

24021
39



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
CAMPUS ACATLÁN**

"ANÁLISIS Y APLICACIONES DE SERIES DE TIEMPO"

T E S I S
PARA OBTENER EL TÍTULO DE:
**LICENCIADO EN MATEMÁTICAS APLICADAS
Y COMPUTACIÓN**
P R E S E N T A :
JOSÉ HUGO MAX NAVA KOPP

Asesora: M. en E. María del Carmen González Videgaray



Noviembre 2003

A



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

AVANCE a la Dirección General de Bibliotecas de la
UNAM a través de la oficina de la Dirección de Bibliotecas de
Ciencias Exactas y Naturales

NOMBRE José Hugo Max

Nava Kopp

FECHA 3-Dic-2003

FIRMA Hugo Nava Kopp

*Esta tesis se la dedico a mis padres,
Christel Kopp de Nava y Hugo Nava López*

*Gracias por las infinitas oportunidades y una excelente educación
Les debo todo.*

Agradecimientos

- A mi papá y mamá por darme su apoyo moral, económico y didáctico siempre: antes, durante y después del trayecto de esta tesis
- A mi abuelita Lola, en paz descanse, por enseñarme lo que escribía
- A mi abuelo Pepe, en paz descanse, por sus buenos libros que pude leer
- A Mari Carmen González Videgaray, por apoyarme con entusiasmo a través de todo el trayecto de esta tesis y darme excelentes ideas
- A mi tío Benito, por mostrarme competencia cuando escribía su tesis
- A mis hermanos y hermanas. Beba, Clem, Peter, Mariano y Alice, por enseñarme a concentrar
- A mi tío Luis, por incentivar me, aunque no se diera cuenta
- A mi Abuelita Alice por interesarse en el término de esta tesis y decirme que escribo bien
- A Rosario Hurtado Ramos, por hacerme creer que podía terminar esta tesis

Índice

| | | |
|---------------------|---|-----------|
| INTRODUCCIÓN | | 1 |
| CAPÍTULO I | Marco de referencia | 19 |
| | 1.1 Necesidad de pronosticar | 19 |
| | 1.2 Métodos de pronósticos. | 20 |
| | 1.2.1 Métodos cuantitativos, 21 | |
| | 1.2.2 Métodos cualitativos, 23 | |
| | 1.3 Series de tiempo | 24 |
| | 1.4 Aplicaciones de los pronósticos. | 30 |
| CAPÍTULO II | Pronósticos de series de tiempo por suavizamiento | 31 |
| | II.1 Suavizamiento exponencial | 39 |
| | II.2 Suavizamiento exponencial doble | 42 |
| | II.3 Suavizamiento exponencial triple. | 46 |
| CAPÍTULO III | Pronósticos de series de tiempo con variación estacional | 49 |
| | III.1 Descomposición multiplicativa | 49 |
| | III.2 Método de Winter | 60 |
| | III.3 Series de tiempo con variación aditiva. | 62 |
| CAPÍTULO IV | Métodos cualitativos | 65 |
| | IV.1 Métodos exploratorios | 67 |
| | IV.2 Métodos normativos | 76 |
| CAPÍTULO V | Pronósticos y planeación | 81 |
| | V.1 Planeación y decisiones | 82 |
| | V.2 Comparación y selección de métodos | 83 |
| | V.3 Pronósticos basados en el juicio personal | 84 |
| CAPÍTULO VI | Conceptos fundamentales | 93 |
| | VI.1 Procesos Estocásticos y series de tiempo | 94 |
| | VI.2 Función de autocorrelación | 95 |
| | VI.3 Función de auto correlación parcial | 98 |
| | VI.4 Periodograma | 99 |
| | VI.5 Procesos de ruido blanco | 99 |
| | VI.6 Estimación de media. ACF y PACF | 101 |
| | VI.7 Procesos lineales generales | 101 |

| | | |
|----------------------|---|------------|
| CAPÍTULO VII | Modelos estacionarios | 103 |
| | VII.1 Procesos autorregresivos | 107 |
| | VII.2 Procesos de medias móviles | 111 |
| | VII.3 Procesos mezclados | 116 |
| CAPÍTULO VIII | Modelos no estacionarios | 119 |
| | VIII.1 Procesos con tendencia | 119 |
| | VIII.2 Modelos autorregresivos integrados y de medias móviles | 122 |
| | VIII.3 Estabilización de la varianza | 125 |
| CAPÍTULO IX | Pronósticos óptimos con los modelos estacionarios y no estacionarios | 129 |
| | IX.1 Pronósticos de error cuadrático mínimo | 129 |
| | IX.2 Cálculo de pronósticos | 138 |
| | IX.3 Actualización de pronósticos | 138 |
| Capítulo X | Metodología de Box-Jenkins | 143 |
| | X.1 Identificación | 146 |
| | X.2 Estimación | 153 |
| | X.3 Diagnóstico | 161 |
| | X.4 Pronóstico | 168 |
| CAPÍTULO XI | Modelos con variación estacional | 173 |
| | XI.1 Modelos puramente estacionales | 177 |
| | XI.2 Modelos multiplicativos | 180 |
| CAPÍTULO XII | Aplicaciones | 185 |
| CONCLUSIONES | | 195 |
| BIBLIOGRAFÍA | | 187 |

INTRODUCCIÓN

"Quien encuentra grandes soluciones es quien enfoca los problemas con más amplitud, quien adopta una actitud filosófica ante la ciencia, es decir, quien sitúa el problema dado en su contexto más amplio y está dispuesto a revisar los fundamentos mismos de las teorías o de las técnicas".

MARIO BUNGE

"Voltear a ver" las cosas puede tener un efecto sobre lo observado. Y no sólo en el sentido de la vista. En "Momo" de Michael Ende se habla de los segundos que dejan de existir con sólo mencionarlos. O más formalmente Jean Paul Sartre nos propone experimentos de tratar de voltear a ver a la conciencia. Pero cada vez que la volteamos a ver o nos referimos a ella, ésta se pasa del otro lado, del lado que está observando dejando del lado observado sólo un reflejo o "espejo" de ella misma como Sartre le llama. Así Descartes, según Sartre, se equivocó al concebir al que habla y de quien se habla como el mismo ente cuando afirmó que "Pienso, luego existo". Incluso añade que de lo que está hablando no es un ente ni es la conciencia, sino una acción ¡un verbo!, lo que el llama una "duda". Así lo que Descartes está tratando de decir según Sartre es "Estoy conciente de que dudo, luego existo" (*Jean Paul Sartre, "Being and Nothingness", 1966, p.XI, Washington Square Press, E.U.*).

Las pruebas de la afectación que sufre lo percibido en la física son bastante interesantes, como se ve con el Principio de Incertidumbre de Heisenberg. En la Sociología se presenta el mismo efecto. Según la maestra MariCarmen González Videgaray esto se puede ver en los casos en los que se realiza una encuesta y la población misma bajo estudio se entera de los resultados, con lo que infaliblemente se ve afectado el objeto estudiado.

Pero como Mario Bunge nos explica, esto acompaña a los científicos en su trabajo y forma parte de su vida diaria. Deben estar completamente concientes de ello para entender que "ninguno de ellos aprehende su objeto tal como es, sino tal como queda modificado por sus propias operaciones...[los científicos] actúan haciendo tácitamente la suposición de que el mundo existirá aún en su ausencia, aunque, desde luego, no exactamente de la misma manera" (*Bunge, 1996*).

Esto puede resultar de interés porque ésta es una tesis en donde se trabaja con muestras y observaciones. Se debe tomar en cuenta el origen epistemológico y gnoseológico de la forma en que se trabaja con el muestreo, para que se puedan comprender mejor los resultados derivados de ello y tener una visión interpretativa más amplia y profunda de las conclusiones a las que llevan las observaciones.

Otro de los principios interesantes para esta tesis es uno universal: la economía de recursos. Una araña que vive en una cueva acuática donde no llega luz solar alguna no tiene ojos. ¿por qué?. Se puede comprobar experimentalmente –repetiendo el experimento en un ambiente controlado– y empíricamente –observando la naturaleza–, que nada en la evolución puede darse el lujo de contar con "cosas de más" (*Herbert Von Ditfurth, "No somos sólo de este mundo", 1983, Planeta, México*).

Esto nos lleva al principio de la navaja de Ockham, que se utiliza, como veremos

más adelante, dentro de los criterios para seleccionar modelos. William de Ockham (1285-1349) fue un monje franciscano que no aprobaba la lujuria con la que vivían los Papas y vivía él mismo con lo mínimo para sobrevivir. Estudioso de la lógica se percató que la ciencia, aunque trabaja con singularidades, al generalizar elimina lo circunstancial y “lo que está de más” conservando únicamente lo *universable* que a su vez tenga un anclaje con la realidad. El principio de Ockham dice textualmente “*Entia non sunt multiplicanda praeter necessitatem*” (Las entidades no deben multiplicarse innecesariamente).

Por extensión de la naturaleza también nuestro cerebro utiliza el mecanismo de ahorro de energía, como se puede observar patentemente en el uso del **contexto**, donde se **toman cosas por sentido**, lo que a su vez nos permite subir en la escala de la abstracción y descripción.

Un ejemplo podrían ser las clases de Álgebra Lineal, en donde el maestro a lo largo de su clase va sentando las bases para una explicación más elevada, de tal manera que alguien que haya permanecido en la clase desde el primer minuto, puede comprender –en potencia– lo que el maestro dice al final de la explicación. Para alguien que arriva a la clase en los últimos minutos sentirá que se ha perdido varias semanas de clase. Sentirá una invasión de términos que desconoce, o que conoce pero están siendo usados en un contexto en el que pierden el significado que él conocía y sentirá por ende que la información es caótica. Es por eso que podemos considerar al contexto como el **anticoas**.

Vale redundar en que si no sucediera este “tomar por sentido” sería imposible avanzar en una dirección de abstraimiento, porque cada vez que se repitiera un término nuevo se tendría que repetir la explicación. Pero esto no es necesario y los términos creados pueden incluso reutilizarse –en conjunción o no con los *viejos y conocidos*– para crear aún nuevas explicaciones para nuevos términos. Se puede seguir así **infinitamente** porque el uso de niveles descriptivos y su capacidad recursiva permiten que no haya límite alguno en la creación de términos.

Esto quiere decir que nuestro lenguaje, por ser capaz de describirse a sí mismo, puede crecer infinitamente en simbología (palabras). “Describirse a sí mismo” significa que una vez que se explicó una nueva palabra, ésta puede pasar del lado *explicado* al *explicativo*, enriqueciendo con esto la capacidad expresiva del lenguaje.

El concepto que se ha expuesto como “voltear a ver” es el principio de todo análisis serio y científico que se desee hacer. Trae implícito el concepto de **objetividad**, que es el de no ser prejuiciosos. Es decir, voltear a ver las cosas tal y como son, sin parcialidad sobre lo que estamos viendo, lo cual nos hace adquirir una actitud impersonal.

La objetividad trae además muchos atributos adicionales consigo. Cuando analizamos objetivamente podemos encontrar enfoques diferentes para lo que se está analizando: verlo con perspectivas varias y relacionarlo con otras que conocemos.

Así por ejemplo alguna vez a alguien se le ocurrió que todo podía ser visto como una cola de espera: la gente que cruza la calle está en espera de la luz verde; las bibliotecas tienen libros que están en una cola de espera para que sean prestados; un silla está a la espera de ser rota o tirada a la basura, etc. Con esto surgió la teoría de colas. Después a alguien se le ocurrió que todo podría ser visto como una red, surgió la teoría de redes. Así

también surgió la teoría de sistemas y otras*. Podemos ver que estos nuevos enfoques se pudieron obtener gracias a que el observador se salió del **punto de vista** con el que se contaba y encontró una relación con algo totalmente diferente en apariencia, pero que para él (o ella) tenía mucha lógica.

La impersonalidad de la objetividad nunca puede ser completa y es considerada por algunos pensadores como un sueño guajiro de los científicos. También todos los enfoques diferentes que se mencionaron son incompletos y ninguno es una respuesta a todos los problemas. De hecho, dependiendo de *para qué* se necesiten, serán algunos modelos más adecuados que otros y no se puede decir que uno sea mejor que otro en general y ni que con el tiempo los modelos se vuelvan mejores. Esto último puede parecer en un principio bastante contraintuitivo.

La objetividad y el "tomar por sentado" se pueden ver como antagonistas, donde ambas tienen algo que aportar al pensamiento científico. El primero puede ser entorpecedor cuando se está tratando de seguir una línea de pensamiento o a la intuición; nos puede hacer ser demasiado metódicos ya que nos aleja de lo ilógico. A su vez el tomar por sentado nos impide ver ciertas cosas a las que estamos acostumbrados y que podrían ser parte de la solución del problema: por ejemplo los judíos no tenían un nombre para *su* religión, aunque si tenían una para las religiones *de los demás*. Asimismo los griegos, cuando fueron ellos los extranjeros por primera vez, se percataron de ciertas cosas de sí mismos que nunca siquiera se habían imaginado en encontrar. Con esto les dio por analizar todo: desde lo sagrado hasta lo real.

Independientemente de que se tenga duda de cuándo es mejor tomar una u otra actitud, no cabe duda de que estos dos conceptos son herramientas para la ciencia y se puede hacer uso de ambas.

