces.

Transformación: no

valor original de la media = 2169.213
desvío estándar original = 18603.4

-----Análisis de Residuales-----

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Período y número de observaciones	Modelo	Parámetros estimados	Intervalos del 95% de Confianza	Correlación entre parámetros >0.5 o <0.5	$m(\hat{\alpha})$	$\tilde{\sigma}_a$	$Q', g.l.$	r_k	Residuales grandes
1821-1913 N=93	AR(2) = ARIMA (2,0,0)	$\hat{\phi}_1 = 1.60$ $\hat{\phi}_2 = -0.519$			$m = 0.001$... = 0.8899	$Q = 26.91$ para 18 g.l.		SI. valores 10 y 19 de ACF de residuales mayores que el umbral.
1821-1913 N=93	ARMA (2,19) = ARIMA (2,0,19)	$\hat{\phi}_1 = 0.973$ $\hat{\phi}_2 = -0.39$ $\hat{\theta}_1 = 0.192$ $\hat{\theta}_{19} = 0.44$			$m = 0.0116$	= 0.7031	$Q = 13.789$ para 16 g.l.		NO.

6.6. Discusión de resultados.

Como se señala en la gráfica de datos originales, se puede apreciar que se trata de una serie estacionaria dado que no presenta una tendencia. En esta misma gráfica observamos que en la serie hay ciclos, por lo tanto es una serie con clara estacionalidad. Los ciclos tienen una periodicidad de 9.6 años aproximadamente, y en el intervalo estudiado, de 1824 a 1913 (89 años) hay 8 ciclos aproximadamente.

El modelo que mejor se ajusta a los datos es el ARMA (2,19), como se puede apreciar en la tabla anterior, donde se comparan los resultados de los modelos estimados. El análisis de residuales del modelo ARMA (2, 19) señala que la varianza de residuales tiene un valor de 0.7029, la gráfica de varianza de residuales es monótona, con un sólo valor que se "dispara" ("outlier"), el autocorrelograma de residuales no presenta valores mayores al límite del umbral, el estadístico de $Q = 13.789$ es menor a sus 16 grados de libertad, y el modelo dá un buen pronóstico.

El pronóstico del primer modelo que se estimó, el ARMA (2,0), no es un mal pronóstico, sin embargo tiene problemas que se manifiestan en el análisis de residuales: La varianza de residuales tiene un valor de 0.8899 que es mayor que la varianza de residuales del modelo ARMA 2,19, lo que indica que el modelo ARIMA(2,0,0) deja mayor variabilidad sin explicar que el modelo ARMA (2,19). Su gráfica de varianza de residuales es no monótona, hay valores que rebasan el umbral en el autocorrelograma de residuales. Y en cuanto a los resultados de las pruebas de la Bondad del ajuste, el estadístico de Q es mayor a los grados de libertad : $Q = 26.9140$, $gl = 18$.

El valor del parámetro "d" se consideró igual a cero, pues aunque al aplicarle una diferencia a los datos la desviación estándar disminuyó, por parsimonia se decidió emplear el menor número de parámetros en el modelo. Cabe señalar que el valor "q = 19", no es muy parsimonioso, sin embargo, los dos elevados valores del ACFP de los residuales del modelo ARIMA (2,0,0), indicaron la necesidad de considerar el componente de promedios móviles (AM), y asignarle valores en los parámetros theta 10 y theta 19.

6.7. Conclusiones.

El modelo ARIMA (2,0,19) obtenido -o ARMA (2,19)-, es un modelo explicativo que resume todo el proceso en función de su historia. Los parámetros del modelo ARMA (2,19) tienen un importante significado biológico. Los ciclos de la serie aparecen en la parte autorregresiva (AR) del modelo, por el signo negativo del coeficiente ϕ_2 . También se nota una retrospcción en la parte de los promedios móviles asociada con los 10 y 19 años atrás, (θ_{10} y θ_{19} tienen valores distintos de cero).

Al igualar el valor de la media con cero, queda bien especificada la parte estocástica del proceso. La varianza, es la varianza de residuales que indica la variabilidad que no se puede explicar con el modelo. Tiene sentido sólo en comparación con la varianza residual de otro modelo.

El modelo obtenido sirve esencialmente como una base de comparación para otros posibles modelos explicativos, como ya se mencionó, aquí se tiene todo el proceso en función de su historia. Esta clase de modelos son de los más eficientes para pronosticar unos cuantos periodos adelante.

Finalmente, cabe señalar que los ciclos están asociados con la dinámica de la población en estudio.

- Capítulo 7: *segunda serie.*-

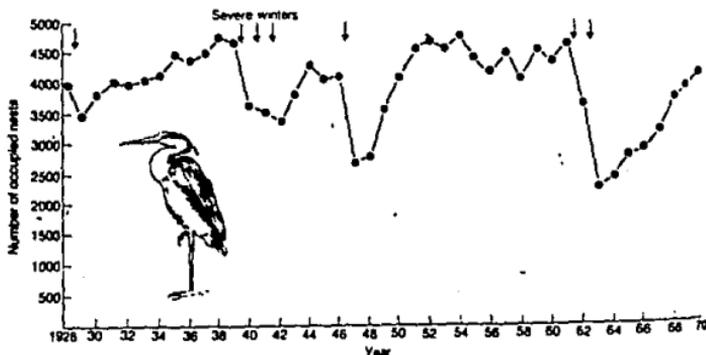
Fase Experimental 1: Comprobación de hipótesis.

Aplicación del análisis estadístico de series de tiempo univariadas a un conjunto de datos de un proceso estocástico ecológico: serie estacionaria:

El caso de las garzas inglesas y galesas "Ardea cinerea".

En 1971, Stafford publicó los siguientes datos sobre la variación en la abundancia poblacional de las garzas "Ardea cinerea" en Inglaterra y Gales, medida como el número de nidos ocupados (Begon, Harper y Townsend 1986):

7.1 Gráfica de los cambios en la abundancia anual de "Ardea cinerea", de 1928 a 1970.



7.2 Tabla de datos.
Cambios en la abundancia poblacional de Garzas "Ardea cinerea" en Inglaterra y Gales.

<u>año</u>	<u>número de nidos ocupados</u>
1928	4000
1929	3400
1930	3800
1931	4100
1932	4000
1933	4100
1934	4200
1935	4500
1936	4400
1937	4550
1938	4650
1939	4600
1940	3600
1941	3500
1942	3400
1943	3800
1944	4250
1945	4000
1946	4100
1947	2600
1948	2800
1949	3500
1950	4000
1951	4500
1952	4600
1953	4500
1954	4700
1955	4400
1956	4100
1957	4450
1958	4000
1959	4500
1960	4400
1961	4600
1962	3600
1963	2100
1964	2400
1965	2800
1966	2950
1967	3200
1968	3800
1969	4000
1970	4200

7.3. Análisis de los datos y construcción del modelo ARIMA mediante el uso del paquete PEST: Metodología. la etapa: Análisis descriptivo.

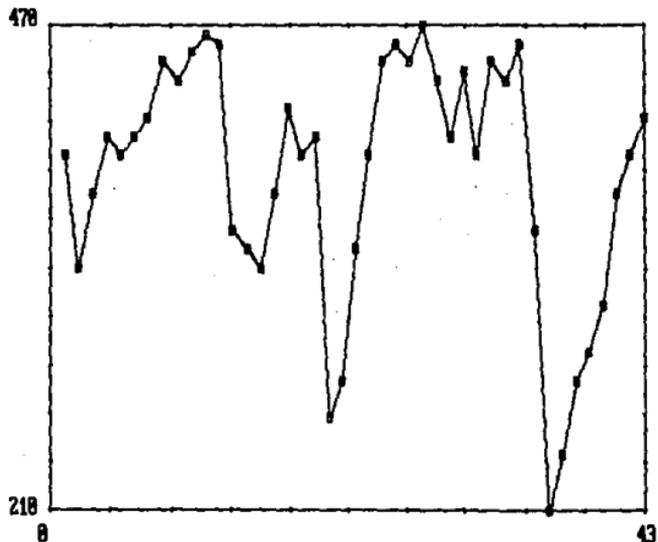
7.3.1a etapa: Análisis descriptivo:

7.3.1. Gráfica de los datos originales.

TOTAL OBSERVATIONS = 43

CHECKING THE FIRST THREE AND LAST DATA POINTS:

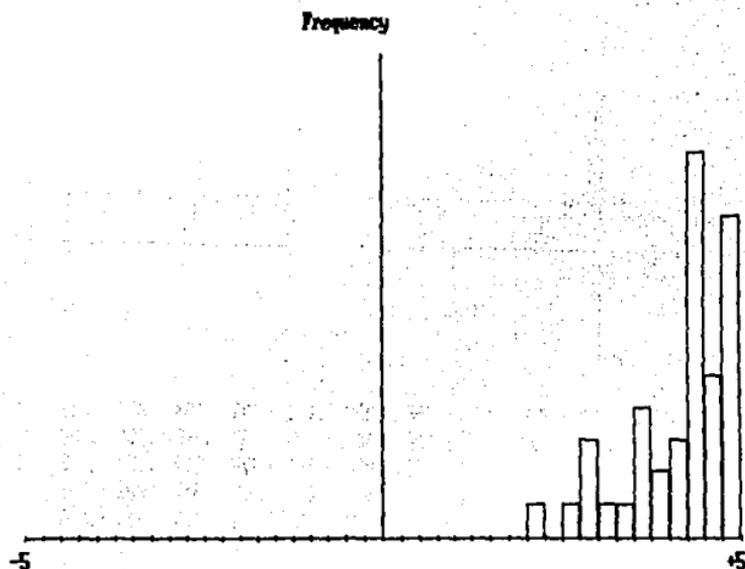
4000.000000 3400.000000 3000.000000 2200.000000



Vertical scale: 1 unit = .100000E+02;
 Max. on vertical scale = .470000E+04; Min. = .210000E+04

Max. on vertical axis = 4700.000000
 Min. on vertical axis = 2100.000000

7.3.2. Histograma y estadísticos.



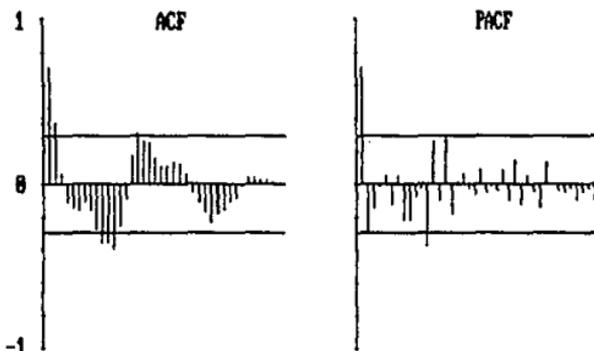
Horizontal Scale : 1 unit = .9488E+03

Max. Frequency : 12 in [4.25, 4.50)

media = 3898.83

desviación estándar = 649.37

7.3.3. Autocorrelograma y autocorrelograma parcial de los datos originales.



ACF:	.718	.363	.064	-.108	-.139	-.152	-.109	-.153	-.270	-.353
	-.352	-.307	-.248	-.006	.169	.305	.259	.248	.151	.100
	.103	.134	.123	.055	-.050	-.104	-.170	-.218	-.182	-.150
	-.101	-.001	-.013	.041	.033	.025	.027	.017	.002	-.005
PACF:	.718	-.206	-.143	-.012	.049	-.123	.051	-.215	-.213	-.076
	-.021	-.373	.261	-.092	.300	-.101	.010	.061	-.024	-.056
	.070	-.040	-.006	-.036	.000	-.096	.130	-.112	.012	-.046
	-.130	.125	-.002	-.035	-.040	-.010	-.097	-.049	-.029	-.090

7.4. Metodología para la construcción del modelo ARIMA de la serie de tiempo: IIa. etapa: Construcción del modelo:

7.4.1. Identificación del modelo:

a) Determinar una serie estacionaria en función de los datos originales:

Estabilización de la varianza, comparando diferencias. Determinar el valor de "d".

1a. Diferencia:

Histograma

DIFFERENCE AT LAG : 1

TOTAL OBSERVATIONS = 42

CHECKING THE FIRST THREE AND LAST DATA POINTS:

-600.000000

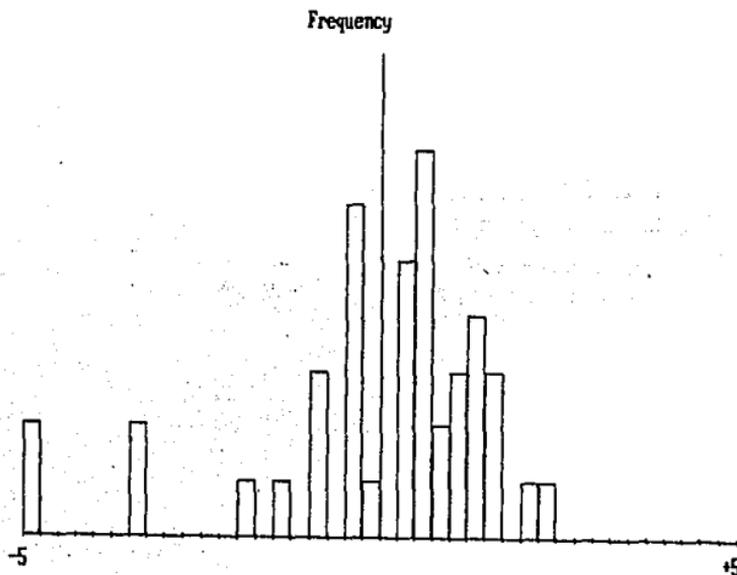
400.0002000

299.9990000

200.0000000

DATA FILE = datosgr

,DIFF'D

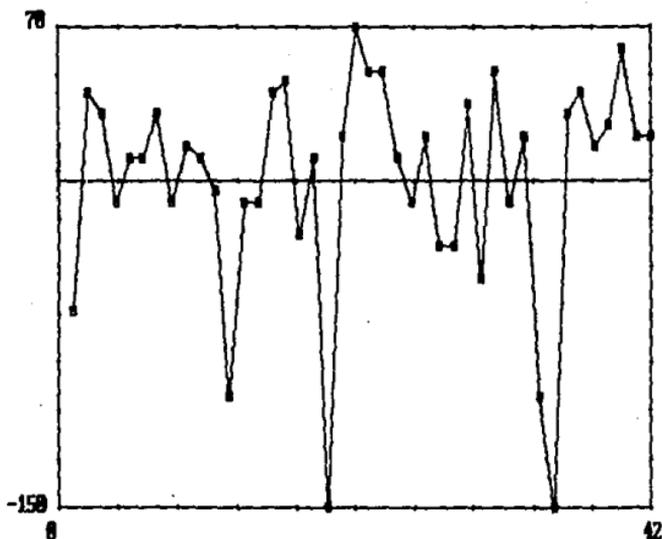


Horizontal Scale : 1 unit = .3000E+03

Max. Frequency : 7 in [.58, .75)

b) Sustraer la media...

Gráfica de los datos originales con una diferencia y media = 0



Vertical scale: 1 unit = .100000E+02;
 Max. on vertical scale = .700000E+03; Min. = -.150000E+04
 Mean of data plotted = -.000005
 Std. deviation = 497.59600

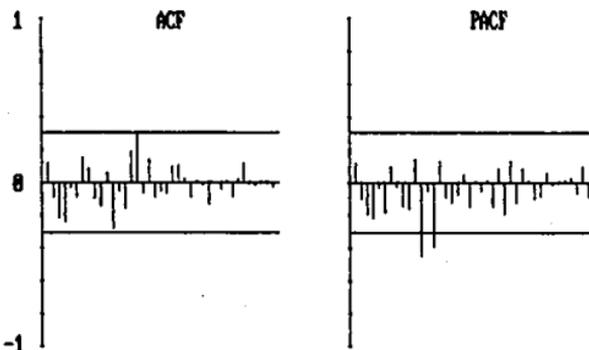
media = -0.0000

Desviación estándar = 497.59

Autocorrelograma de la 1a. diferencia, media = 0

DATA FILE = datosgr

,DIFF'D



ACF:	.112	-.082	-.288	-.237	-.027	-.066	.159	.004	-.098	-.141
	.068	-.278	-.053	-.152	.193	.386	-.054	.147	-.079	-.052
	-.057	.009	.181	.021	-.005	.013	-.017	-.131	.002	-.029
	.017	-.065	.027	.128	.012	-.013	.018	.012	-.022	-.004
PACF:	.112	-.095	-.192	-.211	-.022	-.174	.096	-.019	-.145	-.152
	.137	-.447	-.053	-.392	.129	-.003	-.114	-.065	.051	-.139
	.015	-.043	.000	-.141	.004	-.194	.133	-.120	.001	.014
	-.091	-.008	.059	-.000	.015	.016	.029	-.059	.090	-.079

- c) Obtener una representación ARMA para la serie estacionaria.
 d) Fijar los valores de "p" y "q": **ARIMA (0,1,1)**.

```

Input the order of Autoregression (<?>):
  (input -i if model to be read from file): 0

Input the order of Moving Average (<?>):1

Input the starting values for the coefficients
THETA( 1)= .1

THE ARMA( 0, 1) MODEL IS X(t) = Z(t)
+( .100)*Z(t- 1)

White noise variance = .100000E+01
(Press any key to continue)
  
```

7.4.2. Estimación de los parámetros del modelo: Tablas de resultados.

```

Order of moving average to be fitted (q(<?>)? 1

MA COEFFICIENTS
.1487715
RATIO OF COEFFICIENTS TO (1.96*STANDARD ERROR)
.4919131
(Press any key to continue)

WN VARIANCE ESTIMATE = (RESID SS)/N
.243894E+06

(INNOVATION WN VARIANCE ESTIMATE = .132425E+061

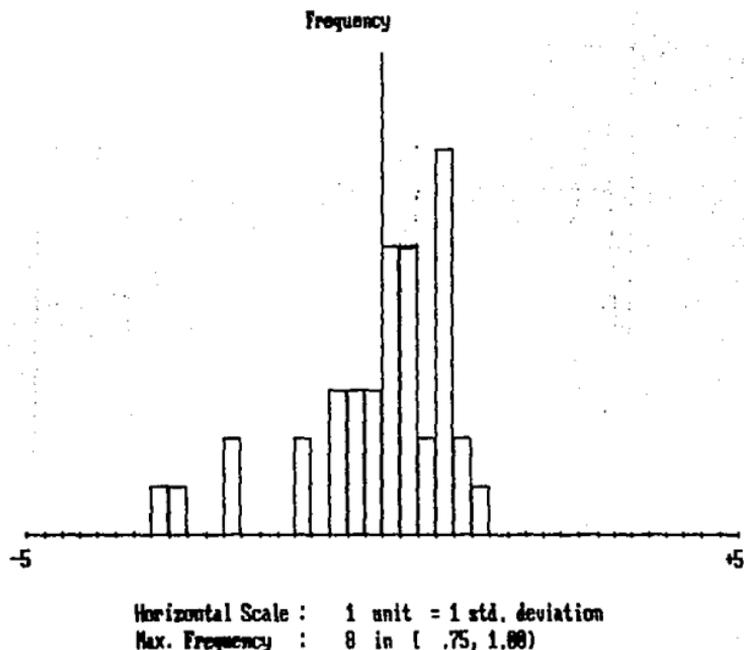
-2 ln(LIKELIHOOD) = 648.28100000
AICC STATISTIC = 644.58950000 PRIOR BEST= 644.58310000
  
```

varianza del ruido blanco = 0.2438

7.4.3. Verificación de supuestos:

a) Análisis de residuales: Gráficas.

Histograma de residuales ARIMA (0,1,1)

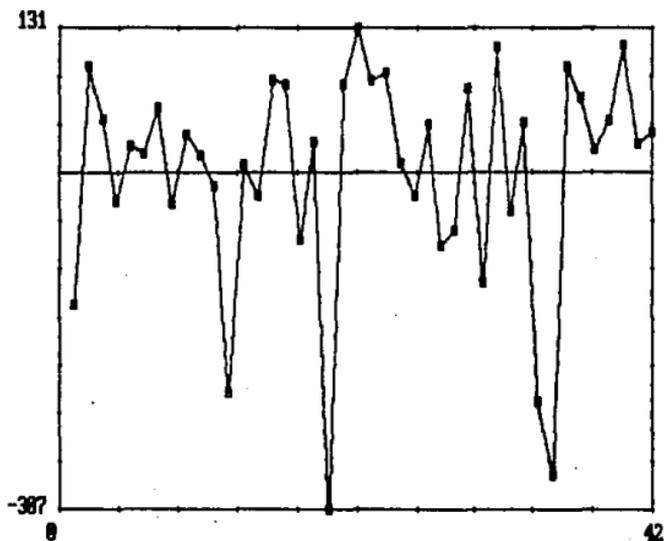


THE ARIMA (0, 1) MODEL IS $X(t) = Z(t)$
 $+ (.130)Z(t-1)$

MOVING AVERAGE PARAMETERS :
 THETA (1) = .12994578

WHITE NOISE VARIANCE = .243643E+06
 BIC STATISTIC = 648.728488
 $-2 \ln(\text{LIKELIHOOD})$ = 648.187588
 AICC STATISTIC = 644.495288

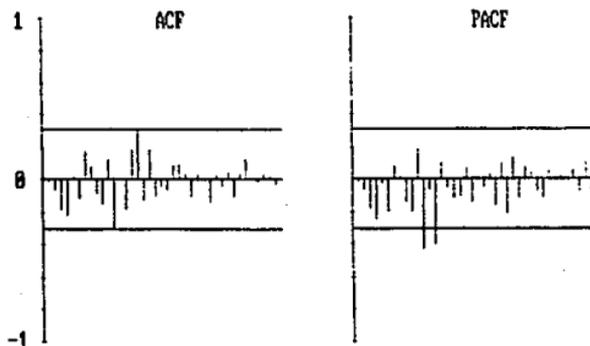
Gráfica de residuales ARIMA (0,1,1)



Vertical scale: 1 unit = .100000E+01;
Max. on vertical scale = .131382E+01; Min. = -.387427E+01

media = 0.0012
Desvío estándar = 0.999

Autocorrelograma de residuales: ARIMA (0,1,1), media = 0



ACF: -.000 -.058 -.172 -.216 .015 -.100 .164 .071 -.003 -.144
 .115 -.292 .006 -.174 .177 .297 -.113 .172 -.097 -.030
 -.063 .000 .000 .018 -.090 .024 -.002 -.133 .023 .036
 .031 -.092 .024 .116 -.001 -.014 .018 .013 -.023 .001
 PACF: -.000 -.058 -.173 -.231 -.024 -.104 .075 .007 -.124 -.190
 .171 -.426 -.058 -.406 .009 -.049 -.111 -.009 .064 -.130
 .005 -.042 .024 -.151 .094 -.197 .126 -.110 .066 .030
 -.071 -.100 .052 -.003 .016 .009 .042 -.072 .093 -.063

b) Pruebas de Bondad del ajuste: Tablas de resultados.

RANDOMNESS TEST STATISTICS (see section 9.4)

PORT'EAU with h= 20: 18.9745 chi sq df= 19

TURNING POINTS = 30. ANORMAL(26.67 2.67**2)

DIFFERENCE-SIGN = 19. ANORMAL(29.58 1.89**2)

RANK TEST = 456. ANORMAL(430.58 130.41**2)

7.4.4. Usos del modelo: pronósticos.

Tabla de los valores estimados en el pronóstico con ARIMA (0,1,1)

MOVING AVERAGE PARAMETERS :

THETA(1)= .12989828

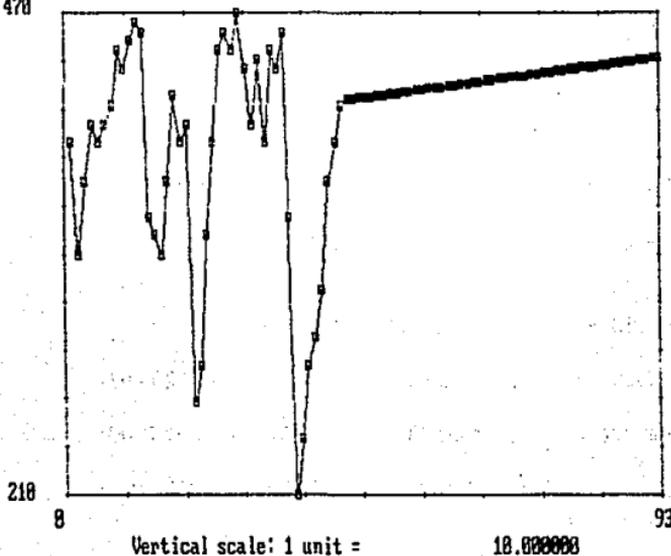
WHITE NOISE VARIANCE = .243836E+06

CONVERGENCE OCCURRED AT STEP 5 OF 42; VALUE OF R(N) = 1.000000

The UNV has been set to (RESIDUAL SS)/n = .243836E+06

#	XHAT	SQRT(MSE)	XHAT+MEAN (= .47619E+01)
43	.233848E+02	.493797E+03	.288659E+02
44	.000000E+00	.497945E+03	.476190E+01
45	.000000E+00	.497945E+03	.476190E+01
46	.000000E+00	.497945E+03	.476190E+01
47	.000000E+00	.497945E+03	.476190E+01
48	.000000E+00	.497945E+03	.476190E+01
49	.000000E+00	.497945E+03	.476190E+01
50	.000000E+00	.497945E+03	.476190E+01
51	.000000E+00	.497945E+03	.476190E+01
52	.000000E+00	.497945E+03	.476190E+01
53	.000000E+00	.497945E+03	.476190E+01

470



7.4.5. Resultados con otros modelos: ARIMA (0,1,14)

Input the order of Autoregression ((27):
 (input -1 if model to be read from file): 0

Input the order of Moving Average ((27):14

Input the starting values for the coefficients

THETA(1)= 0

THETA(2)= 0

THETA(3)= 0

THETA(4)= 0

THETA(5)= 0

THETA(6)= 0

THETA(7)= 0

THETA(8)= 0

THETA(9)= 0

THETA(10)= 0

THETA(11)= 0

THETA(12)= .1

THETA(13)= 0

THETA(14)= .1

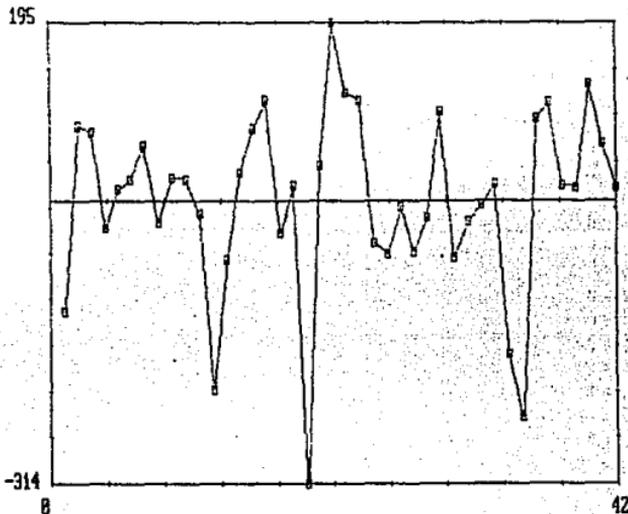
THE ARMA(0,14) MODEL IS $X(t) = Z(t)$
 $+ (-.412)*Z(t-12) + (-.329)*Z(t-14)$

MOVING AVERAGE PARAMETERS :

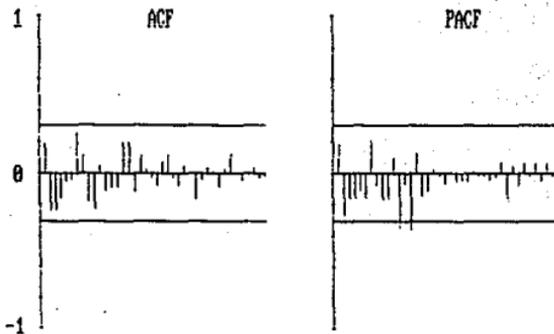
THETA(12)= -.41178570 THETA(14)= -.32921228

WHITE NOISE VARIANCE = .185564E+06
 BIC STATISTIC = 634.548500
 -2 ln(LIKELIHOOD) = 633.115000
 AICC STATISTIC = 639.746600

siguen resultados con otros modelos:
Análisis de residuales del modelo ARIMA (0,1,14)
 varianza:



autocorrelograma de los residuales



ACF:	.172	-.237	-.238	-.151	-.045	-.038	.242	.185	-.181	-.227
	.058	-.103	-.077	-.079	.104	.183	-.121	.107	.025	-.029
	-.071	.076	.111	-.026	-.066	.049	-.005	-.159	-.034	.033
	.012	-.005	.029	.122	.004	-.034	.016	.038	-.024	-.016
PACF:	.172	-.274	-.155	-.159	-.107	-.152	.202	-.067	-.161	-.148
	.096	-.346	-.064	-.356	.114	-.138	-.104	.025	-.011	-.058
	.018	-.046	-.035	-.045	.012	-.015	-.013	-.038	-.022	.069
	-.172	.042	-.070	.058	-.016	.058	-.046	.058	-.018	.008

Pruebas de la Bondad del Ajuste:

RANDOMNESS TEST STATISTICS (see section 9.4)

PORT'EAU with h= 20: 10.5865 chi sq df= 18

TURNING POINTS = 21. ANORMAL(26.67 2.67**2)

DIFFERENCE-SIGM = 21. ANORMAL(28.50 1.89**2)

RANK TEST = 445. ANORMAL(430.50 138.41**2)

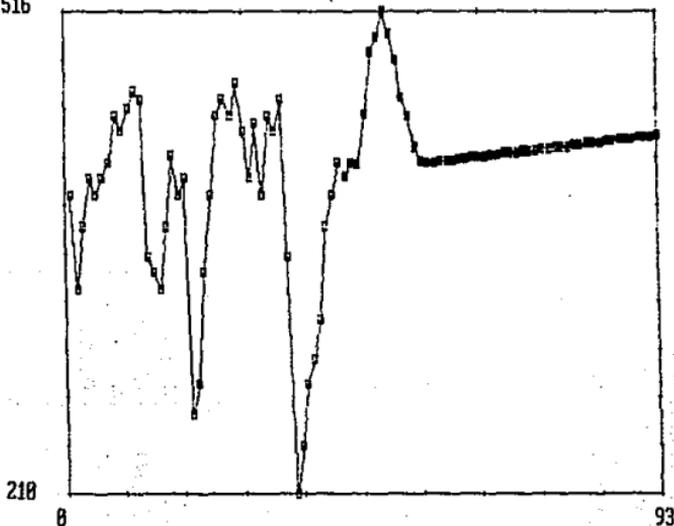
Pronósticos con ARIMA (0,1,14)

MOVING AVERAGE PARAMETERS :

THETA(12)= -.41178578 THETA(14)= -.32921228

WHITE NOISE VARIANCE = .185564E+06

516



The subtracted mean is 4.7619

THE ARMA(0,14) MODEL IS $X(t) = Z(t)$
 $+ (-.412)*Z(t-12) + (-.329)*Z(t-14)$

White noise variance = .185564E+06

Resumen de Resultados de la Estimación de Modelos ARIMA
 7.5 Tabla comparativa de los modelos estimados para la serie de las garzas.

Serie: Z_t = Garzas.

Transformación: no

valor de la media original = 3898.83
 desvío estándar original = 649.37

-----Análisis de Residuales-----

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Período y número de observaciones	Modelo	Parámetros estimados	Intervalos del 95% de Confianza	Correlación entre parámetros >0.5 o <0.5	$m(\hat{a})$	$\bar{\sigma}_a$	$Q', g.l.$	r_k	Residuales grandes
1928 - 1970									
N= 43	ARIMA (0,1,1)	$d = 1$ $\theta 1 = 0.130$			$m = 0.0012$	$\bar{\sigma}_a = 0.2438$	$Q = 18.97$ g.l.=19		SI varianza no monótona. PACF con valores 12 y 14 mayores que el umbral.
N= 43	ARIMA (0,1,14)	$d = 1$ $\theta 12 = -0.4117$ $\theta 14 = -0.3292$			$m = 0.0064$	$\bar{\sigma}_a = 0.1855$	$Q = 18.758$ g.l. =18		SI Varianza no monótona. PACF con valores 12 y 14 mayores que el valor umbral

7.6. Discusión de resultados.

Pese a que se probaron muchas opciones de modelos ARIMA, no es posible ajustar un modelo a este conjunto de datos porque esta serie es una serie con pocos datos, $n = 34$, y con muchas intervenciones climáticas. Se requieren series con más de 50 datos. Se presentan dos modelos de la larga serie de modelos que se intentaron, Sin embargo, los dos modelos que se presentan tienen problemas tanto con los residuales, como con las pruebas de Bondad, y sus pronósticos no son buenos, pese a que el modelo ARIMA (0,1,14) mejora el pronóstico.

Para ajustar un modelo en casos como estos, es necesario tener series con más datos para llevar a cabo un análisis de intervención, identificando los cortes en la serie con eventos significativos ocurridos sobre esta variable en su historia. Sin embargo, no es posible efectuar dicho análisis con el paquete PEST.

7.7. Conclusiones.

En series con tan pocos datos y tantas intervenciones, no es factible realizar un ajuste. En series con intervenciones climáticas y mas datos, se realiza un análisis de intervenciones comparando la serie original con una "serie patrón", que se obtiene realizando cortes en los intervalos entre las intervenciones, y se analiza mediante el análisis de intervenciones. En este caso, las intervenciones son tantas y los datos tan pocos, que no es posible obtener una serie patrón.

Es importante sin embargo, presentar el análisis que se realizó de esta serie porque ejemplifica a una serie con intervenciones. En este caso, las intervenciones corresponden a fuertes inviernos con intervalos sumamente cortos, Las intervenciones se señalan como inviernos severos, con 6 flechitas verticales en la gráfica de la página 152. Para realizar un análisis de intervenciones es necesario tener series mayores y/o con intervenciones menos frecuentes.

- Capítulo 8: *tercera serie.*-

Fase experimental 2: Comprobación de hipótesis.

Aplicación del análisis estadístico de series de tiempo univariadas a un conjunto de datos de un proceso estocástico evolutivo: serie con tendencia:

Los radiolarios Pseudocobus vema.

Análisis del patrón de cambio filogenético de los radiolarios "Pseudocobus vema" observado en las sucesiones cronológicas poblacionales de su registro fósil.

La historia geológica tiene pocos grupos fósiles con un registro tan completo como el de los radiolarios, gracias a que el citoesqueleto de estos protozoarios está compuesto de sílice. Esta composición mineral de los citoesqueletos de radiolarios vivos produce esqueletos preservables. Extendiéndose durante todo el Fanerozoico, estos protozoarios ya eran aparentemente tan diversos y cosmopolitas en el Paleozoico, como lo son actualmente. El registro fósil tan detallado de los radiolarios los hace ser potencialmente uno de los grupos de microfósiles marinos más importantes (Haq y Boersma, 1978).

Los radiolarios son protozoarios exclusivamente marinos; su hábitat comprende un amplio rango de profundidades; su biogeografía es en general cosmopolita y está relacionada con la circulación de grandes masas oceánicas, desde zonas tropicales hasta zonas Antárticas. Se alimentan de diferentes tipos de organismos planctónicos, incluyendo microflagelados y otros protozoarios, diatomeas y aún formas tan grandes y activas como copépodos han sido reportadas dentro de la dieta de los radiolarios (Haq y Boersma, 1978). Estos organismos se distinguen de otros grupos de protozoarios unicelulares, por la presencia de una membrana que divide el contenido celular en una zona interna y una zona externa. Las características de esta membrana son relevantes en la clasificación de los subgrupos de radiolarios, siendo sin embargo, las diferencias fundamentales en la estructura geométrica y composición

del esqueleto las pautas decisivas para la clasificación taxonómica de los grupos de radiolarios.

Uno de los grupos más importantes de radiolarios son los policistinos, que comprende a los espumelarianos, organismos de simetría radial, y los nasselarianos que tienen forma de casco. Ambos grupos son comunes tanto en el plancton, como en sedimentos marinos. Los radiolarios Pseudocubus vema pertenecen al grupo de los espumelarios que son el grupo más antiguo de radiolarios, presentes en todas las colecciones fósiles desde el Paleozoico temprano, hasta nuestros días. Las formas esféricas tienen la historia más antigua y son probablemente el grupo más conservador.

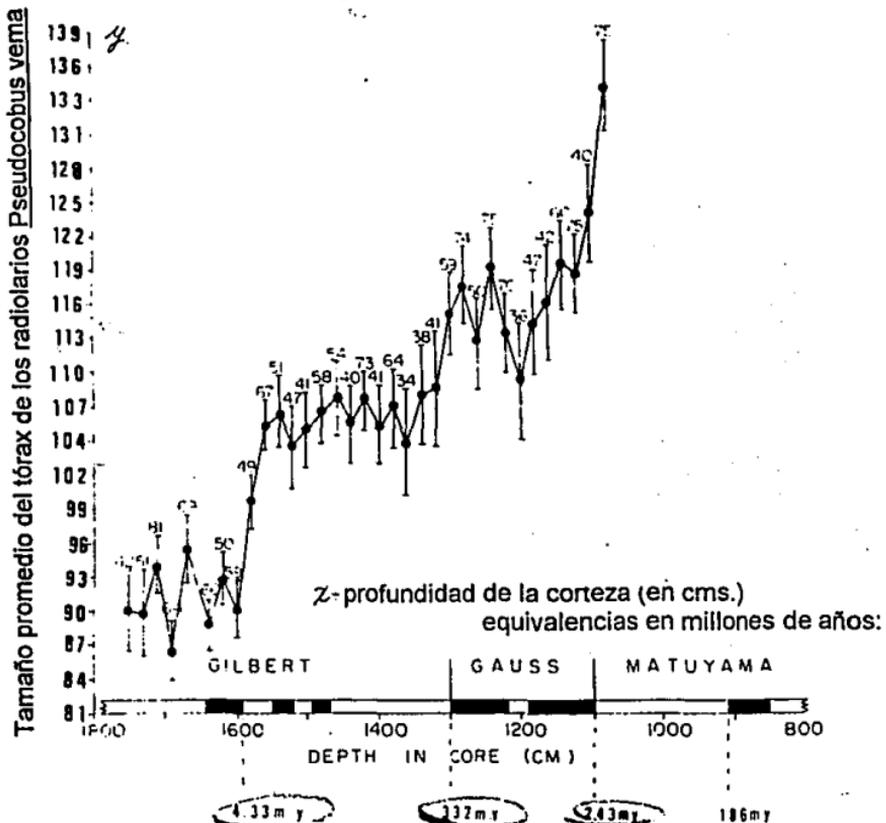
Cabe mencionar que los espumelarios, como grupo, han tenido una gran proliferación desde sus tempranos días en el Paleozoico, y especialmente en el Cenozoico, donde evolucionaron varios linajes. Se ha encontrado una importante correlación entre la abundancia relativa de las especies de radiolarios y las temperaturas promedio de las aguas marinas superficiales (Moore, 1973; citado en Haq y Boersma, 1978).