LEY DE LA CONSERVACIÓN DE LA ENERGÍA

Alguna relación existe entre la energía y la recursividad en el universo. Esta relación es mucho más profunda de lo que uno podría intuir a primera vista y nos lleva a la esencia del estudio científico y a los orígenes del Universo.

Un ejemplo de fórmula recursiva es la ley de gravitación de Newton. Cuando se aplica al movimiento de dos cuerpos celestes, permite determinar la forma de sus trayectorias. Si se define a P_t como la posición de un cuerpo con respecto del otro en el instante t y a V_t como la velocidad de movimiento relativa a otro cuerpo también en ese instante, es posible hallar la fuerza gravitacional que actúa sobre ese cuerpo en ese instante, la cual produce una aceleración $A_t = F_t/M$, donde M es la masa del cuerpo.

Conocida la aceleración A_t y la velocidad V_t del cuerpo es posible determinar la velocidad V_{t+1} para el instante siguiente que se denota por $t+1$, ya que $V_{t+1} = V_t + A_t$. Con una segunda ecuación de movimiento se obtiene la posición del cuerpo para el instante $P_{t+1} = P_t + V_t$. Así P_{t+1} y V_{t+1} han sido definidos a partir de P_t y V_t haciendo uso de la recursividad. Al aplicar este algoritmo suficiente número de veces se puede determinar cualquier posición futura P_{t+n} a partir de la posición inicial P_0 y la velocidad V_0 .

Este comportamiento es extensible a muchos otros fenómenos físicos: partículas

* Ver Anexo I para una definición más detallada de estas teorías

atómicas en los campos eléctricos, la cristalización de un copo de nieve, el agua moviéndose en un canal, etc. De tal forma que con conocer el estado del objeto en un instante es posible determinar su estado en el futuro. Si este comportamiento fuera a su vez extensible al Universo se podría decir que, a pesar de ser muy complejo, el Universo puede ser descrito por un modelo, aunque estático, recursivo.

Piénsese un poco en esto. Un modelo recursivo tiene la capacidad de poderse modificar a sí mismo si así lo dicta él mismo, con lo cuál su estaticidad sería aparente: sólo lo suficiente como para que lo planteemos en un instante dado, pero es *intrínsecamente* dinámico.

LA TESIS DE CHURCH

Un matemático de la Universidad de Princeton, llamado Alonso Church planteó una tesis que determinó los límites de la recursividad en el planteamiento de ecuaciones. Nos dice que cualquier ecuación que puede ser definida de cualquier forma, puede ser definida por una ecuación recursiva que además no contiene ciclos interminables. Esta tesis se sustenta en bases muy sólidas y todos los desarrollos matemáticos posteriores han confirmado su validez (*Revista "Ciencia y Desarrollo", capítulo "La recursividad del Universo", Enrique Calderón Alzate y Jorge Negrete M. p. 39-48*).

Lo interesante del asunto es que entonces se pueden determinar todas las leyes físicas del Universo de manera recursiva. Y se puede ir más lejos y decir que incluso todas estas leyes se pueden determinar a su vez por una fórmula recursiva y afirmar que el Universo está determinado por un modelo recursivo.

Retomando la ley de la gravitación, pensemos en la tierra girando alrededor del sol: después de avanzar cierta distancia regresa al punto de partida P_0 . Es decir uno de los P_{i+n} se convierte P_0 . Sin ser físicos, resulta curioso que después de tanto alejarse de su punto de partida —es más, justamente cuando se encuentra más lejos del punto de partida— el planeta súbitamente empieza *da capo*, de su punto de partida. Eso ya lo conocemos: es un círculo. Pero no seamos demasiado concluyentes y dificultémosnos la vida por un momento.

Gracias a que es posible modelar en la computadora la ley de atracción de Newton, se puede experimentar lo que sucedería en el caso de que la tierra disminuyera su velocidad o que la aumentara. Contando con algunas nociones básicas de física se podría concluir inmediatamente que en el primer caso, desminuyendo la velocidad, la tierra debería formar un espiral hasta chocar contra el sol. En el segundo debería formarse un espiral pero de salida, que dispare tarde o temprano a la tierra lejos del sol. Sin embargo ninguna de estas dos ideas es estrictamente correcta; bajo ciertos límites, la elipse que forma la tierra alrededor del sol cambia su forma —se estira y se encoge— pero en ambos casos la tierra retorna siempre a su punto de partida e incluso ¡lo hace en el mismo lapso de tiempo!

Esto se debe a una ley llamada de la Conservación de la Energía, estudiada en un principio por Galileo. De acuerdo con ella la posición y velocidad inicial de un cuerpo determinan no sólo la posición siguiente, sino incluso toda su futura evolución hasta la eternidad.

SISTEMAS AXIOMÁTICOS Y CONSERVACIÓN DE LA ENERGÍA

Este patrón de comportamiento se puede encontrar en algo mucho más conocido a los matemáticos, que es el de los sistemas axiomáticos. Éstos, con los axiomas y unas

cuantas reglas de formación, pueden crear una innumerable cantidad de teoremas, todos y cada uno de ellos con la misma característica común: la verdad. La verdad se puede ver como la "energía" que es conservada a través de toda deducción: pasa de los axiomas a los teoremas.

Estaremos manejando continuamente fórmulas recursivas en los modelos que se verán en la tesis. Para ello es valioso contar con este contexto de conceptos sobre todo el del balance energético de recursos, ya que se pueden comprender y ver que algunos de los conceptos que utilizan estos modelos como la estacionaridad tienen un respaldo teórico. Por ejemplo unos modelos llamados "AR" (por autorregresivos) son recursivos y si no existe un balance en su manejo de energía, se disparan y se hacen exponenciales.

CONCEPTOS EPISTEMOLÓGICOS

Según Noam Chomsky el lenguaje es el producto de la inteligencia humana que es, por el momento, lo más accesible para ser sujeta a estudio. Sin duda alguna nos encontramos en una época en la que —a diferencia de los siglos XVII, XVIII y XIX en los que formas de expresión indescriptibles, como el arte, formaban parte de la cultura— reina un paradigma positivista lógico donde todo debe ser sujeto al análisis científico y buscar la forma de expresarlo lingüísticamente, o sea, simbólicamente. Lo que ha llevado a que —en lugar de tener un apogeo de obras de arte como hace apenas un siglo— se sujeten a análisis riguroso incluso las obras de arte, el funcionamiento de la mente y en resumen todo lo que nos rodea, con el fin de resolver la problemática que éstas nos plantean: descifrar aquello que en apariencia nos está tratando de decir algo pero que no podemos verbalizar o cuantificar, todavía, con claridad.

Es por eso que se dice que estamos en una "era de la información" donde todo debe poder ser transcrito a alguna forma de lenguaje reproducible.

Es nuestro deber como estudiantes de Matemáticas Aplicadas y Computación entender a fondo ciertos conceptos clave, ya que son conceptos que engloban muchos conceptos y están en los escalones más altos de nuestro contexto de matemáticos aplicados: es decir, forman parte de nuestra jerga y alguien que no estudie lo que nosotros estudiamos no las conoce y difícilmente las puede comprender. Son conceptos a los que han tenido que pasar muchas generaciones de estudio e investigación científica para llegar a ellos y conceptualizarlos. Así que sería un desperdicio no hacerlo —desde un punto de vista pragmático —pero más que nada es nuestro deber —desde un punto de vista ético— el entenderlos en su plenitud e investigarlos a fondo.

Inferencia e inducción: generalizar a alguna ley o teorema a partir de algunas cuantas observaciones. Un caso muy claro es el de Sherlock Holmes. Él *induce* que algo sucedió de una cierta manera a partir de sus observaciones, de manera que pareciera que echa a volar la imaginación con los datos observados en mente y de ahí concluye algo que cuadra perfectamente con lo observado. Pero está claro que lo único real con lo que cuenta es con lo que sus cinco sentidos le proporcionan y no sabe absolutamente nada acerca de lo que **en realidad** sucedió.

Otro caso es cuando decimos que un político tomó tal decisión porque *inducimos* que en las esferas del poder político se está tramando tal o cual cosa. Pero se puede decir que somos "gusanos" y están sucediendo realmente cosas que *ni nos imaginamos*.

Esto de que *ni nos imaginamos* de hecho resulta ser mucho más descriptivo de lo que aparenta. Immanuel Kant nos dice que la razón "extiende en vano sus alas para ir más allá del mundo de los sentidos con el simple poder de la especulación". Es decir, incluso nuestras cavilaciones más *alocadas* se quedan cortas con lo que en realidad está sucediendo, porque estamos limitados por nuestros sentidos e incluso nuestra mente misma.

Dado que la razón es lo mejor con lo que contamos, tenemos que estar constantemente induciendo cosas a partir de lo que observamos. Así decimos que un proceso se comporta bajo las reglas de un cierto modelo, aunque en realidad este proceso podría estar funcionando bajo el régimen de otro modelo, pero por el momento (por un cierto tiempo en el cual el modelo *funciona*) el modelo encontrado nos resulta "verdadero". Aunque después desechemos el actual modelo y progreseemos, acercándonos más al modelo *verdadero*. Esto, lógicamente, suponiendo que se puede encontrar un modelo verdadero, un modelo que no necesite más correcciones en el futuro. Pero como veremos en un instante un modelo estático por formación está condenado al fracaso.

TIEMPO

Hagamos el experimento mental que propone Kant mismo. Imaginemos que no existe nada, dudemos de absolutamente todas las cosas que nos rodean. Por más que tratemos no podremos eliminar ni el concepto de conciencia ni el de tiempo ni el de espacio. Esto puede parecer curioso, ¿por qué no podemos destemporizarnos aunque sea temporalmente?, ¿porqué el tiempo forma parte nuestra de una manera tan innata?. Borges nos dice que el tiempo es una invención nuestra: Federico Reyes Heróles nos dice que esta invención es una invención terrible que nos atrapa, cargamos con ella a todas partes, somos incapaces de bajarnos de ella, somos sus esclavos. "somos incapaces de tener un instante de libertad total, en el cual podamos detener el incontenible flujo". Detener el tiempo, según Reyes Heróles, ha sido una de las grandes tentaciones del ser humano: detenernos y contemplar lo que ocurre "sin ser atrapados por el vértigo del movimiento" para poder "razonar sin prisa, leer con frialdad nuestra propia vida, ordenar nuestra mente para poder, en ejercicio de la razón, débil pero única herramienta que tenemos, optar, decidir y actuar en consecuencia. Abrir un paréntesis"(Reyes Heróles, 1999).

Si nos remontamos a nuestros orígenes evolutivos, veremos que nuestra concepción del tiempo es lo que nos hizo vencer a los animales en las praderas, lo que nos dio ventajas sobre sus mejores dotes para cazar. Aprendimos con ello las consecuencias de los actos y nos fue posible hacer *gedankenexperimenten* (experimentos imposibles de realizar en la realidad como por ejemplo imaginar que vamos en una nave espacial que viaja a la velocidad de la luz), experimentos mentales (como por ejemplo imaginar qué nos sucedería si nos robamos a la cría de un elefante delante de sus narices) o simulaciones (amarrar a alguien que no fuera de nuestra tribu o a algún animal a una liana de la que dudamos que soporte nuestro peso) que de ser llevados a la práctica, resultarían mucho más costosos y riesgosos (Cerejido, 1998).

El que seamos seres temporales trae como consecuencia una cierta secuencialidad de "lo vivido". Una cosa pasa, luego la otra. Dos cosas pasan al mismo tiempo, pero no

pueden pasar en el mismo espacio al mismo tiempo. Y si vemos, después de mucho tiempo de observación, que siempre que pasa A sigue B, sabemos que la próxima vez que suceda A podemos *esperar* que suceda B. Es decir, nuestra mente viene prejuiciada o preparada desde que nacemos a encasillar todo en "si...entonces...". Trae incrustada la causalidad.

CAUSALIDAD

Esto de "esperar" que algo suceda nos debería sonar conocido: el concepto de esperanza estadística. La estadística, aunque parezca improbable a primera vista, está muy relacionada con la filosofía de la ciencia. Karl Popper, un filósofo de la ciencia precisamente, nos dice que la ciencia se puede considerar como una máquina de teoremas, refutables (falseables) todos ellos, ya que están sujetos a lo que la experiencia dicte. Es decir, yo puedo hacer una aseveración científica basada en un par de observaciones, como por ejemplo "El agua hierve a 100°", pero alguien puede llegar a decirme que lo probó en casa y no le resultó, pero yo le digo "No, quítale la tapa, así no es como se debe hacer". ¿Cómo que así no es como se debe hacer?, ¿a qué hora lo dije en "El agua hierve a 100°"? Entonces lo que debería decir es "El agua hierve a 100° en un envase destapado", pero alguien puede telefonarme para decir que está en Quito y que ahí hirvió a menos de 90°. Con esto tendría que añadirle que el experimento sólo funciona a nivel del mar. Y así me puedo ir y cada vez que añada más información a mi aseveración (teorema), más comprobable se vuelve y por ende más falseable. Éste es un principio que manejamos en estadística: a mayor precisión, mayor intervalo de confianza. Según Popper, entonces, la aseveración con mayor nivel informativo de todas sería una descripción exacta del mundo y toda posible observación se tornaría una prueba de ella y una potencial falseación (Magee, 1975, p. 36).