Debido a tan amplia distribución a nivel mundial y a su gran diversidad, los radiolarios son una importante herramienta paleoclimática, especialmente en sedimentos que carecen de sedimentos calcáreos. De hecho en los años setenta los radiolarios fueron sumamente utilizados en estudios paleoecológicos en los suelos oceánicos de la Antártida, que son suelos pobres en carbonatos (Haq y Boersman, 1978).

En 1975 la Doctora D.E. Kellogg publicó un artículo titulado "The role of phyletic change in the evolution of Pseudocubus vema (Radiolaria)" [Paleobiology. vol.1: 359-370]. En 1977 los paleontólogos Stephen Jay Gould y Neils Eldredge retomaron los datos publicados por la Doctora Kellogg, refutando la interpretación que la Dra. dio a sus datos, en una interesante discusión donde proponen un nuevo paradigma dentro de la Teoría de la Evolución, en su artículo "Punctuated Equilibria: the tempo and mode of evolution reconsidered" [Paleobiology. vol.3: 115-151.]

A continuación se presentan los datos publicados por la Doctora Kellogg, y se realiza el análisis estadístico de series de tiempo univariadas a este conjunto de datos.

8.1. Gráfica del cambio filogenético de los radiolarios Pseudocubus vema observado en su registro fósil.



Patrón del cambio filético de los radiolarios Pseudocubus vema

Las líneas verticales sobre los puntos son intervalos para las medias, al 95 % de confianza. Los números sobre las líneas verticales indican el tamaño de las muestras.

8.2. Tabla de datos.

La escala fija es una escala de profundidad en centímetros, que comprende un intervalo de 1000 centímetros (10 metros). El equivalente en tiempo de esta escala de profundidad, es una escala del tiempo con un intervalo de dos y medio millones de años aproximadamente; las fechas más antiguas son de hace 4.33 millones de años, y las fechas más recientes son de hace 2.43 millones de años. Estas fechas corresponden al la época del Plioceno tardío y el Pleistoceno temprano durante la Era cuaternaria, de hecho es el cambio del período terciario al cuaternario.

tiempo (cms de profundidad)	ancho promedio del tórax (micrones)
1740	90
1720	89
1700	94
1680	86
1660	95
1640	88
1620	92
1600	90
1580	99
1560	105
1540	106
1520	103
1500	104
1480	106
1460	107
1440	105.5
1420	107
1400	105
1380	106
1360	103
1340	107
1320	108
1300	115
1280	118
1260	112

1240	119
1220	113
1200	109
1180	114
1160	116
1140	120
1120	119
1100	124
1080	134

8.3. Análisis de los datos y construcción del modelo ARIMA mediante el uso del paquete PEST: Metodología.

1a etapa: Análisis descriptivo:

8.3.1 Gráfica de los datos originales.

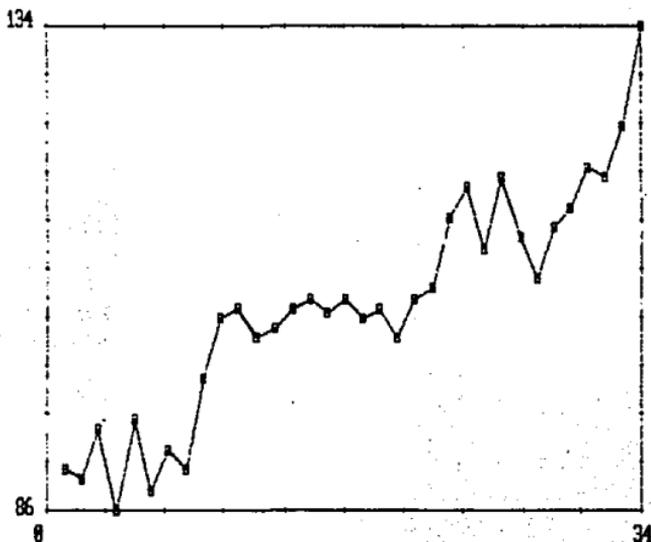
El total de observaciones es de 34, ($n = 34$).

$n1 = 90$

$n2 = 89$

$n3 = 94$

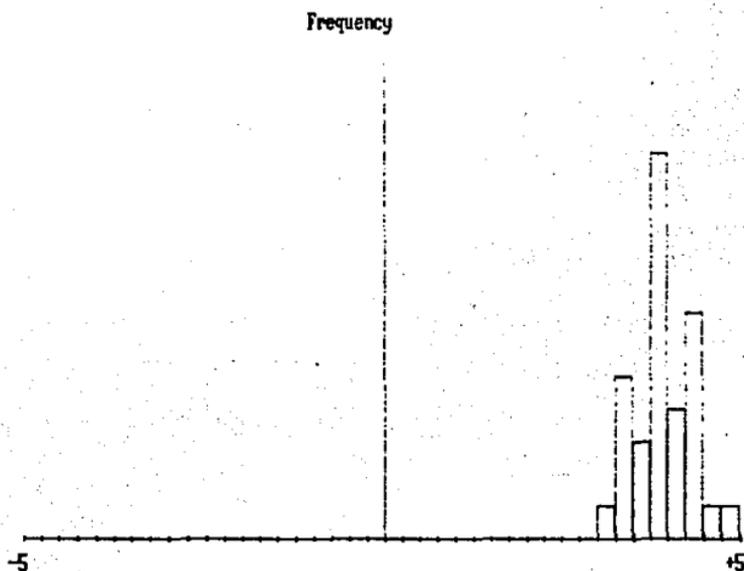
$n\ 34 = 134$



media = 106.132

desviación estándar = 11.1020

8.3.2. Histograma

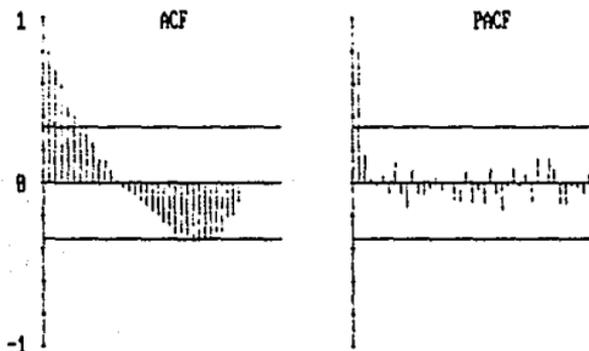


Horizontal Scale : 1 unit = $.2688E+02$
 Max. Frequency : 12 in [3.75, 4.00)

Mean of data plotted = 186.132488
 Std. deviation = 11.182818

8.3.3. Autocorrelograma y autocorrelograma parcial de los datos originales.

ACF decrece exponencialmente y se alternan signos...
 PACF alterna signos y no desaparece...



ACF:	.784	.688	.583	.468	.399	.311	.298	.232	.141	.134
	.878	.816	-.822	-.848	-.865	-.892	-.126	-.178	-.185	-.235
	-.258	-.385	-.296	-.387	-.378	-.368	-.344	-.341	-.292	-.292
	-.283	-.183	-.187	.888	.888	.888	.888	.888	.888	.888
PACF:	.784	.169	.821	-.885	.832	-.858	.112	-.857	-.139	.898
	-.875	-.866	-.818	.828	-.838	.881	-.898	-.188	.859	-.122
	-.832	-.113	.888	-.871	-.178	-.817	.887	-.816	.844	-.188
	.144	-.881	.142	.877	-.116	-.115	-.888	-.819	-.854	.845

8.4

2a etapa: Construcción del modelo:

8.4.1. Identificación del modelo:

a) Determinar una serie estacionaria en función de los datos originales:

Estabilización de la varianza, comparando diferencias. Determinar el valor de "d".

TOTAL OBSERVATIONS = ~~1000~~

CHECKING THE FIRST THREE AND LAST DATA POINTS:

-1.000000

5.000000

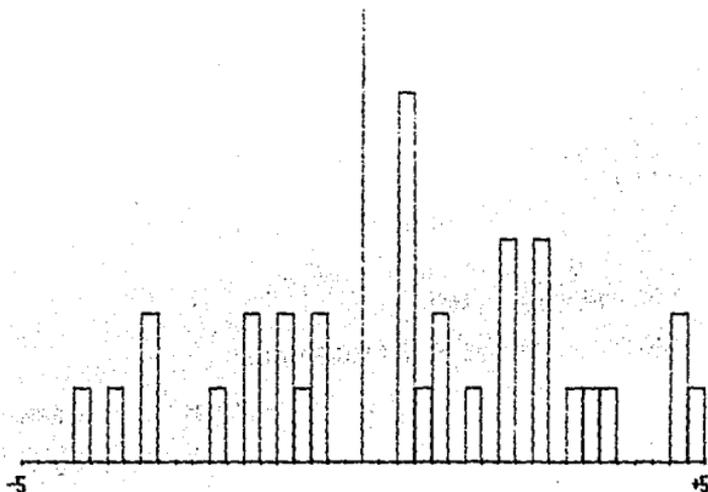
-8.000000

9.9999928

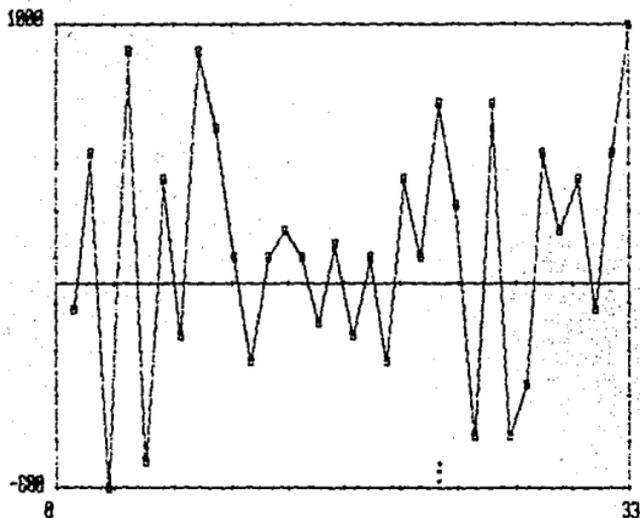
DATA FILE =

,DIFF'D

Gráfica del histograma y de los datos originales con la 1a. diferencia:
(se reduce la tendencia) Frequency



Horizontal Scale : 1 unit = .2000E+01
Max. Frequency : 5 in [.50, .75)



Mean of data plotted =
Std. deviation =

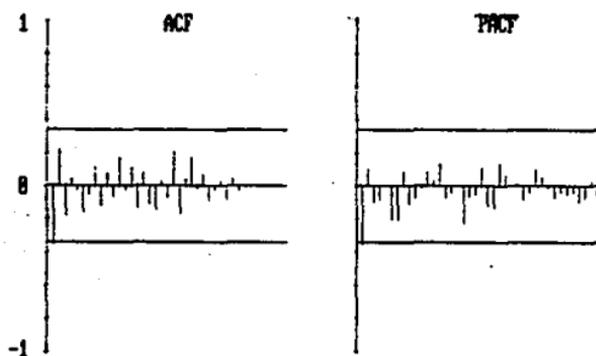
1.33333

4.657459

Autocorrelograma de los datos originales con la 1a. diferencia
(con la 1a. diferencia se reduce la desviación estándar original)

media = 1.333

desviación estándar = 4.657



ACF:	-.356	.211	-.188	.837	-.825	-.153	-.848	.184	-.117	.876
	-.855	.168	-.822	.184	-.133	.873	-.181	-.139	.825	-.876
	.194	-.164	.834	.162	-.888	.856	-.879	-.828	.824	-.887
	.832	-.828	.888	.888	.888	.888	.888	.888	.888	.888
PACF:	-.356	.896	-.898	-.877	-.884	-.199	-.195	.886	-.187	-.875
	-.811	.881	.825	.134	-.875	-.833	-.884	-.225	-.854	-.858
	.184	-.121	-.124	.128	.858	.888	.888	-.886	-.829	.898
	.846	-.815	-.874	-.835	-.851	-.838	-.899	-.886	.828	-.856

¿Qué pasa si diferencio la serie de datos originales dos veces?
 Por la segunda diferencia D''...

TOTAL OBSERVATIONS = 32

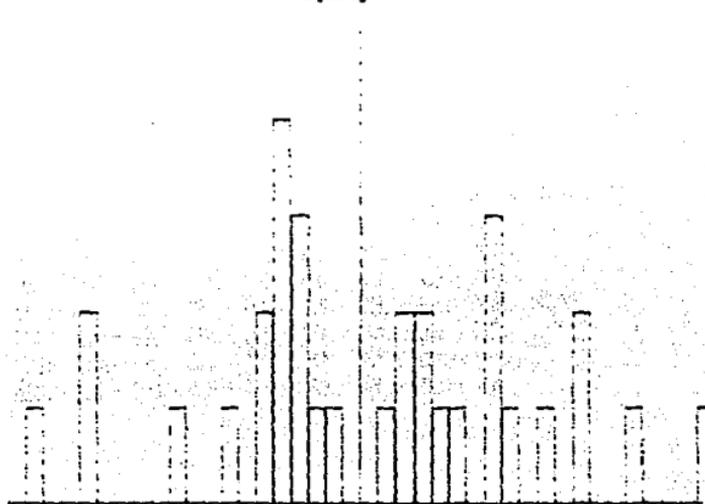
CHECKING THE FIRST THREE AND LAST DATA POINTS:

6.0000000 -13.0000000 17.0000000 4.9999858

DATA FILE = ,DIFF'D

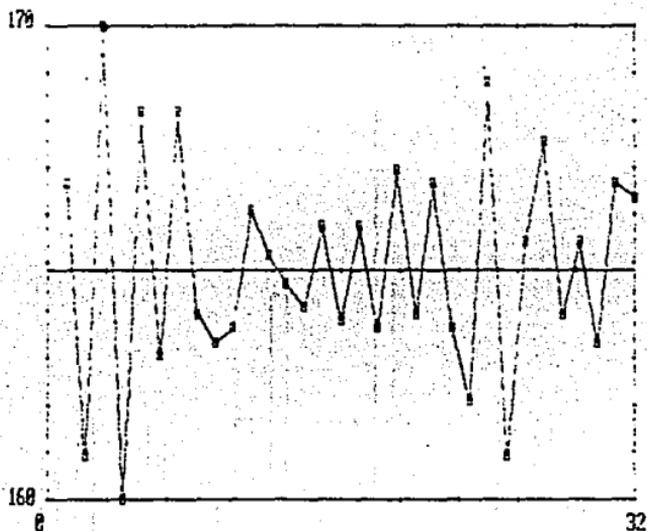
Datos originales, por la segunda diferencia...

Frequency



Horizontal Scale : 1 unit = .3400E+01
 Max. Frequency : 4 in [-1.25,-1.00)

Datos originales, por la segunda diferencia...



2a. diferencia de los datos originales de los radiolarios.

media = 0.343

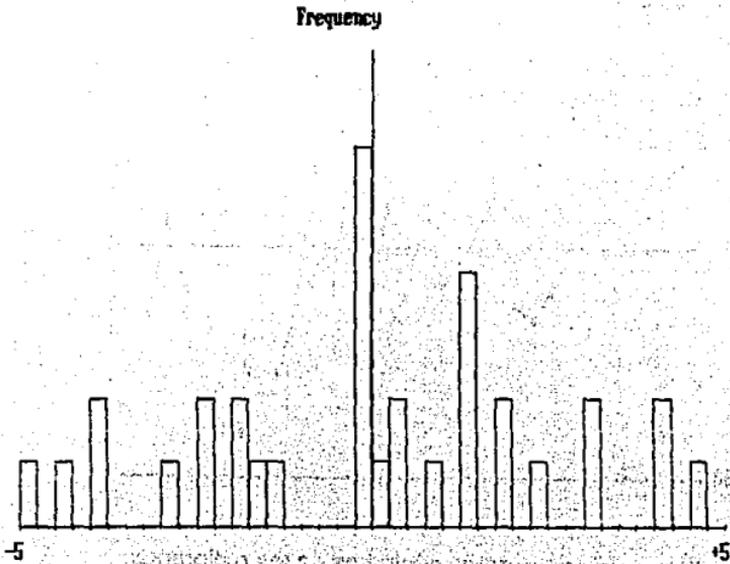
desviación estándar = 7.619

aumenta la desviación estándar,
no conviene sobrediferenciar

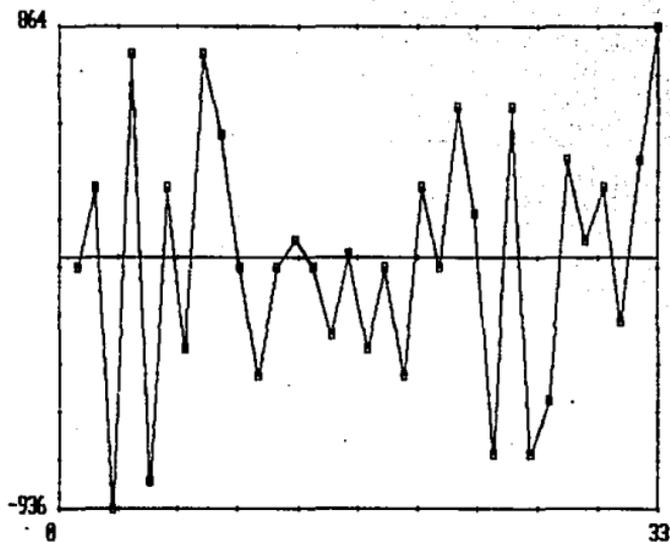
sigue 8.4.1.

$d = 1$

b) Sustraer la media...