El punto de mencionar la filosofía de la ciencia es que ésta concluye que toda la ciencia se remite finalmente siempre a los hechos, las observaciones. Desde cierto punto de vista todo el conocimiento científico depende de predicciones que se hacen sobre lo medible. Las leyes científicas que no se basan en predicciones comprobables, como la psicología, no pueden ser científicas. La ciencia se está basando constantemente en "si...entonces..." que funcionen porque han sido comprobados muchísimas veces, pero si dejan de funcionar en algún caso, veremos lo que sucede.

Si algo no *cuadra*, se analiza este suceso "aislado", porque *debería* cuadrar bajo el esquema teórico actual. O por el otro lado se puede optar por revisar el esquema bajo el cual se está trabajando: los axiomas que se conforman la base y los conceptos que se están tomando por sentado.

Por ejemplo cuando uno de los planetas de nuestro sistema, Urano, estaba comportándose "raro" bajo los supuestos que se manejaban en 1846, la órbita calculada de este planeta no coincidía con la observada es decir no cuadraba con las **predicciones científicas**, no se pusieron a revisión automáticamente todas las leyes, axiomas y teoremas que se tenían en la época. Por el contrario, se buscó la forma en que se ajustara ese hecho aislado al esquema actual. Al esforzarse de esta manera, fue posible resolver la discordancia: se supuso que tendría que existir un planeta más con lo que encajaría al esquema teórico, se pronosticó así la existencia de Neptuno (Pérez Ranzanz, p.62, 1999).

La otra opción sería revisar el modelo teórico, el **paradigma** que le llama Thomas Kuhn, pero como él mismo nos señala, esto sucede solamente en el menor de los casos, a saber, en épocas de crisis de un paradigma.

Por ejemplo, cuando se descubrió que el mundo no era plano con el descubrimiento de América, fue la gota que derramó el vaso de muchas cosas que ya no cuadraban, se vino abajo todo un aparato científico-religioso que tuvo que ser sujeto a revisión por completo. Las crisis de este tipo nunca cesarán en la historia porque la imperfección es inherente a todo paradigma (Pérez Ransanz, p.49, 1999). Otro ejemplo claro fue en la época en la que Santo Tomás llegó a rescatar el poder eclesiástico, ya que los avances de los estudiosos árabes estaban a punto de descalabrar a toda Europa Occidental: la realidad distaba mucho de lo que los religiosos enseñaban y Santo Tomás unificó todas las creencias devolviéndole así a la iglesia su poder abarcador.

Con tantas crisis que estamos pasando en la actualidad, Venezolana, Argentina, guerra en Irak y Corea del Norte, conflicto Palestino-Israelí, crisis energética, administrativa y financiero-contable en los Estados Unidos de Norteamérica, nos deberíamos preguntar si a lo mejor estamos en medio de un cambio de paradigmas: el desgaste de un modelo que ha dejado de ser apropiado

PRONÓSTICOS CIENTÍFICOS

Mencionamos pronósticos científicos. Esto debería llamar la atención, porque esta tesis es de pronósticos precisamente. Los pronósticos científicos son los que hacemos cuando aplicamos los teoremas de un paradigma científico. Simplemente el hacer un experimento aplicando una ley es un pronóstico, porque nadie nos está garantizando lo que sucederá, aunque se haya comprobado un millón de veces.

Este es un talón de Aquiles de la ciencia y una de los meollos también de la epistemología, que es en el que recae toda la estructura científica: contamos con que las cosas sucederán como creemos porque confiamos en la historia. Es decir, estamos utilizando el mismo esquema que en los pronósticos con series de tiempo, ¡pero para la ciencia misma!

Hume asestó un tremendo golpe al comprobar lo anterior: dice que la ciencia está haciendo castillos en el aire y que está tomando por sentado cosas que, en sentido estricto, no tiene por qué tomar por sentado. Y si lo analizamos un poco no suena muy disparatado. Qué tal si cada vez que hemos soltado una manzana a cierta altura del piso ésta siempre ha caído al suelo por pura *casualidad*, y no por causalidad. Causalidad sería decir que hay una cierta fuerza gravitatoria que se ejerce de la tierra con una gran masa hacia la manzana con una menor masa, lo que hace que la manzana sea atraída hacia la tierra y no a la inversa. Por casualidad sería decir que ha sido pura coincidencia y llegará algún día en que la manzana no caiga y se quede flotando en el aire y las leyes de la física que habíamos contraído en realidad no eran leyes de nada, sino simples conclusiones erróneas que trataban de establecer reglas para ciertos patrones observados.

Pero abogemos también para el otro lado. Con el conocimiento de las leyes de atracción de los cuerpos se pudieron poner cohetes en la luna: nunca se había realizado, pero como cuadró con todo lo que se sabía, fue posible. Es posible predecir la existencia de ciertos cuerpos en el espacio al conocer las leyes de atracción de los cuerpos y ver el efecto de un cuerpo que todavía no se descubre. Conociendo las leyes de los gases, se puede

estimar y predecir perfectamente las temperaturas y comportamientos de estrellas que nunca habíamos visto. Y después corroborarlo. Lo que hace que, realidad o no, es hasta cierto punto bastante consistente.

El concepto de predicción efectiva forma parte de las cualidades con las que tienen que tener los modelos científicos o paradigmas para poder establecerse: por ejemplo si en una época intraparadigmática, en donde hay modelos compitiendo, como sucedió con las matemáticas a principios del siglo XX, el que **prediga** mejor es el que se establece como nuevo esquema de trabajo para los científicos de esa área, como sucedió con Bertrand Russell y Kurt Gödel, donde el *Principia Mathematica* del primero sucumbió ante el teorema del segundo. De hecho en este caso el trabajo de Russell fue el escalón para el de Gödel, dándole al primero unos cuantos años de gloria. Pero nunca se pudo establecer como un paradigma científico porque presentó desde el principio demasiados huecos teóricos, como la imposibilidad de la autorreferencia.

De hecho podríamos simular que nos encontramos en esa época de crisis y tratar de encontrar el nuevo paradigma. Por un lado tenemos un sistema axiomático que pretende ser completo y consistente a la vez y clasifica todos los teoremas jerárquicamente, donde los teoremas inferiores no pueden hacer referencia a los superiores. Por otro lado tenemos un teorema verdaderamente genial, en donde las matemáticas mismas (la aritmética) prueban que ellas mismas son incapaces de decir si alguno de los axiomas que la conforman son verdaderos o falsos. Vemos que el mejor paradigma "pronosticador" efectivamente es el de Gödel, porque es absurdo eliminar la autorreferencia de nuestro lenguaje y nuestra forma de pensar, siendo sobre todo esta capacidad autorreferencial lo más valioso que se nos ha proporcionado a los seres vivos y es quizá una extensión de la evolución y el universo mismo. Suena alocado, pero en un momento regresaremos a este punto.

CAOS

Por otro lado está la cuestión del caos. Es un tema para varias tesis, pero por el momento démosle un acercamiento cauteloso y más que nada vertical para no divagar demasiado.

Primero, ¿por qué utilizamos el caos (ruido blanco) para predecir?. En parte ya se respondió esta pregunta. Dijimos que tenemos tres fuentes de incertidumbre: uno, la incertidumbre que siempre acompaña al futuro; dos, la irrepetibilidad de la historia y unicidad de cada uno de los eventos; y tres, el error inherente a la muestra. Pues existe un factor mucho más profundo y que tiene su origen en nada más y nada menos que en la formación del universo!.

La evolución trabaja a base de "intuición" y prueba y error, es decir heurísticamente, y cuando uno de sus intentos le sale "bien", lo guarda por medio de la genética y si le sale "mal", lo vuelve a intentar más tarde. Podríamos decir que la naturaleza no aprende de sus errores pero sí de sus aciertos. Bueno, el caso es que esos constantes e insistentes intentos, mutaciones, tienen algo de mágicos a pesar de ser tortuosamente lentos: son completamente aleatorios. La naturaleza, y el cosmos mismo se han dado cuenta de que el mejor predictor ¡es el caos! (Von Ditfurth, 1983).

Pero ¿por qué?. ¿No es acaso nuestro mundo una representación muy complicada pero representable con unas cuantas ecuaciones matemáticas que alguna computadora podría resolver?. No. El mundo no es un sistema muy complicado, es un *sistema complejo*.

En un sistema complejo las cosas suceden de una manera impredecible, aunque contengan períodos de estabilidad aparente. Es un sistema con interconexiones entre sus partes a diferentes niveles y donde los cambios toman, con el tiempo, un efecto exponencial.

EXPONENCIALIDAD

Por ejemplo el aleteo de una mariposa puede hacer que dentro de un año haya un ciclón en la costa japonesa. Esto es debido a la *inestabilidad exponencial* inherente a los sistemas complejos y la meteorología es precisamente un sistema complejo. Se sabe que en meteorología la amplitud de una perturbación se duplica cada tres días si nada contrarresta su desarrollo (Ekeland, 1992).

La exponencialidad ha fascinado a los pueblos durante siglos. Tómese por ejemplo la leyenda persa sobre el cortesano que ofendió a su rey un tablero de ajedrez y le pidió al rey que le diera un grano de arroz por el primer cuadro, el doble de eso para el segundo cuadro, el doble de eso para el tercero y así hasta recorrer los sesentaycuatro cuadros. Se dio cuenta el rey que tendría que dar al cortesano más arroz del que había en el mundo.

También existe un juego infantil francés que resalta otra peculiaridad de la exponencialidad: el aparente *repentino arribo a un límite*. Los nenúfares duplican su tamaño cada día y en treinta días pueden cubrir un estanque, ahogando cualquier otra forma de vida. Pero lo curioso es que no se nota este crecimiento. El vigésimoquinto día la planta sólo cubre $1/32$ del estanque, pero en el vigésimonoveno ya cubre la mitad, haciéndose notar de esta manera cuando sólo resta un día para salvar la demás vida en el estanque. También en las curvas de aprendizaje de nuestro cuerpo y mente suceden cosas similares.

Es fácil entender que en un proceso que contenga uno de sus componentes con crecimiento exponencial puede *temporalmente* aparentar que ese componente es un factor desdeñable para ser considerado en el modelo o incluso ni se identifica.

Vemos con estos ejemplos la importancia de la dinamicidad de los modelos y de lo poco prácticos que pueden resultar los modelos de predicción estáticos para horizontes que no sean cortos.

Dado que cualquier efecto puede transformarse en un potencial factor decisivo para el futuro de un sistema, resulta imposible poder modelar, y por ende pronosticar, un sistema complejo para horizontes lejanos porque se tendrían que tomar en cuenta todos los factores posibles, y aún esto sería insuficiente porque en el futuro aparecerían nuevas variables.

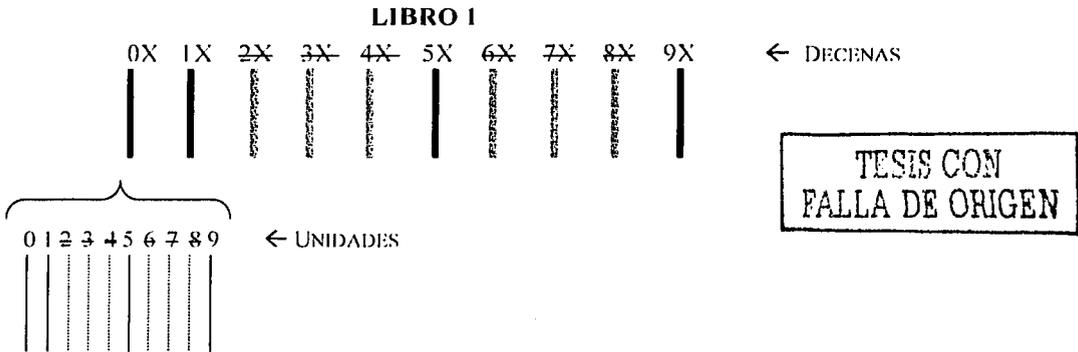
Una de las ventajas de los efectos exponenciales en la tecnología es que se pueden inducir los efectos para poder forjar el futuro. Por ejemplo, un automóvil que pesa varios miles de kilogramos se puede manejar con mover un volante que se puede mover con un solo dedo. Esto se debe a que se aprovecha la inestabilidad exponencial para inducir resultados. Sería como hacer que la mariposa aletee cada vez que queramos un tornado.

Por cierto, a este efecto se le llama "efecto mariposa" que es una metáfora de la sensibilidad a las condiciones iniciales. Esta metáfora la hizo famosa la teoría que fue propuesta en 1979 por Edward Lorenz del MIT con un trabajo denominado "Predictabilidad. ¿Hace el aleteo de una mariposa en Brasil que ocurra un tornado en Texas?".

FRACTALES

Para hacer una referencia a los fractales y su relación con la presente tesis, haré uso de una paradoja milenaria: la paradoja de Zenón. Nos plantea una pregunta muy interesante: ¿Existe el movimiento?. Es decir, si existe la comprobación de que los números reales entre 1 y 2 son infinitamente mayores a los Naturales, ¿cómo puede existir el movimiento entre estos dos puntos?. Imaginemos que tenemos una estación del metro A y la siguiente es la estación B del metro, dentro de un línea del metro llamada N que tiene una cantidad de estaciones infinita aleph cero. Resulta que entre la estación A y la B existen otras estaciones, que no pertenecen a la línea de metro N, pero están entre A y B. Estas estaciones resultan ser exponencial e infinitamente más que todas las estaciones de la línea N misma. Entre A y B existen aleph cero elevado al infinito (\aleph_0^{∞}) estaciones, es decir \aleph_1 . Pero, curiosamente, uno puede pasar del punto A al punto B sin problema.

Este concepto de encerrar algo de dimensiones infinitas en un espacio delimitable, y por ende finito, es extensible al de Fractales: figuras de dimensiones no enteras. Por ejemplo, imaginemos un libro de 100 hojas. Arranquémosle las que contengan los números 2.3.4.6.7 y 8 en sus páginas., con lo que quedarían sólo las que contengan al 0,1.5 y 9. Longitudinalmente los paquetes de hojas quedarían así:



Así tenemos que los paquetes de las decenas en realidad no son hojas, sino **etiquetas** de hojas. Por ejemplo el paquete de los "cincuentaytantos" (5X) contiene algunas de las hojas del 50 al 59, **no** todas: contiene la 50,51,55 y 59. Si se compara este paquete de hojas con el paquete de los paquetes —el libro en sí— veremos que hacen falta exactamente los mismos números. Así todas las decenas contiene la misma información (el mismo patrón) que el libro en su totalidad.

Ahora subamos el nivel de abstracción. Supongamos que el libro ahora tiene una cantidad infinita de hojas. ¿Cuántas hojas contendrá el libro después de arrancarle las hojas que determinamos?.

Empecemos desde abajo.

$$X=4^n$$

$$Y=4^{n-1}$$

Donde X es la cantidad de hojas resultantes después de arrancar; Y es el número de etiquetas; n es el número de "niveles" (en nuestro caso el orden: millares, centenas, decenas, etc.)

La cantidad de hojas que tiene el libro 1 era de $4(4)=4^2=16$, donde las etiquetas (en este caso las decenas) son 4. Si tenemos un libro de mil hojas iniciales, tendrá una cantidad de $4^3=64$ hojas y estará utilizando $4^2=16$ etiquetas después del arrancadero de hojas. Es decir, siempre estará utilizando 25% de recursos para **describirse** y el 75% es información pura, independientemente de la cantidad inicial de hojas en el libro, manteniendo así un uso constante de **energía**. Y lo más importante es que cualquier paquete de hojas a cualquier nivel, a excepción de las hojas reales, contendrá exactamente el mismo patrón.

Por ende, podemos decir que el libro estaría hecho de *muchos libros idénticos a él*, a excepción de las hojas reales, ya que no contienen nada debajo de ellas.

Cuanto mayor sea la cantidad de hojas iniciales, mayor será la cantidad real de hojas. Pero algo muy raro sucede cuando se tiene un libro de *infinita* cantidad de hojas: la **descripción** se vuelve total, del 100%, ¡y el contenido de hojas reales se vuelve nulo!

¿Cómo puede ser que, sin hojas reales, sólo etiquetas, pueda haber *algo* en este extraño libro infinito?.

Es un hecho curioso, pero lo podemos representar con una sencilla ecuación. Recordemos que X es la cantidad de hojas.

$$X=4(4(4(...)))$$

Si hacemos la siguiente sustitución para X:

$$X=\text{"cuatro por cuatro por cuatro así hasta el infinito"}$$

La cual es completamente válida, la ecuación queda:

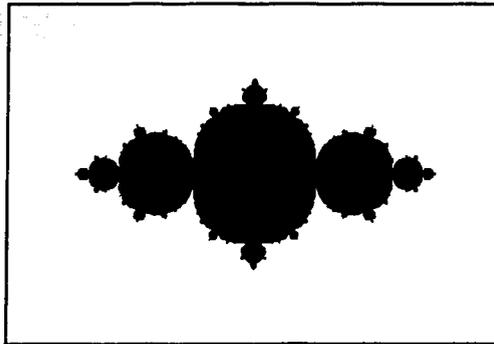
$$X=4(X)$$

Y ¿cuál es el único valor de X que resuelve esta ecuación?: cero. Por tanto no existen ya hojas reales y todo el libro está autodescribiéndose de tal forma que cualquier parte de él que tomemos (cualquier paquete) a cualquier nivel, contendrá el patrón que lo describe en su totalidad. Un tema muy relacionado con el de los fractales es el de la proporción áurea*.

Concluimos así que los fractales tienen una manera muy peculiar de comportarse. Este fractal en específico se llama atractor de Lorenz y tiene una dimensión de orden entre dos y tres.

Los fractales que tienen entre una y dos dimensiones son más fáciles de visualizar y se encuentran mejor relacionados con la paradoja de Zenón. Los fractales, a pesar de que tienen en teoría un perímetro infinito, ya que cualquier segmento de perímetro que se tome contiene a la figura "madre" misma, se pueden delimitar. Por ejemplo el siguiente fractal está delimitado por un cuadrado.

* Ver Anexo 2



TESIS CON
FALLA DE ORIGEN

Evolución, caos y pronósticos

Se mencionó que la Evolución utiliza el caos. Existe algo llamado el citocromo c. Éste tiene la característica de ser el padre de todas las enzimas y biopolímeros que se encuentran en la tierra y que son lo orígenes de la vida. Estos biopolímeros se encuentran incluso en los asteroides que han quedado a nuestro alcance y se encuentran en la misma proporción en los asteroides que en nuestro cuerpo!. Lo maravilloso de estos componentes bioquímicos es que son únicos a pesar de contar con una infinidad de posibilidades de existir. Por ejemplo, el citocromo c tiene una posibilidad de existir de 10^{130} cuando en el universo hay 10^{80} moléculas. Es decir, la posibilidad es casi nula de que surgiera esta citocromo y generara la vida, pero se creó. Hay una corriente de pensadores llamados los vitalistas que consideran esto como una prueba de Dios, pero si se analiza un poco tiene una explicación que el universo mismo nos da.

Comprueban varias cosas, entre ellas el hecho de la irrepitibilidad de un proceso, concepto que se maneja en series de tiempo cuando decimos que un proceso estocástico, si se pudiera repetir, sería completamente diferente. Lo mismo pasa con la Evolución: es imposible que, aunque fuera después de mil repeticiones del comienzo de la existencia del universo, las formas de vida fueran idénticas a las que tenemos. (Von Ditfurth, 1983)

Un flechazo en la oscuridad puede resultar ser más *apropiado* que un flechazo con "dirección". Toda decisión inherentemente contiene un objetivo, ¿quién dice cuál es el objetivo de la evolución?, o mejor dicho: suponiendo que el objetivo principal de la Evolución y del Cosmos existiera ¿cómo podemos decidir cuál es el mejor camino para lograr ese objetivo? ¿quién y con qué bases tomaría esta decisión?. Así vemos que para llegar al objetivo que tiene el Universo hace uso de la aleatoriedad y el aprendizaje. Resulta una sorpresa percatarse que el Universo ha estado sin nuestro grandioso cerebro, que no tiene más de 100.000 años, la inmensa mayoría de su existencia y ha podido arreglárselas perfectamente. No solo eso, sino que las capacidades que lo conforman, el aprendizaje y la creatividad, nos las ha proporcionado a través de nuestra mente. Lo que lleva a una interesante pregunta: ¿nos creó para que lo viéramos, lo analizáramos y nos diéramos cuenta de él? ¿creó el Universo una extensión de si mismo que cierre el círculo de la autodescripción completa, como sucede con los fractales?.

Decir esto tan a secas puede resultar un poco grosero para más de uno (decir que somos hermanos de los monos es una cosa, pero decir que somos fractales, cosas, ya pone

en duda la cordura de quien lo dice). Pero a lo mejor ha hecho falta decir que los fractales no son sólo curiosidades matemáticas; se encuentran literalmente en toda la naturaleza. Desde los árboles, las nubes, las formas de las galaxias, los copos de nieve, hasta nuestro cuerpo.

Con decir que nuestra mente es la única forma de inteligencia que conocemos que es capaz de hacer referencia así misma. La autorreferencia es algo que también se puede rastrear hasta los confines del origen del Universo. Entonces no es tan absurdo pensar que una de las capacidades intrínsecas del universo haya seguido su curso "natural" y creado una forma de vida que puede describirlo, cerrándose así el bucle de la autodescripción total.

Pero por otro lado resulta vergonzoso para muchos, Darwinistas sobre todo, pensar que quizá no somos nosotros los que cerraremos ese bucle: muy probablemente nuestra conciencia no esté todavía al nivel de concebir el *strqwzxyhgi** del Universo. Si ni siquiera hemos podido, después de más de 5.000 años que tenemos de seres culturales y sociales, llegar a un acuerdo unificador de lo científico con lo religioso, una cuestión que a todas luces muestra la limitación psíquica que todavía nos aqueja (decimos "todavía" por si todavía está evolucionando nuestro nivel de conciencia).

INTELIGENCIA

Pero la cuestión que saca a relucir el tema de que somos capaces de **percatarnos** de esto (que la creatividad y nivel de conciencia nos fue heredado por el Universo mismo) es la cuestión de qué es entonces la **inteligencia**, porque estamos en constante búsqueda de *vida inteligente*, no tomando en cuenta así lo "inanimado". Pensemos por ejemplo en el hígado. Este órgano realiza una complejidad de funciones tal que apenas recientemente se empiezan a entender la mayoría de ellas. Pero al hígado no lo consideramos *inteligente*.

Para los padres de la Inteligencia Artificial, como Marvin Minsky, la inteligencia es en parte la **velocidad** que imprimimos al desarrollo de cosas que nos son útiles. La perfección que adquiere el ala de un pájaro toma millones de años, pero a nosotros no nos tomó más de cien años percatarnos de su existencia, su funcionamiento y su utilidad, además de construirla e incluso mejorarla, para que satisficiera nuestras diversas necesidades de vuelo.

Pero esta definición de inteligencia está sujeta a discusión, por supuesto, pero suena interesante. Una característica más de la inteligencia es la de la solución de problemas. Cuando queremos resolver un problema, pero nos sale mal, no intentamos automáticamente empezar por otro camino sino que **analizamos** qué fue lo que no resultó. Es decir, una vez más hacemos uso de nuestro poder de conceptualizar a niveles y hacemos el trabajo de meta-resolvedores-de-problemas (analistas del funcionamiento de la resolución de problemas).

Una computadora, por ejemplo, no tiene esa capacidad, por lo menos no a escala considerable**.

Se considera que cualquier forma de ser inteligente que nos encontremos en el

* Se utiliza esta palabra sin sentido sintáctico aparente para sacar a relucir la fuerza del **contexto**: No fue necesario que dijera una palabra conocida, se *sobreentiende*

** Fuente: Marvin Minsky en Internet: <http://ftp.ai.mit.edu/pub/minsky/ComputersCantThink.rst>

Universo debe ser **resolvidor de problemas**. Todos los resolvidores de problemas estamos sujetos a las mismas limitaciones: de espacio, tiempo y materiales (administración de recursos escasos u optimización). Así que es muy probable que, de existir, comunicarnos con ellos sea posible dado que tenemos preocupaciones en común.

Cito a continuación características que el autor del libro "Gödel Escher y Bach, una eterna Trenza Dorada", Douglas R. Hofstadter, considera como propias de la inteligencia:

- 1) Responder muy flexiblemente a las situaciones.
- 2) Sacar provecho de circunstancias fortuitas.
- 3) Hallar sentido en mensajes ambiguos o contradictorios.
- 4) Reconocer la importancia relativa de los diferentes elementos de una situación.
- 5) Encontrar semejanzas entre varias situaciones, pese a las diferencias que puedan separarlas.
- 6) Descubrir diferencias entre varias situaciones, pese a las semejanzas que puedan vincularlas.
- 7) Sintetizar nuevos conceptos a base de conceptos viejos que se toman y se reacomodan de nuevas maneras.
- 8) Salir con ideas novedosas.

Ahora bien, al intentar crear máquinas inteligentes, los matemáticos, lingüistas y científicos de la computación se dieron cuenta de que no sabemos realmente cómo funciona nuestra propia mente. Desconocemos cómo funcionan los diversos procesos que nos hacen tomar una decisión, por ejemplo, o por qué a veces estamos de un humor y luego de otro y esto afecta nuestro juicio, o porqué cierta música puede transmitirnos mensajes o sentimientos.

La creatividad es la capacidad que menos entendemos. La cultura occidental nos enseña a intimidarnos ante nuestros Beethovens y Einsteins, pero ¿qué puede tener un genio que nosotros no tengamos?. En la cultura islámica, por ejemplo, el estudio científico está mucho más ligado a la deidad, lo que le da una absoluta ventaja evolutiva, filosófica y científica sobre la cultura occidental.

Los genios tienen las mismas habilidades que cualquier otra persona, sólo que en una combinación poco usual. Y también han aprendido, a algún cierto nivel mental, una mejor forma de aprender. Es decir, han aprendido a aprender de una manera muy eficiente.