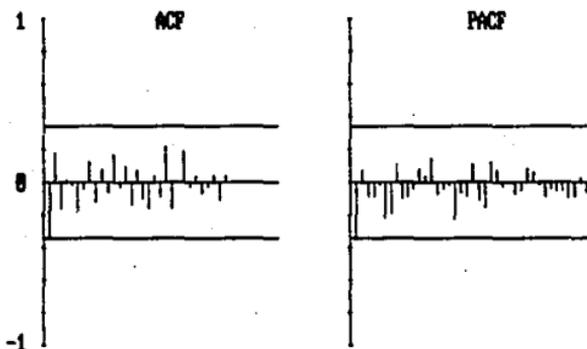
media = 0



Vertical scale: 1 unit = .188888E+01;
 Max. on vertical scale = .863636E+01; Min. = -.936364E+01

Mean of data plotted = .888888
 Std. deviation = 4.619251

Autocorrelograma de los datos con $d = 1$, media = 0



ACF:	-.338	.178	-.158	.011	-.013	-.176	-.034	.119	-.113	.065
	-.058	.165	-.019	.090	-.130	.065	-.090	-.157	.033	-.006
	.211	-.151	.005	.190	-.022	.037	-.071	-.022	.035	-.100
	.030	-.004	.000	.000	.000	.000	.000	.000	.000	.000
PACF:	-.338	.072	-.079	-.001	-.013	-.213	-.109	.103	-.094	-.079
	-.032	.000	.032	.142	-.073	-.037	-.016	-.224	-.055	-.000
	.110	-.109	-.152	.122	.076	-.029	-.003	-.069	-.046	.006
	.059	-.000	-.002	-.033	-.052	-.045	-.091	-.006	.022	-.054

- c) Obtener una representación ARMA para la serie estacionaria.
d) Fijar los valores de "p" y "q". ARIMA (1,1,1)

```
Input the order of Autoregression (Z7):  
(input -1 if model to be read from file):
```

1

```
Input the order of Moving Average (Z7):1
```

```
Input the starting values for the coefficients  
PHI( 1)= .1  
THETA( 1)= .1
```

8.4.2. Estimación de los parámetros del modelo ARIMA (1,1,1)

$$+(-.585)X(t-1) + (.168)Z(t-1)$$

AUTOREGRESSIVE PARAMETERS :

$$\text{PHI}(1) = \quad \quad \quad -.58498918$$

MOVING AVERAGE PARAMETERS :

$$\text{THETA}(1) = \quad \quad \quad .16761829$$

$$\text{WHITE NOISE VARIANCE} = \quad \quad \quad .185634E+82$$

$$\text{BIC STATISTIC} = \quad \quad \quad 193.792488$$

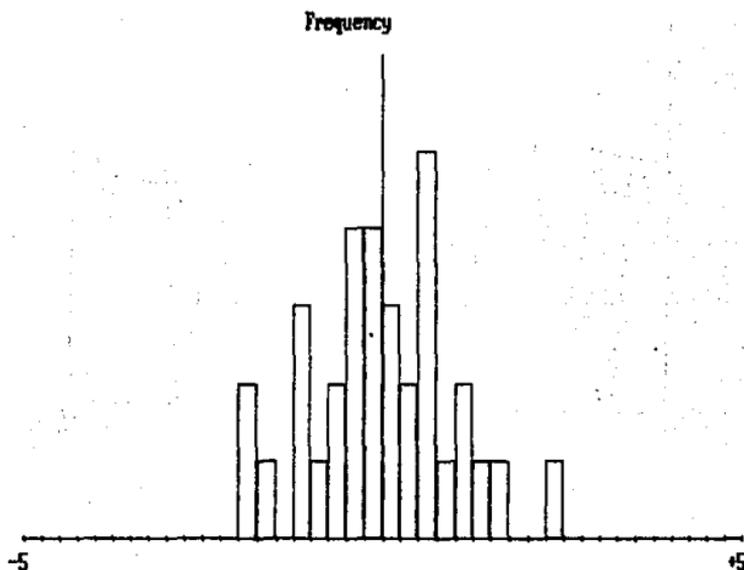
$$-2 \ln(\text{LIKELIHOOD}) = \quad \quad \quad 198.195388$$

$$\text{AIC STATISTIC} = \quad \quad \quad 197.822988$$

8.4.3. Verificación de supuestos.

a) Análisis de residuales:

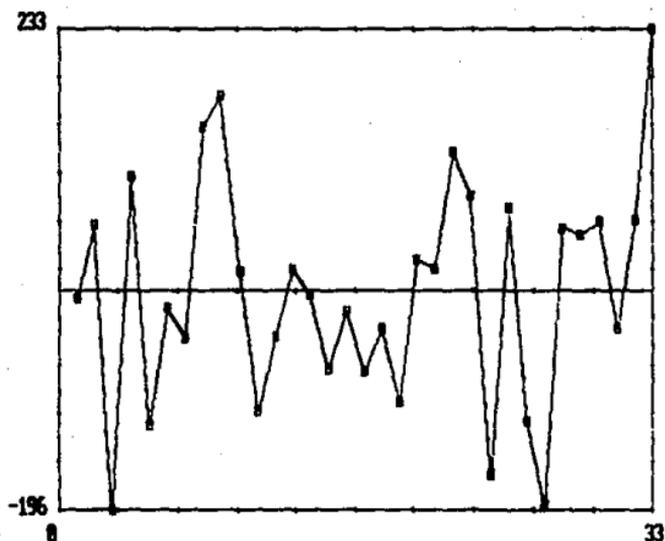
Histograma de residuales del modelo ARIMA (1,1,1)



Horizontal Scale : 1 unit = 1 std. deviation

Max. Frequency : 5 in I (.50, .75)

Gráfica de residuales del modelo ARIMA (1,1,1)



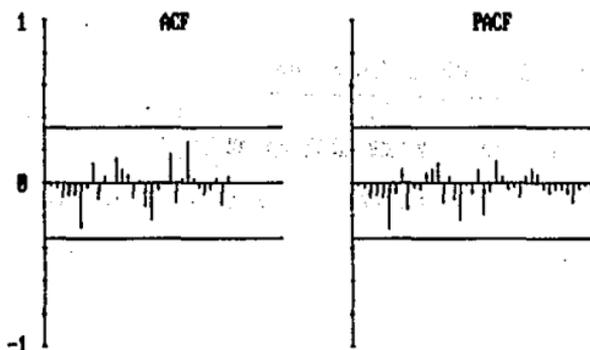
Vertical scale: 1 unit = .100000E+01;
 Max. on vertical scale = .232656E+01; Min. = -.196123E+01

Max. on vertical axis = 2.326565
 Min. on vertical axis = -1.961233

Mean of data plotted = -.016002
 Std. deviation = .999872

media = -0.016
 desviación estándar = 0.995

Autocorrelograma de los residuales del modelo ARIMA 1,1,1)



ACF: -.015 -.023 -.006 -.067 -.070 -.265 -.010 .120 -.090 .036
 -.002 .150 .004 .049 -.001 .007 -.140 -.222 -.040 -.005
 .100 -.112 .028 .242 .026 -.010 -.073 -.036 .020 -.126
 .041 -.004 .000 .000 .000 .000 .000 .000 .000 .000

PACF: -.015 -.023 -.007 -.070 -.070 -.205 -.062 .002 -.167 -.027
 -.041 .050 .001 .113 -.116 .035 -.009 -.224 -.014 -.060
 .070 -.106 -.050 .120 .035 -.036 -.020 -.003 .030 .005
 .050 -.037 -.064 -.039 -.036 -.059 -.113 -.036 -.014 -.051

b) Pruebas de Bondad del ajuste: Tablas de resultados.
ARIMA (1,1,1)

$Q = 7.2969, g.l. = 18$

WILCOXON TEST STATISTICS (see section 9.4)

PORT' TEAU with h= 28:	7.2969	chi sq	df= 18	
TURNING POINTS =	24.	ANORMAL(28.67	2.35**2)
DIFFERENCE-SIGN =	16.	ANORMAL(16.88	1.68**2)
RUNN TEST =	295.	ANORMAL(264.88	96.81**2)

8.4.4. Usos del modelo: pronósticos. ARIMA (1,1,1)

Tabla de pronósticos estimados:

THE ARMA(1, 1) MODEL IS $X(t) = Z(t)$
 $+(-.587)*X(t-1) + (.178)*Z(t-1)$

AUTOREGRESSIVE PARAMETERS :

PHI(1) = -.58686418

MOVING AVERAGE PARAMETERS :

THETA(1)= .16958258

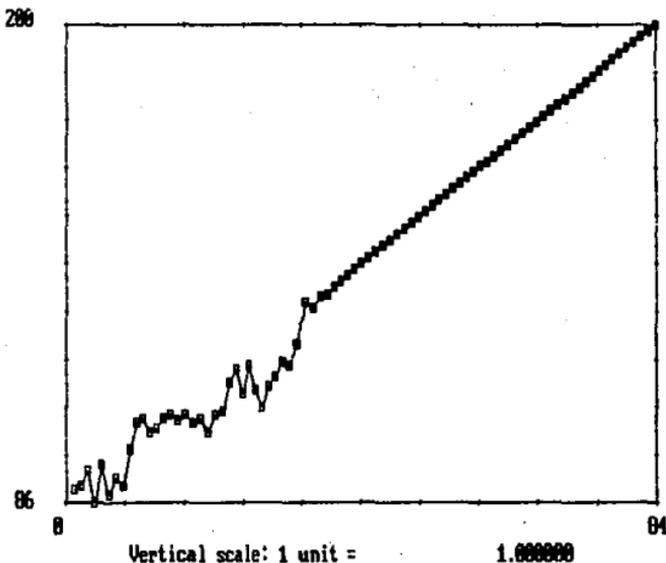
WHITE NOISE VARIANCE = .185632E+02

CONVERGENCE OCCURRED AT STEP 5 OF 33: VALUE OF R(N) = 1.0000000
 The UNV has been set to (RESIDUAL SS)/n = .185632E+02

Do you wish to change it (y/n)?n

#	XMAT	SQRT(MSE)	XMAT+MEAN (=	.13636E+01)
34	-.267837E+01	.438658E+01	-.131473E+01	
35	.135757E+01	.454787E+01	.272121E+01	
36	-.688183E+00	.468637E+01	.675533E+00	
37	.348775E+00	.462148E+01	.171241E+01	
38	-.176781E+00	.462536E+01	.118686E+01	
39	.896841E-01	.462635E+01	.145324E+01	
40	-.454171E-01	.462661E+01	.131822E+01	
41	.238283E-01	.462667E+01	.138666E+01	
42	-.116682E-01	.462669E+01	.135197E+01	
43	.591417E-02	.462678E+01	.136955E+01	
44	-.299768E-02	.462678E+01	.136864E+01	

pronósticos. ARIMA (1,1,1).



The datafile is : Total data points= 33

Difference lag
1 1

The subtracted mean is 1.3636

THE ARMA(1, 1) MODEL IS $X(t) = Z(t)$
 $+ (-.587)*X(t-1) + (.178)*Z(t-1)$

White noise variance = .185632E+02
 (Press any key to continue)

Otros modelos posibles...ARIMA (1,1,0):

Regreso a los datos originales para obtener una representación ARMA de los datos originales con una diferencia... (que ya será un modelo ARIMA dado que ya determiné que $d=1$):

Modelo ARIMA (1,1,0):

Order of autoregression to be fitted (p<Z?)? 1

Order of moving average to be fitted (q<Z?)? 0

AR COEFFICIENTS

-.3379645

RATIO OF COEFFICIENTS TO (1.96-STANDARD ERROR)

-1.8524698

(Press any key to continue)

UN VARIANCE ESTIMATE = (RESID SS)/N

.186417E+02

Estimación de los parámetros para el modelo ARIMA (1,1,0)

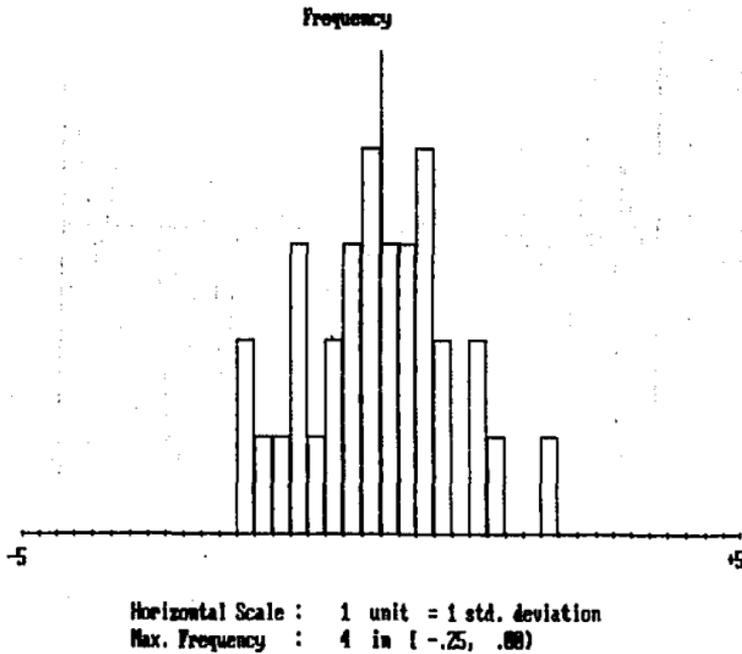
THE ARMA(1, 0) MODEL IS $X(t) = Z(t)$
 $+ (-.366)X(t-1)$

AUTOREGRESSIVE PARAMETERS :

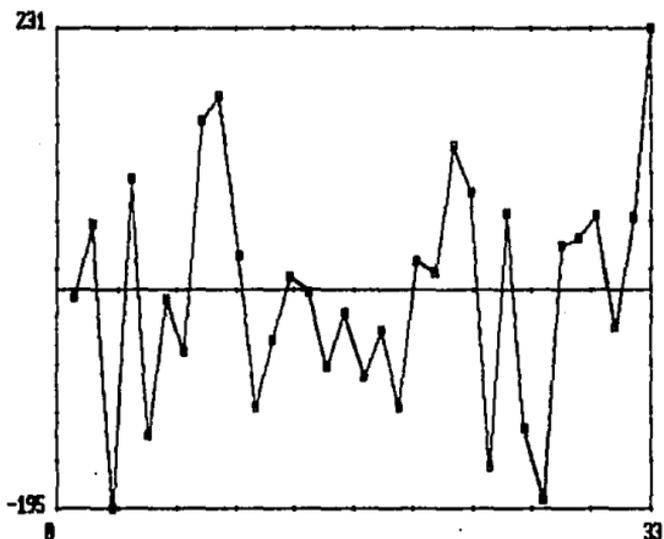
PHI(1) = -.36568048

WHITE NOISE VARIANCE =	.186139E+82
FPE STATISTIC =	.197773E+82
BIC STATISTIC =	192.698288
-2 ln(LIKELIHOOD) =	198.282588
AICC STATISTIC =	194.682588

Verificación de supuestos:
Análisis de residuales, modelo ARIMA (1,1,0)



Varianza de residuales del modelo ARIMA (1,1,0)
 media = 0, desvío estándar = 0.9997.

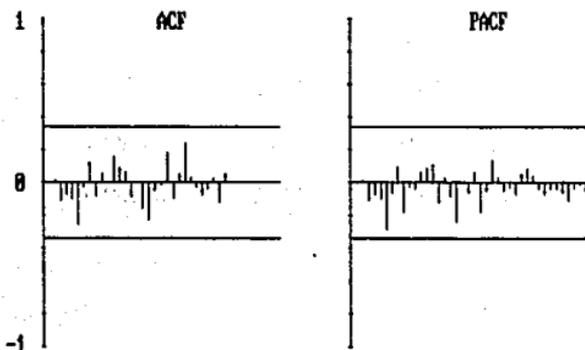


Vertical scale: 1 unit = .18888E+01;
 Max. on vertical scale = .238991E+01; Min. = -.194692E+01

Max. on vertical axis = 2.389918
 Min. on vertical axis = -1.946918

Mean of data plotted = -.822881
 Std. deviation = .999758

Autocorrelograma de residuales ARIMA (1,1,0)



ACF:	.001	.009	-.118	-.072	-.098	-.259	-.027	.128	-.007	.046
	.001	.152	.001	.057	-.079	-.003	-.148	-.227	-.042	-.000
	.173	-.095	.047	.235	.029	-.024	-.072	-.040	.025	-.123
	.041	-.004	.000	.000	.000	.000	.000	.000	.000	.000
PACF:	.001	.009	-.118	-.073	-.098	-.279	-.062	.093	-.100	-.026
	-.035	.053	.001	.109	-.117	.024	-.001	-.236	.001	-.056
	.058	-.100	-.044	.134	.028	-.045	-.025	-.075	.042	.006
	.040	-.030	-.066	-.033	-.039	-.061	-.111	-.035	-.015	-.049

Pruebas de la Bondad del Ajuste: modelo ARIMA 1,1,0

RANDOMNESS TEST STATISTICS (see section 9.4)

PORT' TEAU with h= 28: 7.7866 chi sq df= 19

TURNING POINTS =	22. ANORMAL(28.67	2.35**2)
DIFFERENCE-SIGN =	17. ANORMAL(16.88	1.68**2)
RAND TEST =	291. ANORMAL(264.88	96.81**2)

THE ARMA(1, 0) MODEL IS $X(t) = Z(t)$
 $+ (-.366)*X(t-1)$

AUTOREGRESSIVE PARAMETERS :

PHI(1) = -.36568848

WHITE NOISE VARIANCE =	.186139E+82		
FPE STATISTIC =	.197773E+82		
BIC STATISTIC =	192.698288		
-2 ln(LIKELIHOOD) =	198.282588	LAST=	198.282588
AICC STATISTIC =	194.682588	LAST=	194.682588

. Usos del modelo: Pronóstico con el modelo ARIMA (1,1,0).

THE ARMA (1, 0) MODEL IS $X(t) = Z(t)$
 $+ (-.366)X(t-1)$

AUTOREGRESSIVE PARAMETERS :

PHI (1) = -.36568048

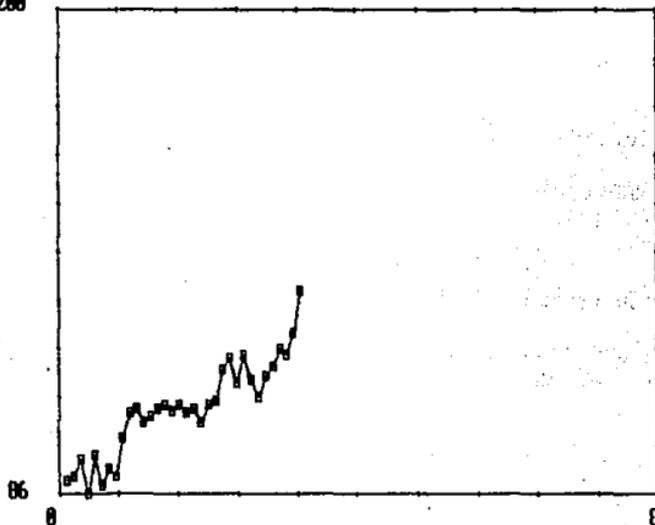
WHITE NOISE VARIANCE = .186139E+02

The WNV has been set to (RESIDUAL SS)/n = .186139E+02

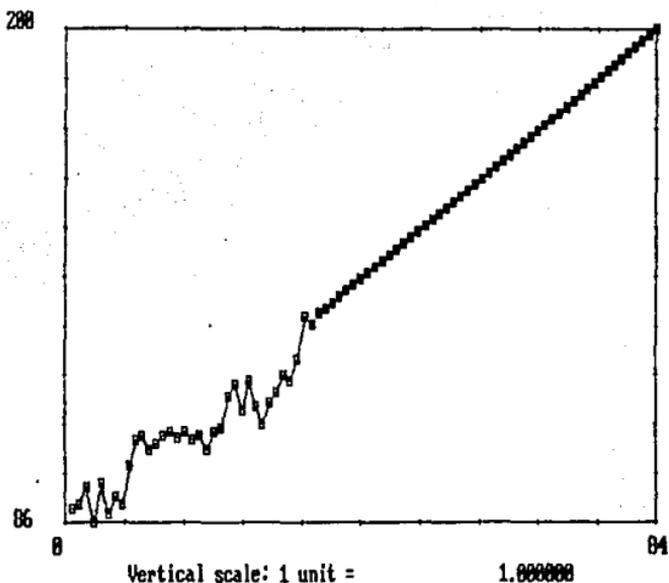
#	XHAT	SQRT(MSE)	XHAT+MEAN (= .13636E+01)
34	-.315753E+01	.431439E+01	-.179389E+01
35	.115442E+01	.459378E+01	.251885E+01
36	-.422865E+00	.462976E+01	.941571E+00
37	.154311E+00	.463456E+01	.151795E+01
38	-.564172E-01	.463528E+01	.138722E+01
39	.286266E-01	.463528E+01	.138426E+01
40	-.754126E-02	.463529E+01	.135618E+01
41	.275715E-02	.463538E+01	.136639E+01
42	-.188884E-02	.463538E+01	.136263E+01
43	.368547E-03	.463538E+01	.136488E+01
44	-.134744E-03	.463538E+01	.136358E+01

<Press any key to continue>

280



Pronóstico con el modelo ARIMA (1,1,0).