Así nos hemos encontrado que nuestra inteligencia no es tan autoconciente como se pensaba en un principio. Se creía esto sobre todo antes de crear las computadoras que tenemos hoy en día. Se consideraba éste el problema: la falta de capacidad de las computadoras. Pero ahora que, en teoría, están listas y esperando a que les programemos nuestra forma de pensar, nos encontramos con que estamos muy pero muy distantes de lograr comprender nuestra propia mente.

Podría ser que las computadoras realmente no están tan listas como pensamos y este sea una parte del problema. Probablemente haga falta un cambio de paradigmas en las ciencias de computación, que nos ponga una escalera en la colina y podamos vislumbrar el camino a seguir.

Se empieza hablar ahora de computadoras cuánticas que en lugar de bits manejan

unas q-bits que en lugar de usar 0 y 1 usan cualquier real entre 0 y 1 (Revista Scientific American, Noviembre 2002 "Rules for a Complex Quantum World" Michael Nielsen, p.49-57). ¿Qué capacidad tendrán semejantes computadoras?

CONTENIDO DE LA TESIS

Esta tesis contiene en su primera parte, que corresponde al temario de Series de Tiempo I, 5 capítulos. En el primer capítulo introduce a los pronósticos: sus orígenes y cómo surgen como una necesidad natural para la sobrevivencia. Se introduce el concepto de serie de tiempo y se introduce también a los diferentes tipos de métodos de pronósticos.

En el segundo capítulo se ven los métodos de pronóstico más intuitivos, como son los de promedios móviles, para escalar a los de suavizamiento. Aquí se introduce al software que se utiliza para aplicar los métodos, que se llama Statgraphics. En el capítulo tercero se ve lo que es la aleatoriedad y de cómo afecta el modelado de las series de tiempo. Se presentan los dos métodos de descomposición, multiplicativa y aditiva, además de criterios de cuándo utilizar una o la otra. Se recurre a ejemplos para facilitar la explicación.

Los métodos cualitativos se estudian en el cuarto capítulo, tanto los exploratorios como los normativos. Algunos esquemas se muestran, obtenidos de otros libros y de aplicaciones sobre todo de casos no mexicanos, ya que no existen casos documentados de que se hayan aplicado estos métodos en México.

En el capítulo quinto, "Pronósticos y planeación" se plantea la relación que guardan las decisiones y los pronósticos, y de cómo convergen en la planeación. Se introducen los métodos de pronóstico que utilizan el juicio personal, que tratan de evadir y simplificar las dinámicas de las relaciones personales a la hora de emitir un juicio sobre una estimación a futuro.

La segunda parte comienza con los conceptos fundamentales necesarios para proseguir en la línea de estudio de las series de tiempo, pero ahora enfocadas a la metodología de Box-Jenkins. Empieza por explicar qué es un proceso estocástico y cómo se relaciona este concepto con las series de tiempo. Se obtienen los estimadores de la ACF y PACF que son los identificadores principales de variables para considerarse en el modelo bajo la metodología de Box-Jenkins. También se enseña lo que es el ruido blanco: un artificio inventado para comprender al máximo los fenómenos reales bajo un enfoque causalista.

En los capítulos 7 al 11 ya se entra a la parte más teórica de la tesis. Se comienza por estudiar los modelos estacionarios y los no estacionarios en los capítulos 7 y 8 respectivamente, que son herramientas necesarias para la metodología de Box-Jenkins. Los modelos estacionarios se pueden llamar "bien portados" ya que se pueden manejar directamente tal y cómo están. En cambio los no estacionarios deben transformarse a estacionarios primero, para poder ser modelables con la metodología de Box-Jenkins.

En el capítulo 9 se introduce el concepto de pronóstico óptimo, concepto que permite que los pronósticos por medio de la metodología de Box-Jenkins funcione. Se muestra cómo se computan los pronósticos automáticamente.

En el capítulo 10 se explican los cuatro pasos de la metodología de Box-Jenkins:

Identificación, Estimación, Diagnóstico y Pronóstico. Dado que hasta esta parte de la tesis se ha tratado exclusivamente el caso no estacional, en el siguiente capítulo se estudia el caso estacional. Por último en el capítulo 12 se van aplicaciones de la metodología de Box-Jenkins en situaciones reales.

**TESIS CON
FALLA DE ORIGEN**

Capítulo I. Marco de referencia

"Entonces Faraón dijo a José: 'en mi sueño me parecía que estaba a la orilla del río; y que del río subían siete vacas de gruesas carnes y hermosa apariencia, que pacían en el prado. Y que otras siete vacas subían después de ellas, flacas y de muy feo aspecto; tan extenuadas, que no he visto otras semejantes en toda la tierra de Egipto. Y las vacas flacas y feas devoraban a las siete primeras vacas gordas; y éstas entraban a sus entrañas, pero no se conocía que hubiesen entrado, porque la apariencia de las flacas era aún mala, como al principio. Y desperté. Vi también soñando que siete espigas crecían en una misma caña, llenas y hermosas. Y que otras siete espigas menudas, marchitas, a batidas por el viento solano, crecían después de ellas; y las espigas menudas devoraban a las siete espigas hermosas; yo he dicho a los magos, más no hay quien me lo interprete'. Entonces respondió José a Faraón: 'El sueño de Faraón es uno mismo; Dios ha mostrado a Faraón lo que va a hacer. Las siete vacas hermosas siete años son, y las espigas hermosas son siete años; el sueño es uno mismo. También las siete vacas flacas y feas que subían tras ellas son siete años; y las siete espigas menudas y marchitas por el viento solano, siete años serán de hambre. Esto es lo que respondo a Faraón. Lo que Dios va a hacer, lo ha mostrado a Faraón. He aquí que vienen siete años de abundancia en toda la tierra de Egipto. Y tras ellas seguirán siete años de hambre; y toda la abundancia será olvidada en toda la tierra de Egipto, y el hambre consumirá la tierra. Y esa abundancia no se echará a ver, a causa del hambre siguiente, que será gravísima. Y el suceder el sueño a Faraón dos veces significa que la cosa es firme por parte de Dios, y que Dios se apresura a hacerla. Por tanto, provéase ahora Faraón de un varón prudente y sabio, y póngalo sobre la tierra de Egipto. Haga esto Faraón, y ponga gobernadores sobre el país, y quite la tierra de Egipto en los siete años de la abundancia. Y junten toda la provisión de estos buenos años que vienen, y recojan el trigo bajo la mano de Faraón para mantenimiento de las ciudades; y guárdenlo. Y esté aquella provisión en depósito para el país, para los siete años de hambre que habrá en la tierra de Egipto; y el país no padecerá de hambre'."

GENESIS XLI, 17-36

I.1 Necesidad de pronosticar

¿Cómo sería el mundo si todos soñáramos lo que va a suceder?. Sin duda el mundo sería muy diferente y el humano no sería humano. La incertidumbre forma parte de nuestras vidas desde el momento en que nacemos y la única certeza con la que contamos es la de que moriremos algún día. La evolución misma tiene más edad que nosotros y aun ella se enfrenta con la misma incertidumbre ante el futuro.

El conocer cómo van a ser los hechos en el futuro ha intrigado al humano desde que tuvo la capacidad de preocuparse por su porvenir. ¿Cómo planearon los primeros homínidos la estructura de sus casas y hogares? No podían hacer una vivienda demasiado cobijada si vivían en una zona cálida o templada, porque se sofocarían y su estancia sería inhabitable. ¿Y que tan resistente tendría que ser?. ¿sería necesario utilizar materiales más resistentes como para soportar el ataque de un animal de gran tamaño, si no habitaban estos en la cercanía?. ¿dónde debería construir, en un valle, en una planicie o en un monte?.

Al pensar en la decisión que debía tomar, el hombre necesitaba información. Podía ver a su alrededor y percatarse que no existía la posibilidad de que su hogar fuera inundado por el desbordamiento de un río cercano. Podía conocer con esto la situación del presente, pero no le era suficiente. Los que no tuvieron la capacidad de darse cuenta de que esto no era suficiente, no lograron sobrevivir. De manera que nuestros ancestros que sí sobrevivieron nos heredaron la capacidad de planear basándose en la **experiencia o conocimiento del pasado**.

Las culturas que se fueron desarrollando a partir de esta capacidad de planear. Desarrollaron artificios como los calendarios, donde se plasma el hecho de que comprendían perfectamente ciertas cosas que se repetían cada determinado tiempo –año con año por ejemplo– como el clima. A pesar de lo impredecible que era día con día, se percataron que en unas épocas hacía más frío que en otras y que en unas llovía y en otras no. Así comprendieron cuándo cosechar ciertos frutos, aunque estuvieron siempre concientes de la posibilidad de que el clima no se comportaría como ellos esperaban.

Existen principios de pronóstico tan palpables, que siguen vigentes incluso en los métodos de pronóstico más avanzados, como es el de que alguien que tiene mayor cantidad de información tiene así tanto de ventaja sobre el que no la tiene.

Cuando uno de los primeros habitantes racionales de la tierra decía a una de sus hijos que no jugara cerca de un barranco e incluso se lo impedía por la fuerza, estaba haciendo uso de información que bien puede haber sido adquirida por él en el pasado, si es que alguna vez resbaló y sufrió lastimaduras y entendió que podría haber muerto, o quizá vio a alguien resbalar y sucederle lo mismo, o fueron sus progenitores mismos que le enseñaron por la fuerza que le podría suceder algo desagradable si se acercaba demasiado al barranco. Por el contrario, si un infante que no tiene quien le cuide y que le diga lo peligroso que es jugar cerca de un barranco tiene una clara desventaja evolutiva sobre los que cuentan con información adicional.

Afortunadamente, el hombre no ha perdido su interés y curiosidad por el futuro a través de la historia, dado que el futuro no ha dejado de sorprenderlo. Así ha logrado mejorar sus técnicas de pronóstico haciendo uso de cuanto recurso sirva para realizar predicciones exitosas. Al buscar en las ciencias encontró en las matemáticas un aliado sin paralelo, debido a que si volvemos toda la información en números (cuantificamos) podemos aplicarles operaciones matemáticas y pasar de lo abstracto a lo cuantificable, creándonos una oportunidad de poder hacer algo, y con algo de suerte, encontrar un patrón.

Al estar escudriñando en otras ciencias no tan exactas, encontró que el juicio de las personas funciona muchas veces tan bien como los pronósticos realizados matemáticamente. Con esto se percató de que el simple hecho de medir algo y encontrarle un patrón no es garantía de nada si no se estudia y entiende el entorno de lo que se quiere pronosticar. Este entorno podría ser social, económico, financiero, de ciencias exactas, etc. Con lo que surgió la necesidad de hacer uso de expertos, que ven las cosas a nivel **micro** y que nos ayudan a entender las cosas a nivel **macro**.

Aunque la mayoría de las actuales técnicas formales de pronóstico fueron creadas en el siglo diecinueve, la aparición de la computadora ha tenido un gran impacto en las técnicas cuantitativas. Debido a la naturaleza repetitiva de muchos de ellos, se ha vuelto una necesidad utilizar la computadora. Es por esto que métodos como el de Box-Jenkins que en un principio se consideraban no aplicables por ser demasiado iterativos y poco prácticos -a pesar del respaldo teórico con el que contaron desde un principio- no pudieron ser implementadas ni ser utilizados a gran escala hasta que apareció la computadora. A finales del siglo pasado se diseñó y programó una gran cantidad de software específicamente para pronosticar. Sumado esto al auge de las computadoras personales, se ha logrado que los métodos más novedosos estén al alcance de todos.

Se puede concluir que pronosticar es un intento de adivinar eventos por medio del análisis del pasado. Las técnicas de pronóstico se pueden basar en la experiencia y opinión de expertos o en modelos matemáticos que describen el patrón de datos del pasado.

1.2 Métodos de pronósticos

Los procedimientos utilizados se pueden clasificar dependiendo de qué tan lejos hacia el futuro se desea pronosticar. Es decir, dependen del **horizonte de pronóstico**. También se pueden clasificar dependiendo de si el tipo de método necesario para realizar el pronóstico es **cuantitativo** o **cualitativo**.

El horizonte de pronóstico se puede aproximar a tres clasificaciones:

- 1) A corto plazo. Va desde un día hasta un año.

Ejemplo 1: ¿Qué tiempo hará mañana, para saber desde ahora qué ropa debo ponerme?

Ejemplo 2: ¿Estaré de humor para levantarme temprano o amaneceré con más flojera de lo normal?

- 2) A mediano plazo. Va de más de un año a cinco años.

Ejemplo 1: ¿Terminaré con calificación aprobatoria el año escolar que estoy empezando?

Ejemplo 2: ¿Cuál será la situación económica en el país el año que viene?

- 3) A largo plazo. Va de cinco a diez años.

Ejemplo 1: ¿Cuántos hijos tendré dentro de 5 años ?

Ejemplo 2: Quiero comprar un terreno. ¿Qué zona me redituará la mayor plusvalía?

Nótese que para cada par de ejemplos, en uno se puede tener una postura activa con respecto a lo que sucederá y en otro se está a la expectativa del futuro. Se puede establecer como regla general y siendo no muy estrictos que cuanto más corto el horizonte, más certero y eficiente resulta cualquier pronóstico.

Todas las metodologías para pronosticar se pueden clasificar en dos grandes categorías: **métodos cuantitativos** y **métodos cualitativos**.

1.2.1 Métodos cuantitativos

Los métodos cuantitativos utilizan el análisis estadístico de datos históricos para poder identificar un patrón en los datos, encontrando una fórmula o modelo matemático que genere el patrón y con éste generar las predicciones. Los modelos cuantitativos toman por sentado el hecho que el patrón identificado se va a comportar igual tanto en el pasado como en el futuro. Los métodos cuantitativos se pueden clasificar en dos tipos: los métodos **de series de tiempo** y los métodos **explicativos**.

Existen modelos de series de tiempo univariados o multivariados. En este trabajo solamente trataremos con el caso univariado. Tómese en cuenta que no se puede proceder a estudiar el caso de más de una variable sin haber comprendido plenamente el caso de una sola y por tanto es con éste último caso con el que se necesita formar una base teórica sólida. En adelante cuando hablemos de "series de tiempo" nos estaremos refiriendo siempre al caso univariado.

Los modelos de **series de tiempo** se basan en el análisis de la secuencia cronológica de observaciones de una variable en específico. Las observaciones se pueden hacer

anualmente, mensualmente, semanalmente, diariamente, etc. El análisis de series de tiempo se basa en la suposición de que augurar los valores que va a tomar cierta variable (evento) en el futuro es posible por medio del análisis del desempeño de esa variable a través del tiempo.

Los métodos **explicativos** trabajan de manera diferente. Suponen que la variable que se desea pronosticar puede ser "explicada" con otras variables, que son las variables independientes. El objetivo de estos métodos es encontrar la fórmula matemática que represente la manera en que las variables independientes afectan a la variable a pronosticar, que es la variable dependiente. Un ejemplo es el de las ecuaciones diferenciales y los pronósticos que se hacen la física con éstos.

Otro ejemplo más práctico son las ventas de un cierto producto. Éstas pueden ser explicadas a través de la cantidad de dinero que se invierte en publicidad, la liquidez de los clientes, el precio del producto, la cantidad de competencia y otros factores.

Existen grandes diferencias que aportan ventajas a cada tipo de método sobre el otro. Los métodos de series tienen un enfoque diferente con respecto al proceso o sistema bajo análisis, porque ven al proceso como una caja negra, donde las operaciones que suceden dentro de ella no le interesan, le importa sólo el resultado que arroja. En cambio los métodos explicativos buscan encontrar todas las operaciones que se realizan entre las variables, sin antes encontrar cuáles son esas variables, y así comprender qué es lo que sucede en el proceso. Buscan ser abarcadores. lo que en teoría les da la ventaja de estar listos para pronosticar un cambio radical de comportamiento. Aunque su aparente robustez les puede resultar contraproducente debido a que hay muchos procesos que no pueden ser asimilados por completo, debido a que su complejidad es tal que se vuelven sistemas, no "muy complicados", sino complejos. Más adelante se verán más a fondo estos, pero por el momento bástenos saber que en dichos sistemas las variables ¡aparecen y desaparecen!.

Por otro lado los modelos de series de tiempo son más simplificadores, lo que les da el potencial de pronosticar a través de lo que *en resumen* ha hecho un proceso. Se preocupa sólo de las salidas y no indaga el por qué de éstas. Esto les imposibilita prever un cambio sin precedente en el proceso.

Las ventajas de utilizar métodos explicativos es que uno entiende el porqué de las salidas del proceso y logra hacer un mapa conceptual de los procesos que afectan al que tenemos bajo estudio. Pero no siempre es necesario comprender el porqué y el cómo de las salidas de nuestra caja negra. Si podemos encontrar patrones en las salidas por sí solas, como lo hacen las series de tiempo, nos estaríamos ahorrando una gran cantidad de operaciones, cálculos y almacenamiento de datos. Ya que por el método de las series necesitamos únicamente almacenar los datos de una sola variable y para los explicativos se almacenan todas las variables del modelo o caja negra.

Por lo tanto, en series de tiempo se puede establecer una función como la generadora de las salidas, cuyas variables son precisamente las mismas salidas del proceso, pero originadas en el pasado.

Así se puede escribir que

$$P = F (S_1, S_2, S_3, \dots)$$

Pronóstico = P

Función = F

Salidas o Resultados del proceso en el pasado = S_n

donde la n, cuanto mayor sea, indica que se está refiriendo a una salida más lejana hacia atrás en el tiempo. Se podría decir entonces que de cierta forma los métodos de series de tiempo trabajan un poco como explicativos, donde se tiene una variable diferente para cada tiempo diferente t de la variable única de salida.

En cuanto a las diferencias en el horizonte de pronóstico para ambos tipos de métodos cuantitativos, se resumen en la siguiente tabla:

MÉTODOS DE PRONÓSTICO Y SUS APLICACIONES

| Aplicación de los pronósticos | Horizonte de tiempo | Exactitud requerida | Número de productos | Nivel administrativo | Método de pronóstico |
|-------------------------------|---------------------|---------------------|---------------------|----------------------|---------------------------------|
| Diseño de procesos | Largo plazo | Mediana | Uno o pocos | Alto | Explicativos y cualitativos |
| Planeación agregada | Mediano plazo | Alta | Pocos | Medio | Explicativos y Series de tiempo |
| Programación de actividades | Corto Plazo | Superior | Muchos | Bajo | Series de tiempo |
| Administración de inventarios | Corto Plazo | Superior | Muchos | Bajo | Series de tiempo |

1.2.2 Métodos cualitativos

Los métodos cualitativos de pronóstico son los que se utilizan cuando se tienen pocos datos registrados acerca del evento que se quiere pronosticar, cuando la información que se tiene no se puede cuantificar o cuando el evento es susceptible de ser afectado por cambios tecnológicos. Los métodos cualitativos pueden ser **exploratorios** o **normativos**.

Los **métodos exploratorios** trabajan de atrás hacia adelante. Utilizan la información del pasado y presente para crear pronósticos de cuándo y qué es lo más factible que suceda. Se analizan exhaustivamente todas las situaciones posibles. Dicen: "Esto va a suceder según nuestro pronóstico, entonces ya que sabemos eso, debemos de sacar el mayor provecho de nuestra información".

Por otro lado, los **métodos normativos** trabajan cronológicamente de adelante hacia atrás. Establecen primero las metas y objetivos, y después trabajan hacia el presente, para ver si es posible lograr las metas establecidas, estableciendo así sub-metas. Analiza la tecnología y los recursos actuales y establece cuáles son las limitaciones y restricciones que se deben superar o eliminar. Su filosofía dice: "Hay que lograr hacer esto (objetivo), entonces hay muchas cosas que debemos cambiar en el presente para poder lograr eso que hemos planeado".

La principal ventaja de la pronosticación cualitativa es que uno tiene a la mano una gran cantidad de información, tanto cuantificable como incuantificable para hacer el pronóstico. Las desventajas son que, en primera, no existe una manera sistemática de

mejorar la exactitud del pronóstico; y en segunda, los pronósticos pueden ser parciales o sesgados.

En la industria es frecuente valerse de los pronósticos cualitativos para averiguar qué zonas de la tecnología y la ingeniería quedarían afectadas por un avance científico; es decir, se emplean para conseguir que la sucesión constituida por descubrimiento, puesta a punto, realización ingenieril avanzada y aplicación se lleve a cabo del modo más eficiente posible.

Para cualquier evento, los métodos cualitativos son los más apropiados para hacer pronósticos a *largo* plazo ya que son capaces de tratar con lo inefable o incuantificable, y ser más "inteligentes" y abarcadores que los cuantitativos.

DIFERENCIAS ENTRE MÉTODOS CUALITATIVOS Y CUANTITATIVOS

Los métodos cuantitativos tienen ventajas sobre los cualitativos, como es el de ser objetivos, ya que una vez que se forma el modelo y se escogen las variables independientes no es necesario "pensar" ni "preocuparse" cada vez que se utilice el modelo para hacer un nuevo pronóstico. Asimismo, con los métodos cuantitativos sí es posible medir la exactitud de las premoniciones y, por tanto, su eficiencia. Aunado a esto los cuantitativos, además de pronósticos puntuales (de un solo valor), pueden generar pronósticos de intervalo (un rango de valores sujeto a un cierto grado de confianza).

Una de las desventajas de los métodos cuantitativos es que son incapaces de pronosticar a largo plazo: en estos casos se deben utilizar los cualitativos. La otra gran desventaja es que es muy difícil que los métodos de pronóstico cuantitativos pronostiquen cambios drásticos de comportamiento, sobre todo si nunca se habían presentado.

Por ejemplo supongamos que habíamos realizado un modelo para pronosticar el precio por galón de la gasolina sin plomo para el período de 1980 a 1989. Aunque hasta antes de 1990 el modelo hubiera funcionado excelentemente, para ese año el modelo resultaría nefasto, dado que en el momento de hacer el pronóstico no se podía anticipar la invasión iraquí a Kuwait y su efecto inmediato en el precio de gasolina sin plomo.

Para poder hacer pronósticos eficaces la gran mayoría de las veces resulta necesario mezclar ambos tipos de metodologías, tanto la cuantitativa como la cualitativa, porque nada permanece exactamente igual a través del tiempo. Así que la mayoría de las veces los métodos cuantitativos son medidos y amoldados con las opiniones de expertos en el área.

La presente tesis se concentra en los métodos de series de tiempo: se estudiarán los métodos de suavizamiento, los de descomposición y la metodología Box-Jenkins, así como métodos cualitativos, tanto exploratorios como normativos. Los cuantitativos-explicativos no se estudiarán en la tesis.

Se pretende que todos los conceptos que se manejen en la tesis sean conocidos para el lector. Si aparece uno que no se ha manejado anteriormente tenga el lector la certeza de que se explicará más tarde.

1.3 Series de tiempo

Una serie de tiempo es un conjunto de observaciones de un evento realizado de manera secuencial a través del tiempo con períodos equidistantes entre observaciones. Las series de tiempo pueden ser tanto discretas como continuas. Con las que nos encontraremos

en la vida real la inmensa mayoría de las veces son las primeras, por lo cual son las que estudiaremos en este texto. Lo ideal en una serie de tiempo es que la distancia entre observaciones sea homogénea, que las observaciones estén ordenadas cronológicamente y que la serie no tenga observaciones faltantes. Es común encontrar series con las carencias de alguno de estos requisitos.

OBSERVACIONES FALTANTES

Si existen datos faltantes en la serie se deben reponer de alguna manera lógica. Podemos reponer por ejemplo con la media de toda la serie o con la media de los datos adyacentes. También se puede interpolar, que no es más que pronosticar hacia dentro de la serie de tiempo con regresión lineal.

DATOS FUERA DE LO COMÚN O DATOS ATÍPICOS

Es posible que en la gráfica de la serie aparezcan datos extraños o atípicos. Esto se puede deber a dos razones: o alguien se equivocó al capturar los datos o sucedió algo fuera de lo común en el proceso observado para esa etapa del tiempo. Puede existir una explicación perfectamente racional al respecto, como es el que para ese periodo (mes, semana, día, etc.) haya ocurrido algo excepcional, como podría ser un simulacro de incendio. Si pensamos en el caso en que la serie representa los datos de la producción diaria de cierta compañía, es fácil concluir que si en ese día la producción se presenta excepcionalmente baja, no querrá decir que la producción del día siguiente, o de cualquier día en el futuro, se vea afectada por éste hecho. A estos datos se les llama datos **discrepantes**. Pero también existen casos en los que una observación excepcional sí tiene una repercusión de mayor trascendencia. Pensemos en que la misma compañía cuenta con más de una fábrica y que una de éstas se incendia y se pierde la fábrica por completo. En la serie de "Producción Diaria" va a existir un recortamiento súbito que va a "jalar" todos los datos, afectando por completo la perspectiva gráfica de la serie. A este tipo de datos (el día del incendio) se les llama datos **influyentes**.

En el primer caso los datos se deben tratar como en el caso de los datos ausentes, ya que la información que nos presentan es defectuosa. Pero en el caso de los datos influyentes no hay nada que podamos hacer y se debe trabajar tal y como está.

LONGITUD DE LA SERIE DE TIEMPO

Como regla, la serie debe contener suficientes datos para una estimación correcta de los parámetros. No existe una forma clara de establecer el mínimo de observaciones. Algunos autores dicen que con treinta es suficiente, otros que se necesitan cincuenta y otros dicen que son necesarios por lo menos sesenta. Si la serie contiene ciclos entonces se debe abarcar suficiente cantidad de estos ciclos para que sea posible capturar su efecto. Si contiene estacionalidad, lo mismo aplica. Así que las series que contienen estacionalidad necesitan más datos que las que no la contienen. Un pocas palabras, entre más datos, mejor.

REPRESENTATIVIDAD DE LA SERIE

Los datos en la serie deben personificar al máximo el proceso bajo estudio. Esto quiere decir que, viendo la serie como una muestra, los parámetros muestrales se deben acercar a los parámetros poblacionales conforme la muestra sea mayor, una propiedad que en estadística se llama **ergodicidad**.

NOTACIÓN

Todas las series de tiempo que estaremos estudiando son discretas. Utilizaremos una notación de subíndice, de operador rezago y del operador diferencia. Cuando es necesario indicar la posición cronológica o temporal en la serie es necesario utilizar un subíndice. Si una variable en la serie de tiempo está indicada y_t , entonces el subíndice t indica la ubicación temporal de ese elemento en la serie. Si t toma valores $1, 2, 3, \dots, T$, esta serie puede ser representada por y_1 hasta y_T .