The datafile is : Total data points= 33

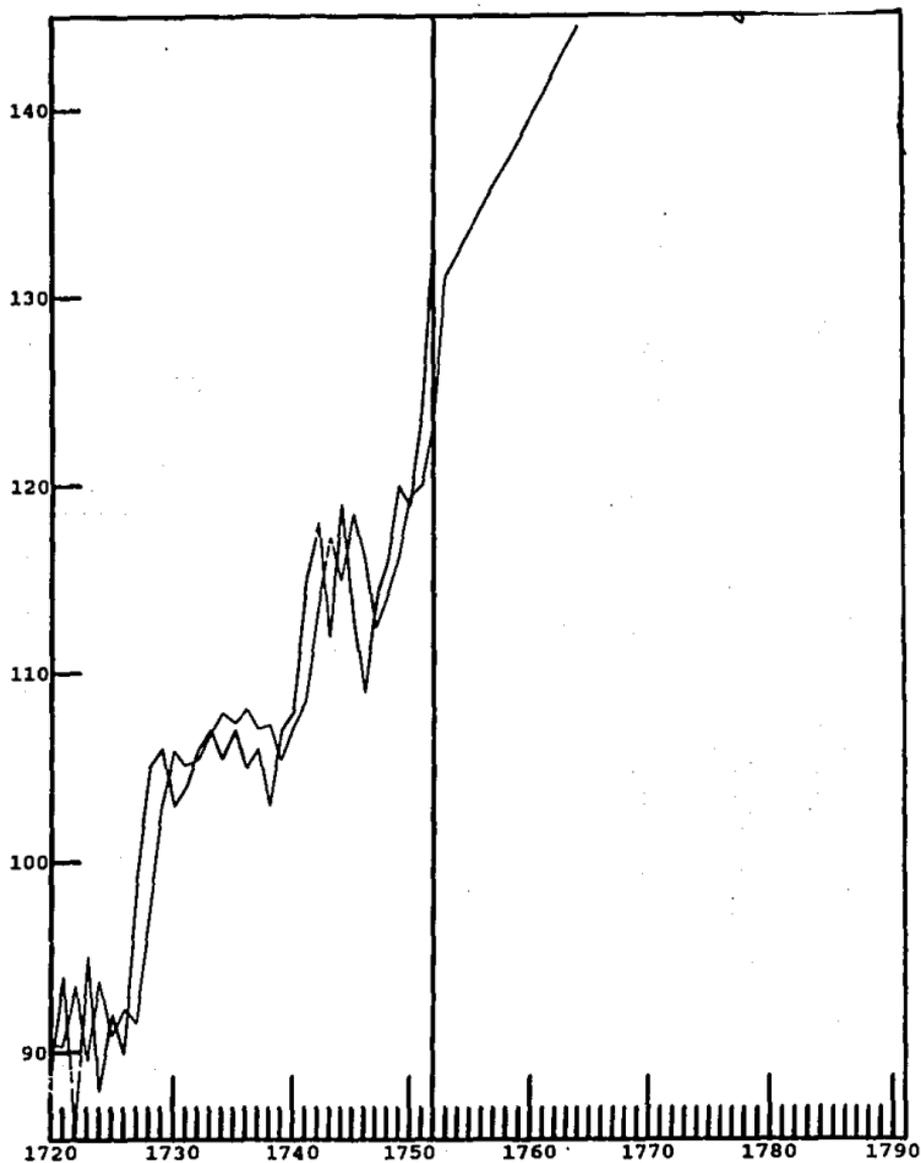
Difference	lag
1	1

The subtracted mean is 1.3636

THE ARMA(1, 0) MODEL IS $X(t) = Z(t)$
 $+ (- .366)X(t- 1)$

White noise variance = .186139E+02

Ajuste exponencial de HOLT, para el modelo ARIMA (1,1,0)



Resumen de Resultados de la Estimación de Modelos ARIMA
 8.5 Tabla comparativa de los modelos estimados para la serie de los radiolarios.

Serie: $Z_t =$ Radiolarios Pseudocobus vema.

Transformación: no

—————Análisis de Residuales—————

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Período y número de observaciones	Modelo	Parámetros estimados	Intervalos del 95% de Confianza	Correlación entre parámetros >0.5 o <0.5	$m(d)$ cociente	\bar{e}_n	Q, g.l.	χ^2	Residuales grandes
1740 - 1080									
N= 34	ARIMA (1,1,1)	$d = 1$ $\phi_1 = -0.505$ $\theta_1 = 0.1676$			$m = 0.016$	$= 0.1856$	$Q = 7.2969$		varianza no monótona
N= 34	ARIMA (1,1,0)	$d=1$ $\phi_1 = -0.3656$			$m = 0$	$= 0.1861$	$Q = 7.7066$		varianza no monótona

8.6 Análisis de intervenciones

Para realizar el análisis de intervenciones, es necesario cortar a la serie original en un período entre las dos intervenciones. Por tanto se cortó a la serie original en el período 1560 cms (105 micrones de ancho promedio de tórax), hasta 1120 cms (119 micrones de ancho promedio de tórax). N = 23 datos.

Serie patrón para los radiolarios: Serie de los datos originales de los radiolarios, sin las intervenciones: n=23 datos.

Histograma de los datos originales con que se alimentó al paquete:

TOTAL OBSERVATIONS = 23

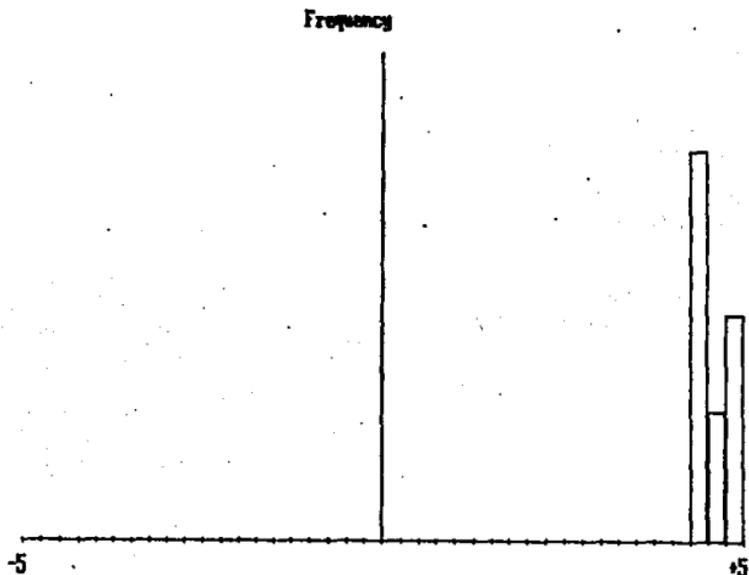
CHECKING THE FIRST THREE AND LAST DATA POINTS:

185.8000000

186.8000000

183.8000000

119.8000000



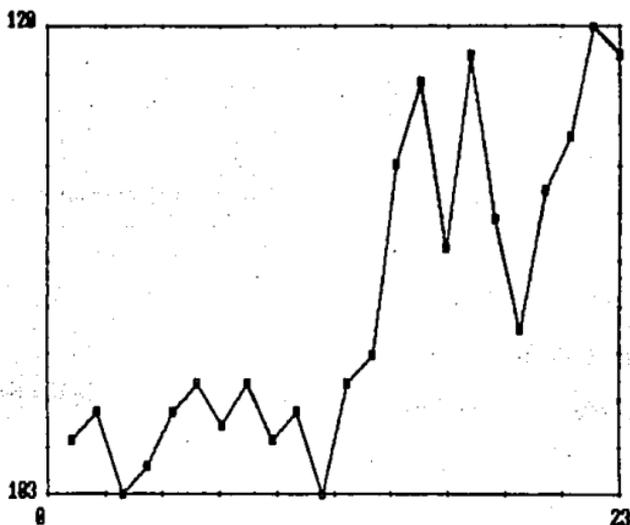
Horizontal Scale : 1 unit = .2400E+02

Max. Frequency : 12 in [4.25, 4.50)

Mean = 180000E+01: Std Dev = 55167E+01: C Skewness =

5.77

Grafica de los datos originales de la Serie patrón para los radiolarios

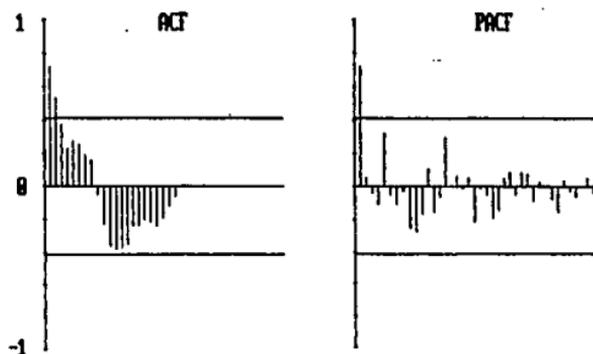


Vertical scale: 1 unit = .100000E+01;
 Max. on vertical scale = .129000E+03; Min. = .183000E+03
 Max. on vertical axis = 129.000000
 Min. on vertical axis = 183.000000

media = 109.891

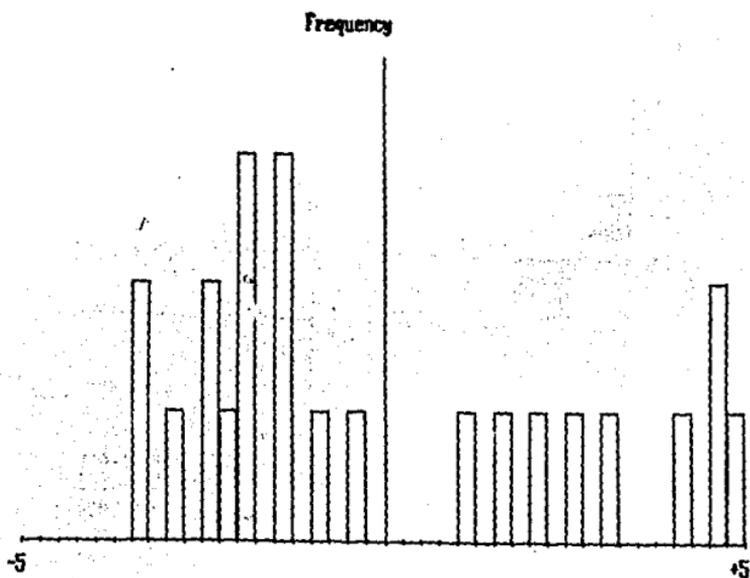
desviación estándar = 5.5166

Autocorrelograma de la serie patrón de los radiolarios.



ACT:	.715	.534	.399	.218	.274	.258	.188	.151	-.046	-.227
	-.353	-.376	-.364	-.348	-.238	-.248	-.282	-.212	-.239	-.189
	-.121	-.064	.000	.000	.000	.000	.000	.000	.000	.000
	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
PACT:	.715	.048	-.038	-.109	.318	-.043	-.105	-.029	-.251	-.266
	-.165	.183	-.149	-.068	.298	-.001	.061	-.012	.046	-.206
	-.013	-.049	-.188	-.142	.043	.003	-.046	.002	.071	-.077
	.018	.004	-.069	-.155	.033	-.029	-.058	-.004	.049	-.032

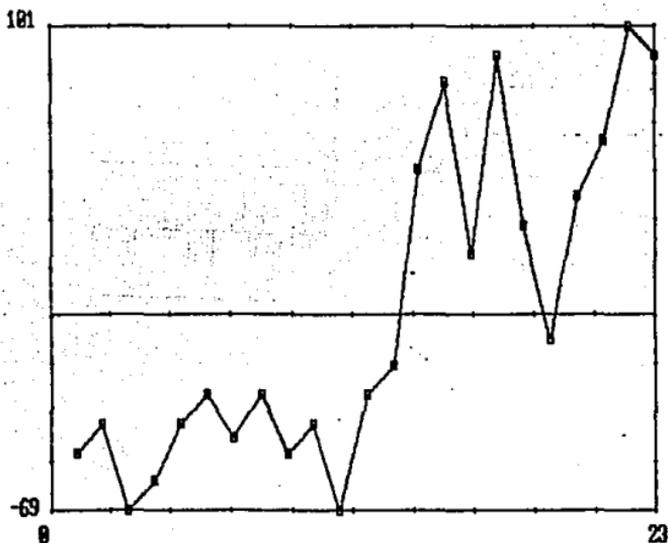
Sustraer la media : Histograma



Horizontal Scale : 1 unit = $.2822E+01$

Max. Frequency : 3 in $[-2.88, -1.75]$

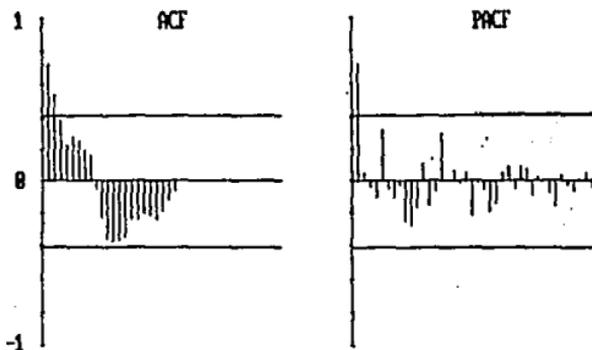
gráfica de los datos originales con media = 0



Vertical scale: 1 unit = .166388E+00;
 Max. on vertical scale = .181087E+02; Min. = -.689138E+01

Mean of data plotted = .000000
 Std. deviation = 5.516687

autocorrelograma de los datos originales con media = 0



ACF:	.715	.534	.388	.218	.274	.258	.188	.151	-.046	-.227
	-.353	-.376	-.364	-.348	-.238	-.248	-.282	-.212	-.239	-.189
	-.121	-.064	.000	.000	.000	.000	.000	.000	.000	.000
	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
PACF:	.715	.048	-.030	-.109	.318	-.043	-.105	-.020	-.251	-.266
	-.165	.103	-.149	-.068	.290	-.001	.061	-.012	.046	-.206
	-.013	-.049	-.188	-.142	.043	.003	-.046	.002	.071	-.077
	.018	.004	-.069	-.155	.033	-.029	-.058	-.004	.049	-.032

Obtener una representación ARMA para la serie estacionaria

```

Order of autoregression to be fitted (p(Z)? 1
Order of moving average to be fitted (q(Z)? 0
AR COEFFICIENTS
.7147183
RATIO OF COEFFICIENTS TO (1.96*STANDARD ERROR)
2.5883428
(Press any key to continue)

MN VARIANCE ESTIMATE = (RESID SS)/N
.125139E+02

(YULE-WALKER MN VARIANCE ESTIMATE = .148879E+02)

-2 ln(LIKELIHOOD) = 124.1835888
AICC STATISTIC = 128.7835888

```

Estimación de los parámetros del modelo ARIMA (1,0,0)

```

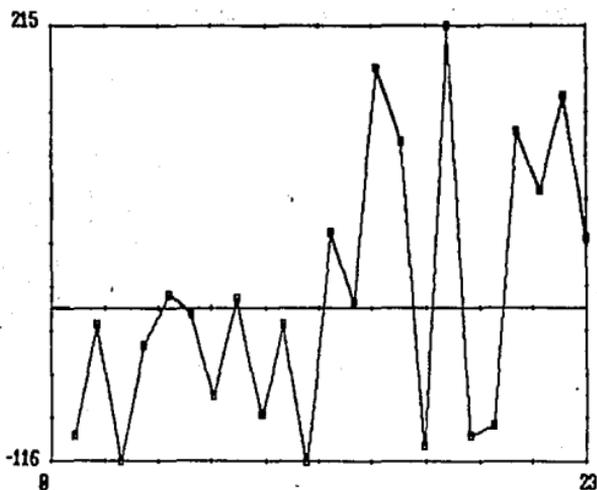
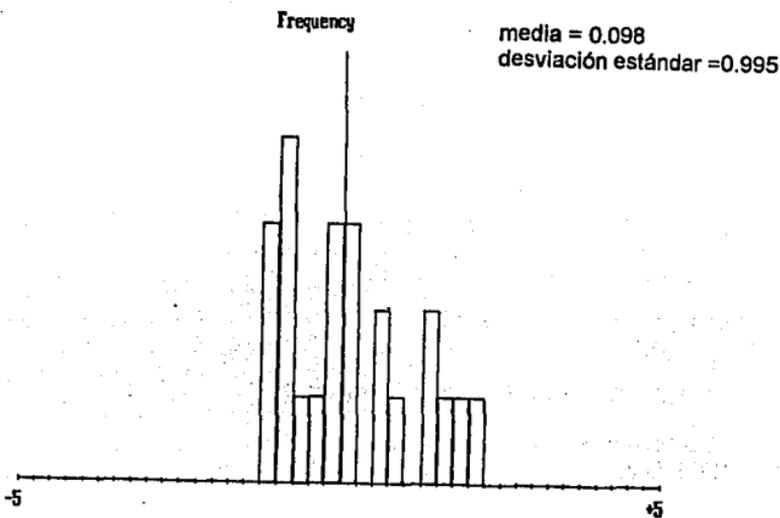
THE ARMA( 1, 0) MODEL IS  $X(t) = Z(t)$ 
+ ( .715)X(t- 1)

AUTOREGRESSIVE PARAMETERS :
PHI( 1) = .71471838

WHITE NOISE VARIANCE = .125139E+02
-2 ln(LIKELIHOOD) = 124.1835888
AICC STATISTIC = 128.7835888

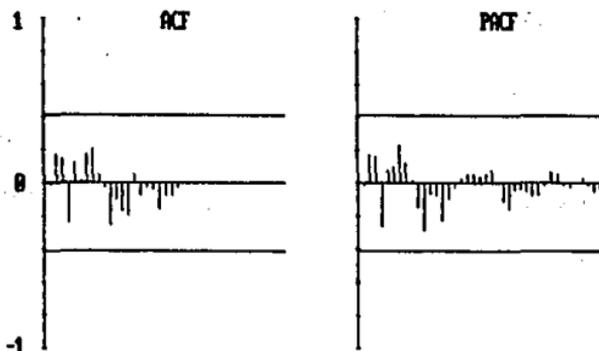
```

Análisis de residuales del modelo ARIMA (1,0,0)



Vertical scale: 1 unit = .100000E+01;
Max. on vertical scale = .214886E+01; Min. = -.116188E+01

autocorrelograma de residuales del modelo ARIMA (1,0,0)