OPERADOR REZAGO

El operador rezago **B** (*backshift operator* en inglés) que opera sobre una variable en el tiempo t está haciendo referencia al valor de la misma variable pero en el momento $t - 1$. Por lo tanto $B y_t = y_{t-1}$. De manera similar $B^2 y_t = y_{t-2}$. Lo que hace el operador rezago es retroceder a la variable un período hacia atrás en el tiempo, es decir hacia el pasado. El álgebra de este operador es similar al del exponencial. Por ejemplo, $B^n B_m(y_t) = B_{n+m}(y_t) = y_{t-n-m}$.

Las potencias del operador rezago se traducen a períodos de rezago: $B^6 = Y_{t-6}$. $B(B Y_t) = B^2 Y_t = Y_{t-2}$. También existen las inversas para el operador: $BB^{-1} = 1$.

De manera análoga $z/(1-B)^{-1} = z/(1-b)$. Resulta interesante observar que la inversa de un rezago puede resultar en una serie de diferencias infinitas, es decir $1/(1-B) = (1-B)^{-1} = (1 + B + B^1 + B^2 + B^3 + \dots + B^{n-1} + B^n + \dots)$.

EL OPERADOR DIFERENCIA

El operador diferencia es representado por Δ . La primera diferencia de y_t está dada por la expresión $w_t = \Delta y_t = y_t - y_{t-1}$. Otra manera de representar esta operación es $w_t = \Delta y_t = (1-B)y_t$. La segunda diferencia es la primera diferencia de la primera diferencia: $\Delta^2 y_t = \Delta(\Delta y_t) = (1-B)(1-B)y_t = (1-2B + B^2) y_t = (y_t - 2y_{t-1} + y_{t-2})$.

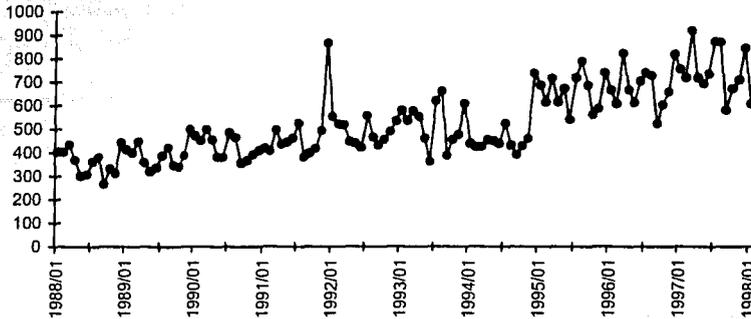
Una serie de tiempo se puede diferenciar y se obtendrá otra serie de tiempo que se dice *diferenciada* de la serie original. La serie diferenciada se obtiene al aplicar el operador diferencia a toda la serie original, dato por dato, empezando por la resta del segundo datos menos el primero, hasta el último dato menos el penúltimo, de tal manera que la serie diferenciada una vez contendrá un dato menos. Una serie diferenciada dos veces, contendrá dos datos menos que la serie original y así sucesivamente.

PRONOSTICAR CON SERIES DE TIEMPO

Una propiedad inherente en las series de tiempo es que las observaciones adyacentes tiene una cierta dependencia. La manera en que se pretende pronosticar es analizando esa relación de dependencia a través del análisis de series de tiempo.

El análisis de series de tiempo utiliza en el tiempo t las observaciones disponibles en ese momento para obtener el valor para el momento $t+h$. La observaciones pueden ser, por ejemplo $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-m}$, donde m es el tamaño de la serie. Al hacer un pronóstico se denota por \hat{y}_t o y_{t+h} . La segunda notación es más explícita en el sentido que está indicando para cuantos períodos en el futuro está pronosticando. El primero es idéntico al de las observaciones y_t con la única diferencia de que tiene un *gorrito* que quiere decir que es un valor estimado. Así los valores observados son los sencillos y_t o también y_{t-h} donde h es la cantidad de períodos que nos movemos hacia el pasado.

A continuación se presenta una serie de tiempo real. Se trata de las llegadas internacionales por mes a México vía aérea de Enero de 1988 a Enero de 1998 en miles de pasajeros.



TESIS CON FALLA DE ORIGEN

Fuente: INEGI, A.S.A. Gerencia de Planeación e Investigación

Vayamos definiendo cada una de las características que podemos identificar en esta serie, de manera que reconozcamos cuáles nos sirven para pronosticar.

Primero, los datos son mensuales. Es decir, si tomamos cualquier dato en la serie y avanzamos 12 datos para adelante o para atrás, sabremos que estaremos avanzando hacia un año posterior o anterior.

También podríamos avanzar 6 datos para adelante o 6 datos para atrás. En este caso no nos estaríamos saliendo del rango de un año necesariamente, bien podríamos movernos dos veces consecutivas de esta manera sin salirnos del rango de un año. Si encontramos algún patrón dentro del período de un año se llama patrón **estacional**. Si se repite un cierto comportamiento fuera del rango de un año, se dice que tiene un patrón **cíclico**.

Situémonos en el mes de Diciembre del año 1991. Ahora realicemos nuestro salto de 6 lugares hacia adelante. Repitamos esto 5 veces. ¿Qué se observa?. En los datos que hemos estado cayendo al realizar los saltos siguen una cierta pauta o patrón que los hace diferentes a los demás y hace que resalten. Que suceda esto no nos debe asombrar: recordemos que estamos hablando de turismo y suena lógico que éste aumente en periodos de vacaciones, como son las de navidad y verano (diciembre y julio), y que disminuya y se mantenga más o menos estable en los demás meses cuando no hay vacaciones largas.

TIPOS DE DATOS

Bien, podríamos pensar ¿pero qué hay de los fines de semana y puentes vacacionales?. Ahí también hay comportamientos que se repiten. Desgraciadamente nuestra serie de tiempo está tratando solamente con datos mensuales, por lo que es incapaz de mostrarnos como se comporta el interior de cada uno de los meses. Aunque nos está fallando en presentarnos más detalle, está mostrándonos una visión más global. Éste es un punto crucial al estar realizando un pronóstico. Debemos asegurarnos que el tipo de datos que estemos usando sea el adecuado para nuestro fin, ya que si estamos realizando el pronóstico para una empresa, ésta podría por ejemplo demandarnos que le indiquemos cuánto turismo se va a estar presentando en Semana Santa del próximo año, de manera que

pueda realizar los ajustes necesarios. Aunque podríamos obtener el pronóstico para abril, seríamos incapaces de indicar precisamente cuánta gente esperaríamos para esa semana en específico, con lo cual estaríamos fallando en cumplir nuestro objetivo.

ESTACIONALIDAD O VARIACIÓN ESTACIONAL

Ya vimos que los valores aparecen similares cada seis meses a lo largo de toda la serie. Luego entonces, esperaríamos que en el futuro los datos se sigan comportando de manera semejante. Esta periodicidad que descubrimos tan sencilla de identificar es de gran valor para armar nuestro pronóstico.

Este comportamiento es muy común en las series de tiempo y se llama **estacionalidad** o **variación estacional** y se dice que la serie de tiempo es una serie estacional.

CICLICIDAD O CICLO

Si ya tratamos con los patrones de comportamiento dentro de un año, ahora vayamos a los mayores a un año. Si realizamos el mismo experimento que hicimos con la estacionalidad, podríamos preguntarnos: ¿qué se repite cada 13 meses, cada 14, cada 15, cada 20, que afecte al turismo de alguna forma?.

Situémonos en el contexto político, por ejemplo. Lo hemos vivido la mayoría de los mexicanos, y lo sabemos de hecho, que cada 6 años con el cambio de presidente comienzan una serie de interrogantes en cuanto a quién va a conformar su gabinete, cómo se va a manejar el presupuesto, que si éste presidente sí va a cumplir sus promesas de campaña, etc. Esta situación de incertidumbre nos afecta a todos nosotros, incluyendo a la economía. Hemos visto la moneda devaluarse de la noche a la mañana, hemos visto guerrillas levantarse en armas, gente asesinada, manifestaciones en las calles, todo girando alrededor de un simple hecho: el comienzo de un nuevo sexenio.

Considerando entonces el contexto económico, podríamos deducir que nuestra serie de turismo también se vería afectada por este suceso de alguna manera, como podría ser que la gente pierda su trabajo y por tanto su poder adquisitivo, con lo que tendría que aparecer una disminución en nuestra serie alrededor de ésta época. También el país pierde su atractivo a los extranjeros cuando empobrece el país porque hay más delincuencia, lo cual conllevaría otra disminución.

Recordemos que casi todos los presidentes tratan de dejar el peso lo mejor posible, así tengan que inflarlo desmesuradamente, por lo que el último año del sexenio será cuando la gente pueda viajar más. Todo esto, cada 6 años.

Una nota muy importante con respecto a la ciclicidad: debido a que la ciclicidad es difícil de identificar, no siempre será posible modelarla.

Pero ¿se nos podría estar escapando algo todavía, además de estacionalidad y ciclicidad?. Sí. Se trata de esa inclinación en la serie, la **tendencia**: ese firme acrecentamiento en las cantidades de los datos. Es fácil entender de donde viene este constante aumento.

Gracias a la medicina, la esperanza de vida de los mexicanos ha aumentado, como lo ha hecho con muchos otros países. También el grado de natalidad de los mexicanos es

uno de los más altos del mundo. Todo esto conlleva a que cada vez más gente habite este país y por ende que cada vez exista más gente viajando.

Cabe mencionar que los dos tipos de tendencia más comunes son las lineales, exponenciales y con forma de S.

Se tienen así en resumen tres propiedades identificadas en total con las que podemos tratar de predecir el comportamiento de un cierto proceso. Ya hemos dado con todas las propiedades que se utilizan en el análisis de series de tiempo. Bueno, con casi todas. De hecho nos falta una que hemos dejado para el final adrede, por su importancia tan grande en la comprensión de procesos reales.

¿Que sucedió en septiembre de 1985? Un terremoto que mató a 30,000 personas. Esto obviamente afectó el turismo en más de una manera. ¿Cree el estudiante que algún pronosticador hubiera podido predecir este suceso? Es obvio que no.

También los huracanes que aterrizan en las costas súbitamente no son pronosticables sino ya cuando están muy cerca y es demasiado tarde. Las tormentas tropicales que inundan comunidades enteras y que hacen que se desgajen cerros cobrando con ello muchísimas vidas son también, para desfortuna de muchos mexicanos, completamente imposibles de pronosticar. Si el futuro es una fuente de incertidumbre, ¿no hay manera de prepararse para ello? ¿no deberíamos incluir un componente que precisamente prepare al modelo para el futuro?

La forma en que se prepara el modelo para el futuro es con un componente más llamado **ruido** o **componente aleatorio**. Existe incluso una forma de medirlo e incluso generarlo. Se puede medir en el sentido de que sabemos cómo se comporta, cómo se ve, ya que en verdad es muy fácil identificarlo dada su principal característica, que es la de ser completamente impronosticable. Podemos aislar este componente de la serie de tiempo y saber cómo se ha comportado en el pasado y conocer su valor actual, pero ningún método de pronóstico sabrá cual es su siguiente valor. No tiene patrón alguno que podamos encontrar, dado que es un proceso completamente aleatorio.

Y es por eso que sabemos como identificarlo: cuando nuestros métodos localizan la tendencia, la ciclicidad y la estacionalidad y las conjuntan para crear el pronóstico y aún así no logran acertar completamente en sus pronósticos, ese error cometido se llama ruido.

FUENTES DE INCERTIDUMBRE EN LA FORMULACIÓN DE MODELOS

Para comprenderla enteramente tendremos que considerar tres hechos muy importantes que forman la base de la creación de cualquier modelo de pronóstico a través de series de tiempo de un proceso real.

La serie de tiempo es una **muestra** de un proceso. Éste es un gran factor de incertidumbre porque sólo estamos hablando de una pequeña parte observada del proceso y no de su totalidad que sería una muestra de tamaño infinito. Tener una muestra infinita (que no llamamos población por que el tiempo nunca se detiene) sería ideal y haría que esta incertidumbre desapareciera. Ahí va nuestra primer fuente de aleatoriedad.

Si fuera posible regresar en el tiempo y repetir el proceso bajo exactamente las mismas condiciones, tendría un comportamiento completamente diferente. Es la segunda

fuente de incertidumbre, a la que podemos llamar de **unicidad** del proceso, ya que el proceso tomó uno de tantos caminos que pudo haber tomado.

Por último, el tercer factor que origina incertidumbre: **el futuro** por sí solo contiene bastante incertidumbre. Es decir, imaginemos que ya tenemos un modelo formado basándose en datos infinitos e imaginemos que hemos viajado en el pasado y que de hecho el modelo sí se ha repetido en la realidad. Es decir, somos "dueños del proceso" y sabemos exactamente cómo se comporta. Aún así, la muestra se termina en el presente y el futuro traerá nuevas observaciones que se añadirán a la infinita muestra, por lo que en cierto sentido no sabemos absolutamente nada acerca de estos nuevos datos.