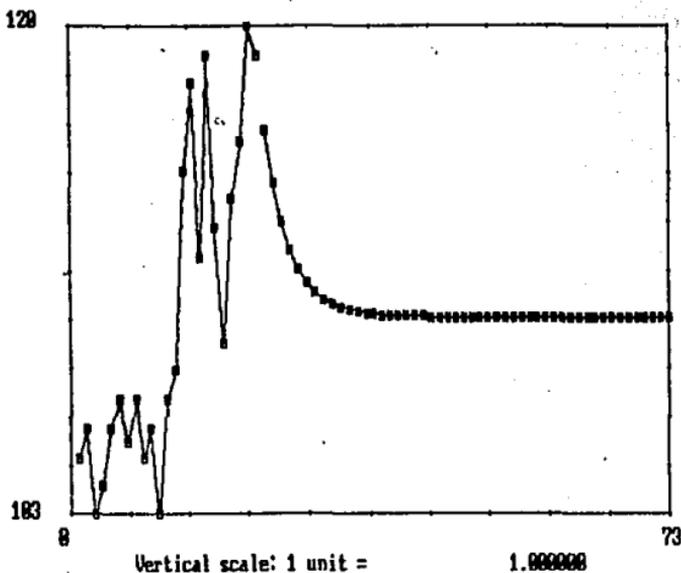
ACF:	-.013	.166	.147	-.230	.117	.015	.161	.204	.052	-.010
	-.243	-.093	-.162	-.192	.053	-.076	-.026	-.035	-.157	-.075
	-.075	-.020	.000	.000	.000	.000	.000	.000	.000	.000
	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
PACF:	-.013	.166	.155	-.264	.067	.096	.223	.114	.010	-.144
	-.201	-.056	-.069	-.219	-.090	-.020	.021	.050	.045	.031
	.052	.075	-.003	-.109	-.155	-.035	-.034	-.052	-.067	-.067
	-.007	.057	.042	-.015	-.020	.004	.020	-.017	-.040	-.040

Pruebas de la Bondad del ajuste

RANDOMNESS TEST STATISTICS (see section 9.4)

PORT'VENU with h= 28:	8.2462	chi sq	df= 19	
TURNING POINTS =	17.	ANORMAL(14.88	1.94**2)
DIFFERENCE-SIGN =	11.	ANORMAL(11.88	1.41**2)
RANK TEST =	166.	ANORMAL(126.58	56.88**2)

Pronóstico con el modelo ARIMA (1,0,0)



THE ARMA(1, 0) MODEL IS $X(t) = Z(t)$
 $+ (.715)X(t-1)$

AUTOREGRESSIVE PARAMETERS :

PHI(1) = .71471830

WHITE NOISE VARIANCE = .125139E+02

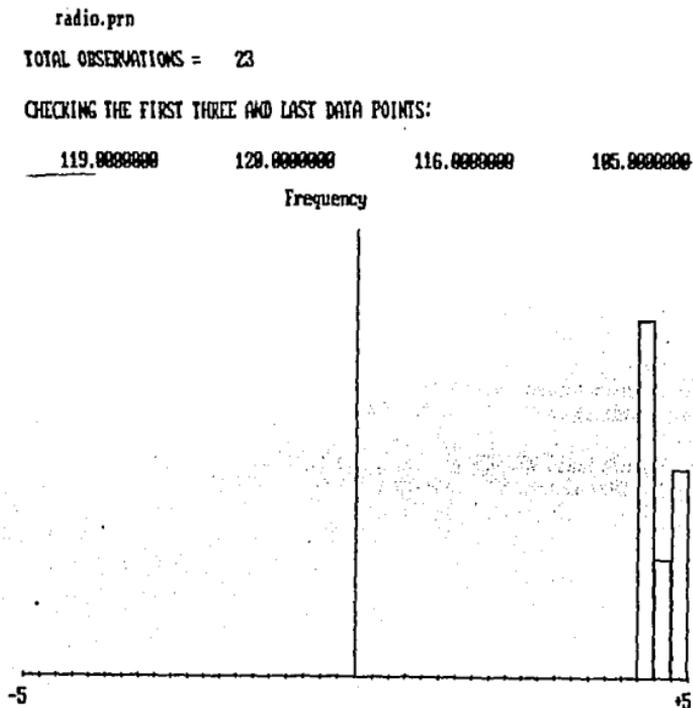
The RW has been set to (RESIDUAL SS)/n = .125139E+02

sigue 8.6 análisis de intervenciones

Para ver como sería la historia de la serie sin las intervenciones, se realiza el análisis de intervenciones, pero ahora "en espejo", es decir, se alimenta al paquete con la serie de datos originales recortada en el período entre las dos intervenciones mencionada, pero invertida. (ahora el valor inicial es 1120 cms (119 micrones de ancho promedio de tórax, y el valor final es 1560 cms (105 micrones de ancho promedio de tórax), N = 23 datos.

Serie patrón "en espejo" para los radiolarios (invertida): Serie de los datos originales de los radiolarios, invertidos y sin las intervenciones: n=23 datos.

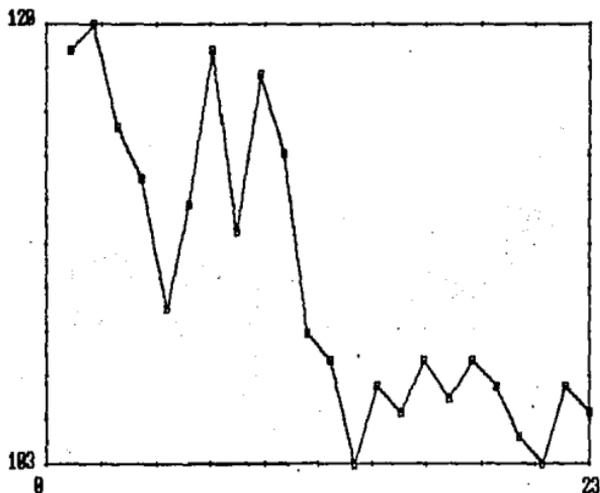
Histograma de los datos originales con que se alimentó al paquete:



Horizontal Scale : 1 unit = .2400E+02

Max. Frequency : 12 in (4.25, 4.50)

Grafica de los datos originales de la Serie patrón para los radiolarios, invertida.

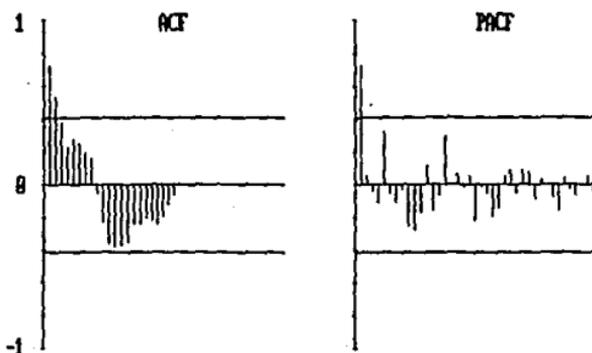


Vertical scale: 1 unit = .180000E+81;
 Max. on vertical scale = .129000E+83; Min. = .183000E+83

Max. on vertical axis = 128.000000
 Min. on vertical axis = 183.000000

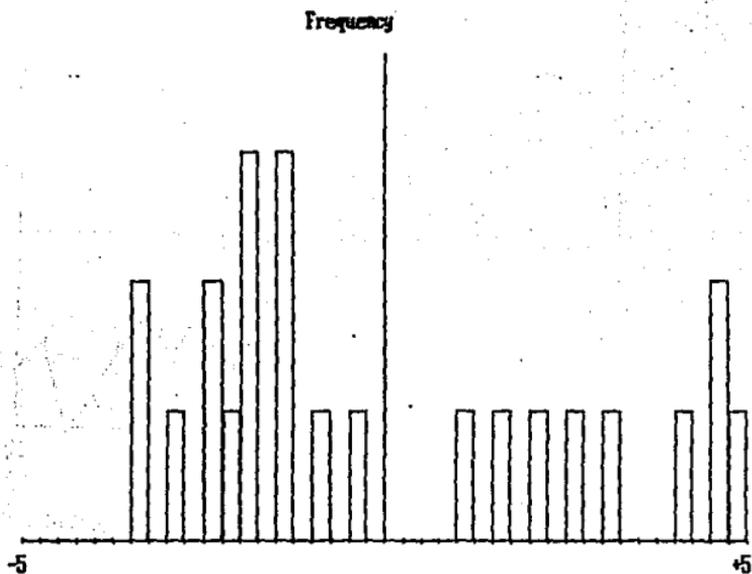
media = 109.891
 desviación estándar = 5.5166

Autocorrelograma de la serie patrón de los radiolarios, invertida.



ACF:	.715	.534	.398	.218	.274	.258	.188	.151	-.046	-.227
	-.353	-.376	-.364	-.348	-.238	-.248	-.282	-.212	-.239	-.189
	-.121	-.064	.000	.000	.000	.000	.000	.000	.000	.000
	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
PACF:	.715	.848	-.838	-.189	.318	-.843	-.185	-.828	-.251	-.256
	-.165	.183	-.149	-.068	.298	-.081	.061	-.012	.046	-.286
	-.813	-.849	-.188	-.142	.843	.883	-.846	.882	.871	-.877
	.818	.884	-.863	-.155	.833	-.829	-.858	-.884	.849	-.832

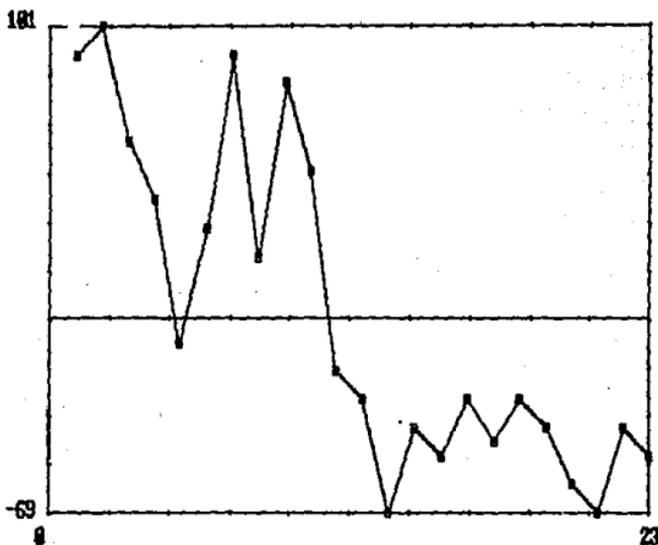
Sustraer la media : Histograma



Horizontal Scale : 1 unit = .2822E+01

Max. Frequency : 3 in [-1.58,-1.25]

gráfica de los datos de la serie patrón invertida con media = 0

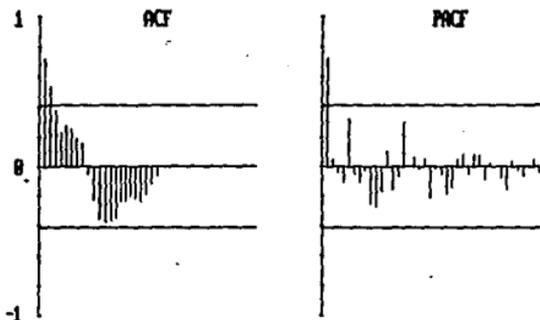


Vertical scale: 1 unit = .100000E+00;
 Max. on vertical scale = .101007E+02; Min. = -.689130E+01

Max. on vertical axis = 10.100700
 Min. on vertical axis = -6.891304

Mean of data plotted = .000000
 Std. deviation = 5.516687

autocorrelograma de los datos de la serie patrón invertida con media = 0



ACF:	.715	.534	.398	.218	.274	.258	.188	.151	-.046	-.227
	-.353	-.376	-.364	-.348	-.238	-.248	-.282	-.212	-.239	-.189
	-.121	-.064	.008	.008	.008	.008	.008	.008	.008	.008
	.008	.008	.008	.008	.008	.008	.008	.008	.008	.008
PACF:	.715	.048	-.038	-.109	.318	-.043	-.105	-.028	-.251	-.266
	-.165	.183	-.149	-.069	.298	-.081	.061	-.012	.046	-.206
	-.013	-.049	-.108	-.142	.043	.063	-.046	.062	.071	-.077
	.018	.004	-.069	-.155	.033	-.029	-.058	-.084	.049	-.032

Obtener una representación ARMA para la serie estacionaria

RATIO OF AR COEFFICIENTS TO (1.96*STANDARD ERROR)

1.0003438

RATIO OF MA COEFFICIENTS TO (1.96*STANDARD ERROR)

-.2786845

<Press any key to continue>

MIN VARIANCE ESTIMATE = (RESID SS)/N

.128819E+02

-2 ln(LIKELIHOOD) = 123.41550000

AICC STATISTIC = 130.67860000

Estimación de los parámetros del modelo ARIMA (1,0,1)

THE ARMA(1, 1) MODEL IS $X(t) = Z(t)$

+ (.819) * X(t- 1) + (-.174) * Z(t- 1)

AUTOREGRESSIVE PARAMETERS :

PHI(1) = .81919238

MOVING AVERAGE PARAMETERS :

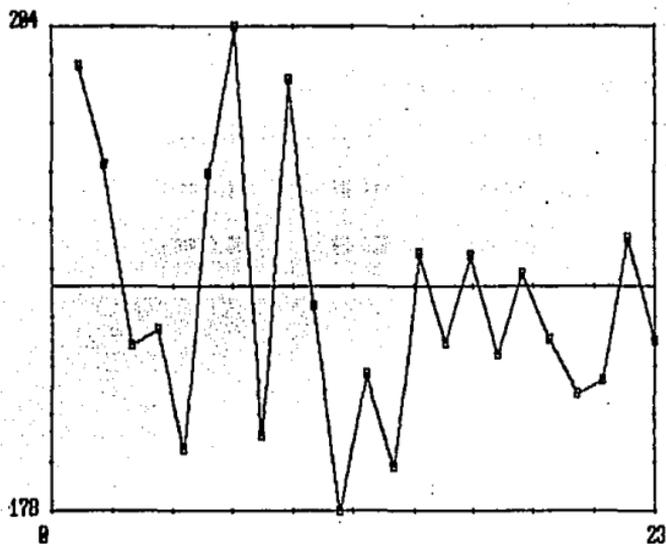
THETA(1) = -.17485368

WHITE NOISE VARIANCE = .128819E+02

CONVERGENCE OCCURRED AT STEP 5 OF 23; VALUE OF R(N) = 1.0000000

The MIN has been set to (RESIDUAL SS)/n = .128819E+02

Análisis de residuales del modelo ARIMA (1,0,1)



Vertical scale: 1 unit = .100000E+01;
 Max. on vertical scale = .284122E+01; Min. = -.177556E+01

media = 0.098
 desviación estándar = 0.995

Pruebas de la Bondad del ajuste

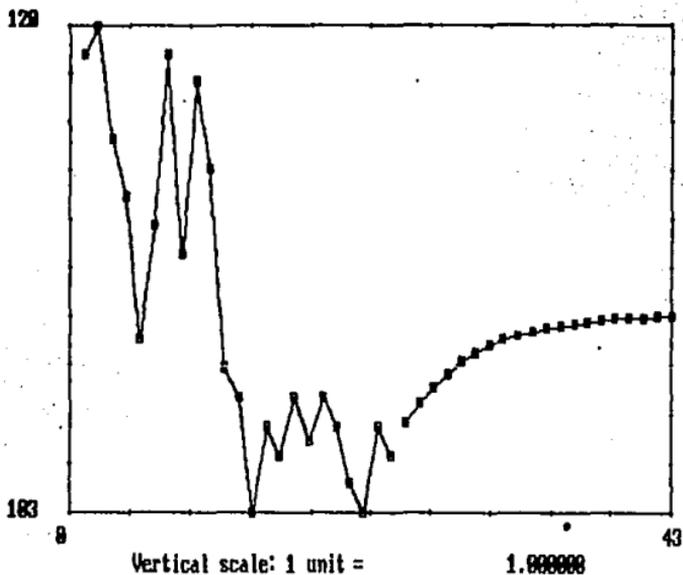
RANDOMNESS TEST STATISTICS (see section 9.4)

PORT'EAU with h= 28:	6.8943	chi sq	df= 18	
TURNING POINTS =	16.	ANORMAL(14.00	1.94**2)
DIFFERENCE-SIGN =	18.	ANORMAL(11.00	1.41**2)
ROWX TEST =	186.	ANORMAL(126.58	56.00**2)

Tabla de valores estimados para el pronóstico

#	XHAT	SQRT(MSE)	XHAT+MEAN (=	.18989E+83)
24	-.374994E+01	.347590E+01	.186141E+83	
25	-.387192E+01	.413648E+01	.186819E+83	
26	-.251649E+01	.452683E+01	.187375E+83	
27	-.286149E+01	.476965E+01	.187830E+83	
28	-.168876E+01	.492638E+01	.188283E+83	

Pronóstico con el modelo ARIMA (1,0,1)



THE ARMA (1, 1) MODEL IS $X(t) = Z(t)$
 $+ (.819)X(t-1) + (-.174)Z(t-1)$

AUTOREGRESSIVE PARAMETERS :

$\Phi(1) = .81919238$

MOVING AVERAGE PARAMETERS :

$\Theta(1) = -.17485368$

WHITE NOISE VARIANCE = .128819E+02

$-2 \ln(\text{LIKELIHOOD}) = 123.415588$

AICC STATISTIC = 138.678688

CONVERGENCE OCCURRED AT STEP 5 OF 23; VALUE OF R(N) = 1.888888

8.7. Discusión de resultados.

La primera etapa para la identificación y construcción de un modelo ARIMA, es observar las gráficas y los parámetros que describen a la serie. Si se observa la serie original es evidente en primer lugar, que se trata de una serie con dos marcadas intervenciones. Son dos eventos en los que se altera el comportamiento histórico de la serie, por lo que es posible que los residuales presenten valores más altos de los esperados en el intervalo 2σ , y la varianza de los residuales puede ser no monótona o tener crecimiento o decrecer. Es de esperarse también un valor alto en el estadístico de Q

Además se observa una tendencia en la serie después de la segunda intervención. Se identificó a $d=1$ como el valor óptimo para reducir la tendencia. Se aplicó una sola vez el operador diferencia a los datos originales dado que la segunda vez que se diferenciaron aumentó la desviación estándar por sobrediferenciación. La primera diferenciación redujo la desviación estándar original de 11.1020 a 4.65. La segunda vez que se diferenciaron los datos originales, el valor de la desviación estándar fue de 7.6193.

Una vez establecida la serie estacionaria sin tendencia, con el valor del operador "d" definido, (es decir, se "estaciona" a la serie a un nivel constante para poder buscar una representación ARMA), se pasó a la siguiente etapa en la construcción del modelo, la etapa de determinación de los valores de "p" y "q", mediante el análisis del autocorrelograma y de la función del autocorrelograma parcial. El autocorrelograma de los datos originales decrece sinusoidal y exponencialmente, con cambio de signo en la curva sinusoidal. Como el autocorrelograma no desaparece, es decir, no tiene un corte, sugiere un componente AR. En cuanto al autocorrelograma parcial tiene un elevado valor inicial, y no decrece ni desaparece, aparentemente tiene una cola. Esto sugiere un modelo ARMA. Por tanto los modelos iniciales que se probaron fueron ARMA (1,0); y ARMA (1,1). No se probó un modelo ARMA (0,1) porque el autocorrelograma inicial no presenta corte, y el FACP presenta un valor significativo en el primer parcial, lo que sugiere que el valor de p es 1, es decir, es un AR (1). Así se definió que se trata de un modelo

(1,1,q). El autocorrelograma y el autocorrelograma parcial de los datos con la primera diferencia también sugieren modelos (1,1,1) y (1,1,0). Así, los valores que se probaron para "q" fueron 0 y 1.

Para definir el valor de "q", se pasó a la siguiente etapa de la construcción del modelo ARIMA, la estimación de los parámetros. Primero se probó el modelo ARIMA (1,1,1) asignándole valores a ϕ_1 y a θ_1 de .1. Después se le dio selección automática y el resultado de esta estimación automática fue de $\phi_1 = -.5049$, y de $\theta_1 = .1676$.

Los dos modelos dieron pronósticos similares, y sus residuales y altos valores del estadístico de Q indican problemas en los dos ajustes realizados. Los modelos contra los que compitió dieron también pronósticos interesantes, pero no fué posible realizar un ajuste satisfactorio. Finalmente se trató de obtener un mejor pronóstico con el modelo ARIMA (1,1,0), realizando un ajuste exponencial "Holt" con que se logró el pronóstico con tendencia que se presenta al final del análisis,

El hecho de no poder ajustar un modelo a este conjunto de datos, evidencia claramente la presencia de intervenciones. Esta situación era de esperarse dada la presencia de las dos intervenciones que se mencionan al inicio del análisis. En estos casos, el análisis de intervenciones indicado es para intervenciones de tipo 3, que alteran el comportamiento histórico de la serie, rompiendo no sólo con la estructura estocástica de la serie, sino también con la estructura determinista de la serie, modificando el valor de los parámetros que tenía la serie antes de la intervención.

Pese a que esta serie es un caso similar al caso de las garzas analizadas en el capítulo 7, porque se trata también de una serie de 34 datos, en esta serie hay solamente dos intervenciones, por lo que se realizó el análisis de intervenciones del segmento de la serie comprendida entre las dos intervenciones.

El análisis de intervenciones se realiza comparando a la serie original con una serie patrón, que se obtiene cortando a la serie original en el período entre las intervenciones, para investigar cómo hubiera sido la historia de la serie original sin las intervenciones y

tener un patrón de referencia. Por tanto, se realizó el análisis de la serie "patrón" con los 23 datos comprendidos entre las dos intervenciones, y se buscó pronosticar los datos con el modelo obtenido. También se realizó el análisis de la "serie patrón invertida", (imagen en espejo) para tratar de proyectar la imagen invertida de la serie en el período considerado y tener más elementos para realizar el análisis de las intervenciones.

Los logros obtenidos del análisis de intervenciones son modestos, porque los pronósticos obtenidos con los modelos ajustados no son buenos. Esta situación era previsible desde el análisis de la varianza de residuales porque en ambos casos es no monótona, aunque el estadístico de Q tiene valores aceptables en ambas series (para la serie patrón: ARIMA (1,0,0): $Q = 8.24$; y para la serie en espejo: ARIMA (1,0,1): $Q = 6.89$ para 18 g.l.). El paquete PEST no realiza este tipo de análisis y para realizar el análisis de intervenciones es necesario tener series mayores, de más de 50 datos como ya ha sido mencionado, para poder partir a la serie en períodos anteriores y posteriores a la intervención que resulten significativos. Sin embargo, cabe subrayar, que la presencia de las intervenciones queda clara mediante el análisis realizado

8.8. Conclusiones.

Se trata de una serie no estacionaria, con dos marcadas intervenciones. Los valores iniciales de la serie tienen un comportamiento estacionario, que es interrumpido bruscamente por la primera intervención en el tiempo 4.33 millones de años. Después de la primera intervención se observa un período estacionario de aproximadamente un millón de años, seguido por una etapa con tendencia ascendente. Finalmente en los últimos valores de la serie se puede señalar a una tercera intervención que rompe con el valor de pendiente de la tendencia que se venía presentando y altera la serie aumentando de manera drástica el tamaño de las fluctuaciones del valor de la variable (tamaño de tórax), con respecto a las fluctuaciones anteriores. En esta serie no se observan ciclos, no es una serie estacional.

El modelo que mejor se ajustó a los datos es el ARIMA (1,1,0) este modelo se eligió por parsimonia dado que el modelo ARIMA (1,1,1) tiene mayor número de parámetros. Entre más parámetros tiene el modelo más aumenta la desviación promedio (estándar). Entre más datos tiene una serie es mejor porque el desvío estándar disminuye y la estimación es más precisa. Por tanto la serie de los radiolarios paleozoicos también padece de pocos datos ($n = 34$). Un número mayor a 50 es recomendable para obtener mejor precisión en la estimación como ya se ha mencionado.

Ahora bien, la serie analizada en particular, no es una prueba contundente a favor del cambio por puntuado. Si bien hay un evidente momento puntuado, es decir, un cambio de nivel drástico en la serie provocado por la primera intervención, -hecho que apoya el cambio por equilibrios puntuados-; en la segunda intervención, que determina otro cambio muy evidente de nivel, se observa una etapa inicial con una tendencia "suave" a incrementar del volumen del tórax. La pendiente "suave" que inicia en la profundidad 1280 (118) con ascensos y descensos al incremento del volumen del tórax, culmina con un cambio brusco de nivel que aumenta en promedio 5 micrones al tamaño promedio del tórax.

La explicación de la etiología de las intervenciones que se presentan en la serie de tiempo sale de los límites del análisis estadístico. Su explicación se ubica dentro de un contexto paleontológico. Hemos dicho que las intervenciones son eventos exógenos que alteran la estructura histórica de la serie. Si ubicamos a las intervenciones de nuestra serie de radiolarios en la escala de tiempo geológico, se observa que, la primera intervención se presentan entre los 4.33 y 4.00 millones de años, mientras que la segunda intervención inicia a los 3.12 millones de años aproximadamente. Estos periodos en la escala geológica se ubican en la Era Cenozoica, en el período Terciario, en la época del Plioceno-Pleistoceno.

Ahora bien, toda la historia Cenozoica está marcada por evidentes cambios ciclos del nivel del mar, que ocurrieron a escala global, en todo el planeta. Esta evidencia se basa en que los márgenes continentales tienen cambios simultáneos y de magnitudes relativas similares. Los cambios cíclicos del nivel del mar que se

observan en el registro geológico, son debidos a procesos de gran magnitud: los glaciares y los eventos geotectónicos (Vail, Mitchum y Thompson, 1977).

Los eventos geotectónicos y los glaciares, implicaron importantes cambios climáticos que afectaron no sólo la temperatura ambiental global de planeta, sino que coinciden con los cambios globales cíclicos del nivel del mar (Vail, Mitchum y Thompson, 1977).

Los cambios en el nivel del mar, probablemente provocaron corrientes marinas superficiales que no sólo aumentaron la temperatura del agua, sino que también acarrearón mas nutrientes al medio en que vivían los radiolarios. Todos estos cambios favorecieron el incremento de tamaño de los radiolarios.

En suma, el análisis estadístico de esta serie de tiempo no es una evidencia contundente de que el cambio evolutivo total ocurrió por saltos. Si bien la primera intervención tienen un claro efecto puntuado en el tamaño promedio del tórax de los radiolarios, antes de que ocurriera la tercera intervención, ya se observaba una tendencia a incrementar el volúmen del tórax. Es necesario buscar series mayores para poder aplicarles un análisis de intervenciones.

Capítulo 9:

Análisis de los resultados generales del trabajo:

Alcances y limitaciones de los modelos ARIMA obtenidos mediante la aplicación del análisis estadístico de las series de tiempo univariadas a los procesos ecológicos y evolutivos considerados.

1. El análisis estadístico de series de tiempo univariadas Si es una herramienta aplicable a datos numéricos del registro fósil. Pese a que la mayoría de los datos disponibles del registro fósil son en general datos categóricos (Montellano M.1994 *Comunicación oral), es importante buscar registros fósiles con datos numéricos y estudiar a estos datos del registro fósil como series de tiempo, en donde el tiempo equivale a la dimensión espacio (profundidad), en la escala del registro fósil. Es importante repetir que los intervalos de "tiempo" considerados deben ser intervalos fijos y de igual magnitud.

2. No se trata de reducir las cuestiones evolutivas a meras "elucubraciones estadísticas". Si los procesos evolutivos son series de tiempo, el tratamiento estadístico debe ser precisamente el análisis de series de tiempo, cuando se tengan registros numéricos y con bases de datos de más de 50 datos. Las series de tiempo se deben estudiar justo como series de tiempo.

3. El análisis estadístico de series de tiempo univariadas es una herramienta muy importante pues es un argumento estadístico y matemático que puede ayudar a explicar "saltos" o "eslabones perdidos" que sin un análisis formal se convierten en una discusión sin fundamentos sólidos. Si la variable dependiente en estudio se relaciona con un eje de abscisas con intervalos de tiempo (o las unidades de profundidad consideradas) fijas, mediante el análisis estadístico de series de tiempo se tiene un fundamento real que indica la presencia y el tipo de cambios en la estructura histórica de la serie, es decir, marca claramente la presencia de intervenciones como factores exógenos, que modifican la estructura histórica de la serie en estudio; y define el tipo de cambios en dicha estructura histórica de la serie, porque marca una diferencia entre tendencias e intervenciones.

4. La hipótesis de trabajo de la presente investigación es en primer lugar investigar si el análisis estadístico de series de tiempo univariadas es aplicable a series de tiempo de procesos estocásticos biológicos. Dado que si lo es, tomando en cuenta las consideraciones anteriormente mencionadas, las siguientes preguntas son: ¿cuáles son los alcances y limitaciones de esta técnica estadística aplicada a series de tiempo de procesos evolutivos estocásticos? ¿Qué tanta información nos pueden dar? ¿Para qué?

Series de tiempo es una herramienta formal para explicar el cambio evolutivo porque permite clasificar el modo como ocurrió el proceso evolutivo en términos de pendientes, de velocidades del cambio en relación al tiempo. Así un cambio brusco, puntuado, implica un cambio instantáneo (en términos geológicos), mientras que las pendientes "suaves" indican tendencias, cambios graduales. Por tanto, es importante considerar esta técnica en estudios de procesos evolutivos biológicos, dado que a la fecha de realización del presente trabajo, no hay trabajos publicados sobre estudios de evolución con análisis de series de tiempo. Por tanto, propongo considerar a la evolución como procesos estocásticos deterministas y estudiarla con la herramienta estadística apropiada, para identificar en el tiempo si el cambio fué instantáneo, en términos geológicos.

En los objetivos específicos de esta tesis, se postula dar un argumento matemático estadístico mediante el modelo ARIMA obtenido, a favor del cambio gradual o del cambio puntuado. Para lograr esto, es necesario realizar un análisis de intervenciones. El análisis de la serie de los radiolarios apunta un cambio puntual en la primera etapa de la serie y un cambio "gradual-puntuado" en la etapa final de la serie. Sin embargo, la conclusión que indica este resultado es que es necesario analizar más series de tiempo de procesos evolutivos, (de más de 50 datos, de preferencia), bajo esta óptica, y realizar un análisis de intervenciones para evaluar el intervalo de confianza en el que cae el pronóstico resultante del análisis de intervenciones, y así poder evaluar el tipo de intervención que tuvo lugar en la serie. De esta manera, es posible confirmar cómo ocurrió el cambio, si por saltos bruscos, por tendencias o si en realidad los procesos evolutivos han ocurrido mediante la combinación de los dos

tipos de cambios que tradicionalmente se han considerado como antagonicos en la visión tradicional.

5. En cuanto a los alcances y limitaciones de esta técnica aplicada al estudio de procesos estocásticos evolutivos, en los objetivos específicos también se plantearon las preguntas: ¿Es posible distinguir los componentes deterministas de los componentes estocásticos mediante este análisis estadístico? ¿Es posible distinguir deriva génica de selección natural?

El análisis estadístico de series de tiempo, como se ha visto en el desarrollo de esta tesis, es un procedimiento tanto estadístico como matemático en el que los aspectos deterministas de las series de tiempo se definen mediante los parámetros de los polinomios de las ecuaciones en diferencia presentados en el capítulo 2. El aspecto estocástico de una serie, queda bien especificado al igualar la media a cero. El aspecto estocástico de la serie se observa en la amplitud y frecuencia de las fluctuaciones del proceso alrededor de la media igualada a cero. Por tanto el análisis si separa los componentes deterministas de los componentes estocásticos de la serie.

Ahora bien, dado que si es posible distinguir componentes deterministas de componentes estocásticos en un conjunto de datos de series de tiempo, ¿es posible explicar en términos biológicos qué significa cada parte de la serie?. Su aspecto estructural, la varianza, ¿a qué se debe el que la serie tenga determinada estructura? ¿qué factores pueden alterar esta estructura? ¿cuál es el significado en términos evolutivos de la media y de la varianza? ¿qué factores aumentan su amplitud y cuáles la restringen? ¿porqué?.

Estas últimas preguntas me llevaron a estudiar los conceptos fundamentales de la Teoría de la Evolución, bajo una perspectiva especial: tratar de entender a el proceso evolutivo como cambios resultantes de dos tipos de fuerzas antagonicas, componentes conservadores, de memoria o de mantenimiento de las valiosas características que han resultado exitosas para los organismos, y por otra parte los componentes de cambio, dirigidos o determinados por la selección natural, de acuerdo con la hipótesis de la reina roja que todo el tiempo corre para mantenerse siempre en el mismo lugar, es decir que las poblaciones están en cambio constante sin llegar nunca

a un punto óptimo, porque los propios organismos modifican constantemente la topografía adaptativa, y los cambios no deterministas, es decir los azarosos o estocásticos debidos a las mutaciones, deriva génica, migraciones, y demás factores impredecibles que son en realidad eventos exógenos o intervenciones en la historia de un proceso estocástico evolutivo.

Y... ¿Es posible distinguir deriva génica de selección natural?. Esta es una pregunta que rebasa los alcances del análisis de las series consideradas en que este trabajo. Este trabajo es un ejemplo del tipo de estudios que se pueden realizar con esta técnica estadística. Con este enfoque es necesario realizar más estudios de procesos evolutivos estocásticos a diferentes niveles, alteraciones en el número de comunidades, cambios en la diversidad, a nivel de genética de poblaciones como variaciones en la adecuación de los organismos, etc. Este enfoque estadístico, puede ser un auxiliar en la interpretación de los datos de procesos estocásticos evolutivos abordados a diferentes niveles por las disciplinas científicas correspondientes. Para diferenciar si los cambios son debidos a deriva génica o a selección natural, es necesario diseñar experimentos específicos con esta finalidad, o analizar otros registros fósiles, bajo esta perspectiva para entender la secuencia de la serie. Esta herramienta estadística define claramente cuándo un evento exógeno, es decir una intervención, "rompe" o altera la secuencia histórica de una serie, sin embargo, el análisis estadístico de series de tiempo "per se" no explica a qué se debe la intervención. La explicación causal de la intervención corresponde a otras disciplinas científicas. Dependiendo del tipo de proceso estocástico evolutivo en estudio, la explicación será del orden de eventos catastróficos como glaciares, meteoritos, tectónica de placas; o bien la explicación será la aparición de mutaciones si se trata de series de tiempo de frecuencias génicas y adecuación.

6. Por lo anterior, en este trabajo, se propone considerar el análisis estadístico de series de tiempo en los protocolos y diseños de investigaciones sobre la evolución de las especies, y asimismo, que la evolución puede ser entendida mediante un modelo estocástico-determinista que considere las fuerzas antagónicas mencionadas.

Es necesario realizar más estudios de procesos estocásticos evolutivos, a diferentes niveles de complejidad, analizando los datos bajo la perspectiva de este nuevo prisma estadístico matemático, y digo .prisma pues es un nuevo instrumento, una nueva herramienta estadística matemática que puede ayudarnos a entender los datos del registro fósil bajo un nuevo enfoque.

Capítulo 10

Conclusiones generales.-

1. El objetivo general de esta tesis si se cumple, porque el análisis estadístico de series de tiempo univariadas Si es una herramienta aplicable a datos numéricos del registro fósil.
2. Es importante buscar registros fósiles con datos numéricos y estudiar a estos datos del registro fósil como series de tiempo, en donde el tiempo equivale a la dimensión espacio (profundidad), en la escala del registro fósil. Es importante repetir que los intervalos de "tiempo" considerados deben ser intervalos fijos y de igual magnitud.
3. El análisis de intervenciones realizado con los modelos ARIMA obtenidos mediante el análisis estadístico de series de tiempo aplicado a conjuntos de datos numéricos de series de tiempo evolutivas, es un procedimiento que puede ayudar a formalizar en términos estadístico matemáticos el proceso de cambio evolutivo, porque permite discernir las intervenciones puntuales (saltos, cambios bruscos de un nivel a otro), de los cambios ocurridos por tendencias.
4. Se sugiere considerar el análisis estadístico de series de tiempo en los protocolos de investigaciones sobre procesos estocásticos evolutivos, dado que la herramienta estadística para analizar datos en el tiempo, es justamente el análisis estadístico de series de tiempo

Biol. Mariana Núñez Zúñiga
Otoño de 1994.

BIBLIOGRAFIA.

- Ayala F. J. 1980. Evolución Molecular. Ediciones Omega, S.A. España.
- Begon M., Harper J.L. y Townsend C. 1986. Ecology: Individuals, Populations and Communities. Editado por la Facultad de Ciencias de la UNAM.
- Berg H.C. 1993. Random walks in Biology. Princeton Univ. Press. USA.
- Boyce W.E. y DiPrima R.C. 1977. Elementary Differential Ecuations. Wiley and sons Ed. N.Y. 3rd, ed.
- Coleman. R. 1976. Procesos estocásticos. Limusa.México.
- Darwin Ch. 1859. On the Origin of Species. (A facsimil of the first edition) Harvard Univ. Press, 1979.
- Degroot M.H. 1988. Probabilidad y Estadística. Ed. Addison-Wesley Iberoamericana.USA.
- Diggle P.J. 1990. Time Series: A Biostatistical Introduction. Oxford Science Publications. Great Britain.
- Dobzhansky T., Ayala F.J., Stebbins G.L. y Valentine J.W. 1980. Evolución. Ediciones Omega S.A. España.
- Eguiarte L. E. 1986. "Una guía para principiantes a la Genética de Poblaciones". Revista "Ciencias", número especial: Polémicas contemporáneas en Evolución. pp 30-38.
- Eguiarte L. E. y Souza V. 1993. "Evolución experimental en bacterias: diez mil generaciones del experimento de Irvine California". En: Tópicos de Biología Evolutiva. (Núñez F.J. y Cordero C. eds.) Centro de Ecología de la UNAM. México.
- Eldredge N. 1985. Time Frames. Princeton University Press.
- Elton, C. Nicholson M.J. 1942. "The ten year cycle in numbers of the lynx in Canada". J.Anim. Ecol. 11:2115-244.
- Feldman, M.W. (ed). 1989. Mathematical Evolutionary Theory. Princeton University Press.
- García R. C. 1991. Aplicaciones de las Matemáticas a la Biología. CINVESTAV:VII Coloquio del Departamento de Matemáticas (29 de Julio al 16 de Agosto de 1991).M,México.
- Gould S.J. y Eldredge N. 1977. "Punctuated Equilibria: the tempo and mode of evolution reconsidered". Paleobiology. 1973: 115-151.
- Gould S.J. y Eldredge N. 1986. Punctuated Equilibria at the third Stage Syst Zool. vol. 35. pp 143-148.

- Guerrero V.M. 1991a. Análisis estadístico de Series de Tiempo económicas. Universidad Autónoma Metropolitana. México.
- Guerrero V.M. 1989. "Optimal conditional ARIMA Forecasts". Journal of Forecasting, Vol 8, 215-229 (1989).
- Guerrero V.M. 1991b. "ARIMA Forecasts with restrictions derived from a structural change". International Journal of Forecasting". 7(1991)339-347. North Holland.
- Guerrero V.M. 1991c. "Restricted ARIMA Forecasts with account for parameter changes". División Académica de Matemáticas Documentos de Trabajo. ITAM.
- Guerrero V.M. 1993. "Combining historical and preliminar information to obtain timely Time Series data". División Académica de Matemáticas ITAM. (en prensa).
- Haq B. U. y Boersma A. 1978. Introduction to marine micropaleontology. Ed. Elsevier. España.
- Kuhn, T.S. 1970. La estructura de las revoluciones científicas. Fondo de Cultura Económica. México.
- Lazcano A.A. 1994. "¿Y si Darwin no hubiera muerto?". La Jornada semanal. No.256. 8 de Mayo de 1994.
- Makridakis, S. Wheelwright, S. y Mc Gee V. 1992. Forecasting Methods and Applications. Ed. John Wiley and Son. New York.
- Martínez M. J. 1986. "¿Qué es la evolución molecular?". Revista "Ciencias" Número especial: Polémicas, Contemporáneas en Evolución. pp. 24-29.
- Núñez F.J y Cordero C. 1993. Tópicos de Biología Evolutiva, Diversidad y adaptación. Editado por el Centro de Ecología de la UNAM. México.
- Núñez Z.M. 1984. Método de aislamiento de leucocitos polimorfonucleares en el estudio de receptores beta ardenérgicos. Tesis de licenciatura en Biología. Facultad de Ciencias. UNAM.
- Olea F. A. 1986. "La Teoría del Equilibrio Puntuado. Una alternativa al Neodarwinismo". Revista "Ciencias" Número especial: Polémicas Contemporáneas en Evolución. pp. 46-59.
- Peitgen H.O., H. Jürgens y D.Saupe. 1992. Fractals for the Classroom Part One: Introduction to Fractals and Chaos. Springer Verlag. New York. Cap.7: The Chaos Game: How Randomness Creates Deterministic Shapes.

- Pielou E.C. 1977. Mathematical Ecology, 2nd. ed. Wiley and Sons, New York.
- Piñero D.D. 1987. De las bacterias al hombre: La evolución. Fondo de Cultura Económica, colección la ciencia desde México.
- Ridley, M. 1993. Evolution. Blackwell Sci. Pub. Boston.
- Rosenblueth, A. 1971. El Método científico. CONACYT. México.
- Rostand J. 1945. Introducción a la Historia de la Biología. Ediciones Península. Barcelona. 1979.
- Sarukhán J. 1988. Las musas de Darwin. Fondo de Cultura Económica. México.
- Schaffer W. 1984. "Stretching and folding in Linx fur returns: evidence for a strange attractor in nature?" *The American Naturalist* 124, 6 : 798-820.
- Smith M. 1982. Evolution now. A Century after Darwin. Ed. Nature. London.
- Soberón M.J. 1986. "Hacia una visión jerarquizada de los fenómenos evolutivos". Revista "Ciencias", número especial: Polémicas contemporáneas en Evolución. pp 60-63.
- Soberón M.J. 1987. Ecología de Poblaciones. Fondo de Cultura Económica, colección la ciencia desde México.
- Shumway R.H. 1988. Applied Statistical Time Series Analysis. Prentice Hall, New Jersey.
- Vail R. P., Mitchum JR y Thompson S. 1977. Seismic Stratigraphy and Global Changes of Sea Level, Part 4: Global Cycles of Relative Changes of Sea Level. Seismic Stratigraphy Applications to Hydrocarbon Exploration (C.E. Paytoneel.) Amer. Assoc. Petrol. Geol. Bull., v.60 p.83 -97.
- Vrba E.S. 1984. What is Species Selection?. *Syst. Zool.* 33(3): 318 -328.
- Wei S.W. 1990. Time Series Analysis, Univariate and Multivariate Methods. Addison Wesley Publ.Co.Canada.

CURSOS

Apuntes del curso de "Aplicaciones del análisis de series de tiempo", impartido por el Dr Jorge Martínez Collantes en el Modulo VII del Diplomado en Estadística Aplicada del ITAM., Viernes 23 de Julio al Sábado 21 de Agosto de 1993.

Anexo a: Comportamientos típicos de la Función de autocorrelación FAC y de la Función de autocorrelación parcial FACP.

Guerrero V.M. 1991a.
Análisis estadístico de series
de tiempo económicas.
capitulo 4, pp 115-123.

En conclusión, para llevar a cabo la etapa de identificación, es conveniente recordar las características de la FAC y de la FACP; que hemos resumido en el cuadro 3 de la pp 73 que se reproduce a continuación:

Cuadro 3:
Comportamientos típicos de la FAC y la FACP (ACF Y PACF)

Proceso	FAC	FACP
AR (p)	Convergencia a cero, con comportamiento dictado por la ecuación para (tienen cola o son sinusoidales)	Solamente las primeras p autocorrelaciones parciales son distintas de cero. (su FACP corta en p)
MA(q)	Sólo las primeras q autocorrelaciones son distintas de cero.	Sucesión infinita convergente a cero.
ARMA (p,q)	Comportamiento irregular de las primeras q autocorrelaciones y después convergen a cero de acuerdo con	Sucesión infinita convergente a cero.

Para facilitar la identificación de modelos a partir de valores muestrales, se reproducen a continuación las figuras 4.5 "a", "b", "c", "d", y "e", del capítulo 4 del libro del Dr. Guerrero (Guerrero 1991a), donde se muestran los comportamientos típicos de las FAC y de las FACP tanto teóricas como muestrales (con muestras de tamaño $N = 100$) para algunos modelos que ocurren con frecuencia en la práctica. En la misma figura aparece lo que se denomina como "región admisible" para los parámetros del modelo, dicha región se obtiene como consecuencia de las condiciones de estacionariedad e invertibilidad para los modelos respectivos. Por ejemplo, un modelo AR (1) es estacionario, si $-1 < \phi < 1$, por lo cual la región admisible puede representarse gráficamente como el intervalo:

Región admisible para un modelo AR (1)

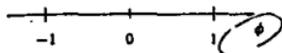


Figura 4.5(a)

Funciones de autocorrelación y de autocorrelación Parcial
teóricas y muestrales para modelos admisibles AR(1)
 MODELO FAC FACP

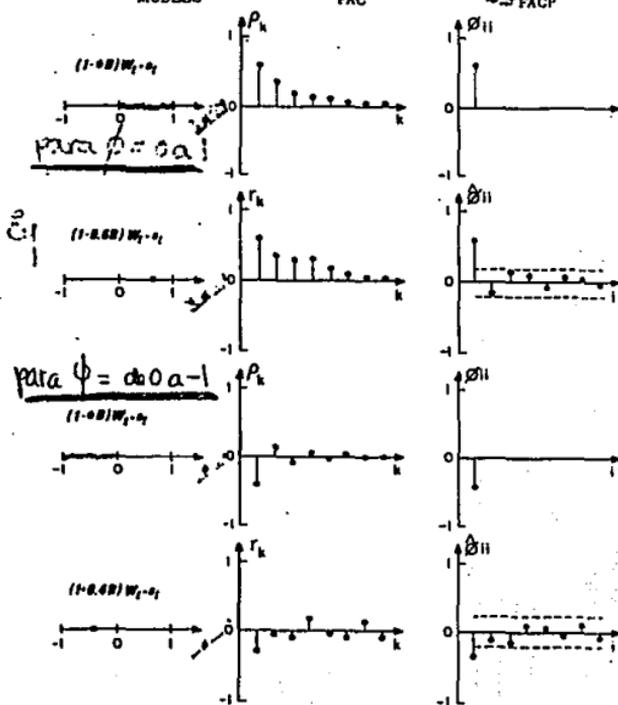


Figura 4.5(b)
 Funciones de autocorrelación y de autocorrelación parcial
 teóricas y muestrales para modelos admisibles MA(1)

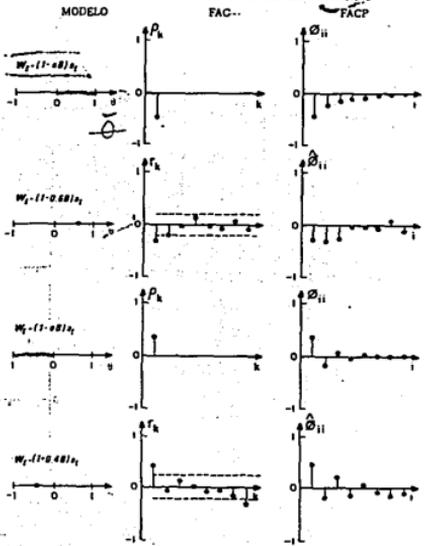


Figura 4.5(c)
 Funciones de autocorrelación y de autocorrelación parcial
 teóricas y muestrales para modelos admisibles AR(2)

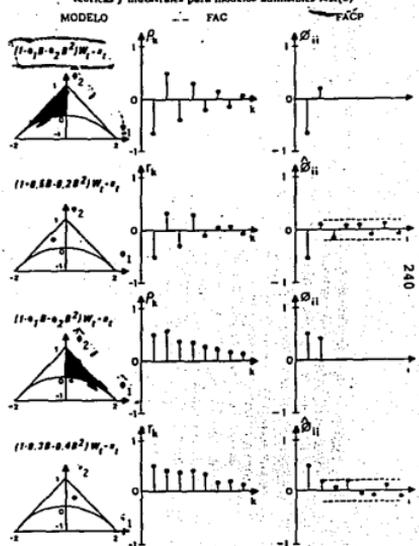


Figura 4.5(c)

Funciones de autocorrelación y de autocorrelación parcial teóricas y muestrales para modelos admisibles AR(2)

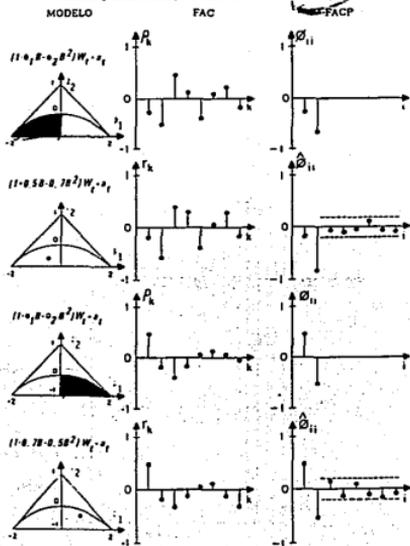


Figura 4.5(d)

Funciones de autocorrelación y de autocorrelación parcial teóricas y muestrales para modelos admisibles MA(2)

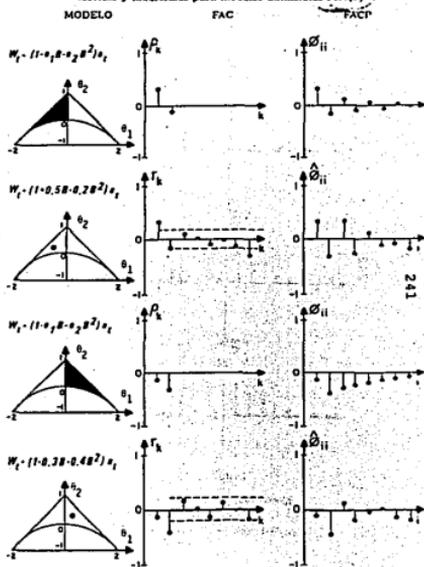


Figura 4.5(d)
Funciones de autocorrelación y de autocorrelación parcial
teóricas y muestrales para modelos admisibles MA(2)

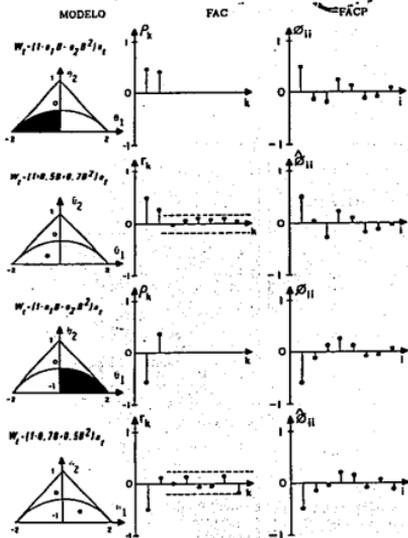


Figura 4.5(e)
Funciones de autocorrelación y de autocorrelación parcial
teóricas y muestrales para modelos admisibles ARMA(1,1)

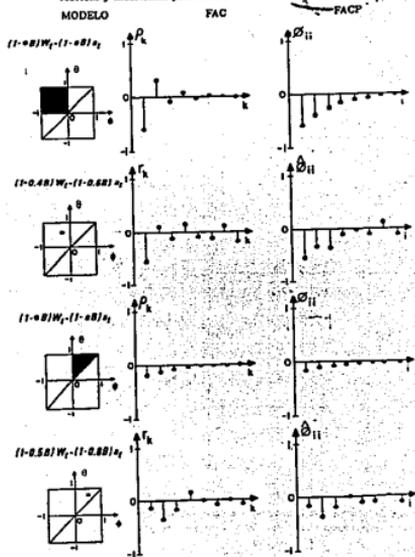


Figura 4.5(a)

Funciones de autocorrelación y de autocorrelación parcial
teóricas y muestrales para modelos admisibles ARMA(1,1)

MODELO

FAC

FACP

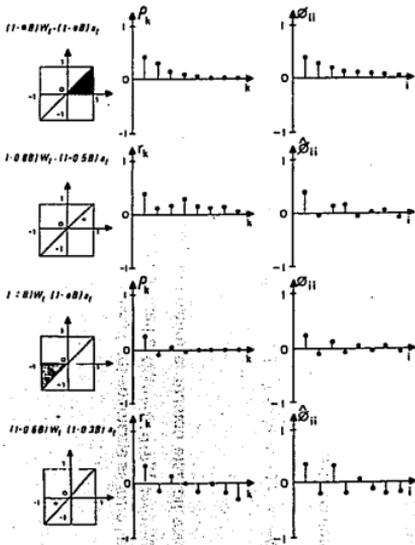


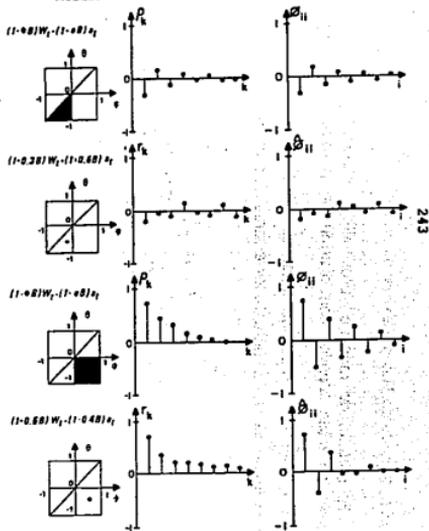
Figura 4.5(b)

Funciones de autocorrelación y de autocorrelación parcial
teóricas y muestrales para modelos admisibles ARMA(1,1)

MODELO

FAC

FACP



Anexo b: Método de Máxima Verosimilitud.

El método de máxima verosimilitud, supone que $\{a_t\}$ es un proceso de ruido blanco, con distribución normal, media cero y varianza σ_a^2 . A partir de estas consideraciones, se obtiene la función de densidad conjunta de los errores aleatorios:

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

Ahora bien, si se considera a la variable

$$W_t = \nabla^d T(z_t) \quad \dots (b.2)$$

entonces la expresión

implica que

$$a_t = W_t + f_1 W_{t-1} - \dots - f_p W_{t-p} - \theta_0 + \theta_1 a_{t-1} + \theta_q a_{t-q} \quad \dots (b.3)$$

esto permite obtener la función de densidad conjunta de la variable $W_{d+p+1}, W_{d+p+2}, \dots, W_N$ a partir de la función de densidad conjunta de los errores aleatorios (b.1) como

$$f(W_{d+p+1}, W_{d+p+2}, \dots, W_N)$$

$$= f(a_{d+p+1}, a_{d+p+2}, \dots, a_N) \prod_{t=d+p+1}^N \left| \frac{dW_t}{da_t} \right|$$

$$= (2\pi)^{\frac{(-N-d-p)}{2\sigma_a^2}} \exp \left\{ -\frac{1}{2\sigma_a^2} \sum_{t=d+p+1}^N (W_t + \phi W_{t-1} + \dots + \phi_p W_{t-p} - \theta_0 - \theta_1 a_{t-1} + \dots + \theta_q a_{t-q})^2 \right\}$$

.....(b.4)

Esta función de densidad permite calcular probabilidades de la distribución normal multivariada una vez que se conoce a los

parámetros $\bar{\phi} = (\phi_1, \dots, \phi_p)'$, $\theta_0, \bar{\theta} = (\theta_1, \dots, \theta_p)'$ y σ_a^2 .

En realidad, lo que se conoce es una muestra de

$$\bar{W} = (W_{d+p+1}, W_{d+p+2}, \dots, W_N) \quad \text{.....(b.5)}$$

y lo que se desconoce es $\bar{\phi}, \theta_0, \bar{\theta}$, y σ_a^2 , por ello se considera que la (b.4) es una función de verosimilitud de los parámetros, que depende del vector de observaciones \bar{W} , esto es,

$$L(\bar{\phi}, \theta_0, \bar{\theta}, \sigma_a^2 / \bar{W}) = (2\pi)^{\frac{(-N-d-p)}{2\sigma_a^2}} \exp \left\{ -\frac{1}{2\sigma_a^2} \sum_{t=d+p+1}^N \left(\frac{\bar{\phi}, \theta_0, \bar{\theta}}{\sigma_a^2} \right)^2 \right\}$$

.....(b.6)

con

$$S(\bar{\phi}, \theta_0, \bar{\theta}) = \sum_{t=d+p+1}^N a_t^2$$

=

$$\sum_{t=d+p+1}^N (W_t - \phi_1 W_{t-1} - \dots - \phi_p W_{t-p} - \theta_0 - \theta_1 a_{t-1} + \theta_q a_{t-q})^2$$

.....(b.7)

dicha función de verosimilitud $L(\bar{\phi}, \theta_0, \bar{\theta}, \sigma_a^2 / \bar{W})$ debe

maximizarse con respecto a los parámetros, para obtener la representación (b.1) mas apropiada de la serie $\{W_t\}$ en estudio.

La estimación de parámetros de modelos ARIMA está conformada por una parte teórica muy fuerte. Para profundizar sobre el método de máxima verosimilitud, debe consultarse el libro del Dr. Guerrero, en el capítulo, 4.