Así es como se suman las tres importantes fuentes de incertidumbre que hacen del ruido una parte muy importante a considerar en nuestros modelos. Esta aleatoriedad, afortunadamente, se puede medir o ponderar para ser considerada en el modelo. Aunque por construcción este factor de aleatoriedad o incertidumbre no se puede pronosticar, sí se puede delimitar*.

1.4 Aplicaciones de los pronósticos

Los pronósticos, como ya vimos, se obtiene con el fin de poder tomar decisiones. Los campos en los que se aplican son superabundantes. A continuación se presentan algunas de las áreas en donde se utilizan y una breve descripción de las decisiones que se apoyan en ellos.

1. Ventas y producción.- A las empresas les interesa conocer los prospectos de cierto producto, para saber si producir más o producir menos. Cuando se tiene un producto en inventario que no está siendo vendido, resulta en una pérdida para el fabricante. Lo mismo sucede si producimos de menos, ya que al no tener el producto a la mano para su venta se está perdiendo no sólo la ganancia de ese producto, sino incluso el renombre de la compañía. De igual forma si los dueños de la compañía conocen cuanto van a estar vendiendo los próximos meses, les da la oportunidad de planear las contrataciones de personal y la cantidad de turnos que serán necesarios.

2. Mercadotecnia. Si se conocen los gustos de la gente se puede decidir cuáles productos necesitan más publicidad, o qué precio se le debe fijar al producto.

3. Economía. Para el gobierno y las empresas es de vital importancia conocer el comportamiento que tendrá la economía en el futuro. Al planearse el gasto público para el siguiente año se necesita tener suficiente información y pronósticos que resulten eficaces, porque un error puede tener consecuencias catastróficas.

4. Demografía. El pronóstico demográfico es muy útil en muchas áreas. A un gobierno le interesa éste pronóstico para planear gastos gubernamentales en asistencia y seguridad social, entre otros.

5. Creación de carteras de inversión. Al conocer la solvencia del cliente y su probabilidad de crisis financiera para cierto período en el futuro, se pueden tomar decisiones acerca de a quién hacer préstamos y a quién no y en qué momento. Asimismo para poder decidir en que bolsa de valores se debe invertir, se deben conocer las proyecciones de ese mercado.

* Ver Anexo 3.

Capítulo II. Pronósticos de series de tiempo por suavizamiento

Una vez que tenemos una idea de lo que es una serie de tiempo y que estamos listos para comenzar a pronosticar, el paso siguiente es precisamente:

¿Cómo pronosticar?

Si un proceso cuenta con una historia registrada de su conducta, es decir un conjunto de resultados observados, es posible que estos cuenten con ciertos patrones en su comportamiento que podrían ayudar a divisar como podría actuar en un futuro.

La cuestión será entonces si somos capaces de identificar estos patrones, en el caso que en realidad existan, con la ayuda de alguna herramienta matemática. Ésta es la función principal del análisis de series de tiempo: predecir el futuro a través de la información del pasado.

Una vez que tenemos nuestra serie de tiempo, podemos comenzar.

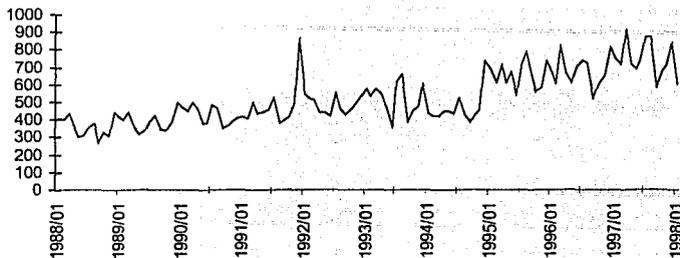
Lo primero que podemos hacer con nuestra serie de tiempo es observarla. En muchos sentidos el ojo humano es por mucho más eficaz en encontrar patrones de comportamiento que cualquier herramienta matemática. Estos patrones podrían ayudarnos a formular un modelo. Sin embargo también el ojo en ocasiones es fácil de engañar, pero es lo mejor para empezar.

Tomemos la serie de tiempo del turismo como ejemplo y tomemos la siguiente tabla con los datos de la serie:

| Mes \ Año | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 |
|------------|------|------|------|------|------|------|------|------|------|------|
| Enero | 405 | 413 | 471 | 419 | 552 | 579 | 439 | 687 | 666 | 756 |
| Febrero | 404 | 400 | 450 | 407 | 521 | 535 | 425 | 614 | 607 | 719 |
| Marzo | 434 | 447 | 499 | 497 | 517 | 577 | 424 | 717 | 823 | 919 |
| Abril | 370 | 359 | 455 | 435 | 447 | 552 | 454 | 614 | 667 | 719 |
| Mayo | 301 | 319 | 382 | 443 | 441 | 462 | 450 | 673 | 613 | 693 |
| Junio | 308 | 336 | 380 | 461 | 420 | 362 | 439 | 541 | 704 | 736 |
| Julio | 362 | 387 | 486 | 525 | 556 | 620 | 525 | 718 | 742 | 873 |
| Agosto | 381 | 422 | 464 | 380 | 466 | 660 | 432 | 788 | 728 | 872 |
| Septiembre | 265 | 347 | 354 | 398 | 432 | 388 | 390 | 687 | 523 | 581 |
| Octubre | 333 | 339 | 367 | 418 | 456 | 454 | 427 | 562 | 603 | 673 |
| Noviembre | 311 | 388 | 390 | 493 | 490 | 476 | 459 | 587 | 658 | 709 |
| Diciembre | 444 | 499 | 409 | 865 | 534 | 608 | 738 | 739 | 819 | 845 |

Esta lista de datos carece de utilidad visualmente hablando ya que es más fácil ver la serie que imaginarnos el patrón de los datos a través del tiempo.

TESIS CON
FALLA DE ORIGEN



¿Qué se ve en la gráfica de la serie? en realidad lo único relevante que alcanzamos a percibir en esta gráfica de la serie de tiempo es una línea temblorosa con ciertos sube y bajas. Nos pudiera recordar a las gráficas de sonido o ruido auditivo, de las gráficas de los electrocardiogramas, etc. Aunque parezca extraño, en realidad todas estas gráficas están muy relacionadas y tienen bastante en común con las series de tiempo.

Si no supiéramos que el eje de las ordenadas representa el tiempo, en realidad nos podríamos imaginar muchas cosas. Las gráficas a las que estamos acostumbrados a ver en cálculo o en geometría analítica, por ejemplo, no tienen esa constante vibración como de shock eléctrico. Pero sabiendo que las abscisas corresponden al tiempo, sabemos que el pronóstico queda a la derecha, el futuro.

Trátase de completar 12 datos más, o sea un año adicional, a la serie de tiempo. No pensará seriamente en incluir los shocks en su dibujo. Uno podría hacerlo, simplemente dibujando con una mano temblorosa, pero no sería lo más sensato, ya que en verdad no sabríamos si en realidad estamos imitando las vibraciones al 100%. Además, como no desvía demasiado lo que pareciera ser el rumbo verdadero de la serie, sería más lógico dibujar solamente este rumbo "medio", sin la temblorina. Algo así como el valor esperado que estudiamos en estadística: sabemos que no estamos dibujándolo como debería de ser al 100% porque no podemos emular la parte "electrificante", pero estamos haciendo nuestro mejor esfuerzo por equivocarnos lo menos posible.

Todos los métodos de pronóstico por medio de series de tiempo se basan en éste sencillo principio. Ya con estos antecedentes podremos proceder a conocer nuestro primer método de pronóstico: el suavizamiento.

Cabe mencionar que no es el único método ni necesariamente el más eficiente, pero tiene ventajas que ninguno de los otros tiene como lo es su simplicidad, por eso será el primero que estudiaremos.

El suavizamiento se utiliza en líneas de producción donde se necesitan pronósticos a muy corto plazo para una cantidad muy grande de productos, donde resultaría poco rentable utilizar métodos muy sofisticados que nos proporcionaran alta precisión o que pronosticaran a un plazo más largo; necesitamos algo bueno y rápido.

Los resultados que han presentado de su aplicación a procesos reales a lo largo de la historia han hecho de los pronósticos por suavizamiento una herramienta altamente utilizada, debido a un excelente nivel de eficiencia y sencillez.

Los modelos de suavizamiento tienen su origen en los modelos de promedios y éstos a la vez en los modelos no formales, por lo que empezaremos por estos últimos hasta llegar a los de suavizamiento.

MODELOS NO FORMALES

Los modelos no formales son todos aquellos que podemos crear nosotros mismos sin la necesidad de hacer uso de fórmulas muy complicadas. El más sencillo es el llamado *pronóstico ingenuo*:

$$y_{t+1} = y_t$$

El pronóstico es realizado en el período t para una unidad de tiempo adicional, es decir, a una **etapa**. Un pronóstico para h etapas es el que se hace para h unidades de tiempo en el futuro o para un horizonte h . El ejemplo más sencillo es:

$$\hat{y}_{t+h} = y_t$$

Tomemos por ejemplo el tiempo $t=1$ donde $y_1 = 405$ que sería el primer valor de la serie. Nuestro pronóstico entonces para el día 2 sería $\hat{y}_2 = y_1 = 405$, o sea el valor observado en el primer mes. Si ahora pasamos a $t = 2$, que será nuestro nuevo presente, el pronóstico será $\hat{y}_3 = 404$ que es el valor observado en y_2 .

¿Qué hacemos pronosticando valores que ya conocemos?. Se llama pronosticar dentro de la muestra y es muy útil para verificar la eficiencia de un modelo de pronóstico. Por supuesto que pronosticar fuera de la muestra es el objetivo. Pero prosigamos así por ahora.

Cabe hacer notar que aunque en el momento $t=2$ ya contábamos con dos valores que podríamos haber utilizado para pronosticar (y_1 y y_2) sólo usamos uno, ya que así nos lo indica el modelo.

Es necesario recalcar que cuando realicemos la predicción en el momento y_{100} , o sea al predecir para el momento y_{101} , sólo estaremos utilizando la centésima parte de la información. Lo que quiere decir que estamos dejando de utilizar 99 datos que nos podrían servir de mucho para entender mejor la serie

Resulta fácil comprender como "piensa" este modelo: es un modelo de mucha fe y poca experiencia. Presupone que el proceso ha llegado a estabilizarse en el valor que se encuentra y que nunca va a cambiar, o por lo menos que no va a cambiar mucho. De manera que considera que si el proceso se comportó de una manera el día de hoy, el día de mañana se comportará de la misma manera.

Otro modelo que podemos formar es el de esperar que el día de mañana tenga el mismo comportamiento que el que tuvo ese mismo día la semana pasada. Por ejemplo, si hoy es jueves nuestro pronóstico sería el valor que tuvo la serie el jueves pasado.

Formalmente:

$$\hat{y}_{t+1} = y_{t-6}$$

o lo que es lo mismo

$$y_{t+7} = y_t$$

“Predigo que el próximo lunes se comportará exactamente como lo hizo el día de hoy”, lunes. No cuesta mucho trabajo darse cuenta de la potencial ineficiencia de estos pronósticos ya que desperdician una considerable cantidad de información por lo que siempre tienen una reacción retardada. Reacciona demasiado tarde.

Por ello debemos preocuparnos por darle más información a nuestros modelos, es decir, si los hacemos que crezcan en su memoria y que hagan un uso mayor del pasado nos percataremos que es necesario sistematizarlos, aunque sea un poquito, para que mejoren en su precisión de manera que puedan prever los cambios. Es así como transitamos al método de **promedios móviles**.

PROMEDIOS MÓVILES

Los modelos con promedios simples no hacen más que utilizar los últimos “n” valores de la serie, promediarlos por medio de la media aritmética y pronosticar. Esta “n” es un número entero que permanece fijo a través de todo el procedimiento.

$$\hat{y}_{i+1} = \frac{1}{n} \sum_{i=0}^n y_{i-1}$$

En otras palabras: primero toma las últimas n observaciones, las promedia y crea su pronóstico a una etapa. Una vez que esta etapa “sucede”, recoge el valor observado y la añade al modelo, eliminando el dato más primitivo. De esta manera se vuelve a quedar con n valores.

La fórmula para el pronóstico a una etapa queda:

$$\hat{y}_{i+1} = \frac{y_i + y_{i-1} + y_{i-2} + \dots + y_{i-n+1}}{n}$$

y el pronóstico para el período t está dado por la ecuación:

$$\hat{y}_t = \frac{y_{t-1} + y_{t-2} + y_{t-3} + \dots + y_{t-n}}{n}$$

entonces podremos reformular nuestro pronóstico a una etapa como

$$\hat{y}_{i+1} = \frac{y_i}{n} + \hat{y}_i - \frac{y_{i-n}}{n}$$

Obviamente cuanto mayor sea “n” mayor será el suavizamiento de la serie, por que se reaccionará poco con cada nuevo valor. Estará confiando más en la información con la que cuenta y considerará el nuevo valor como un valor al que sí hay que tomar en cuenta para las siguientes premoniciones, pero no demasiado.

Sucede lo contrario si tomamos una n pequeña: querrá decir que el modelo no confía tanto en su propia información y por ende tenderá a reaccionar de una manera más drástica con cada nuevo dato que se le agregue.

Tomemos por ejemplo una “n” de tamaño 10 y tomemos en cuenta que no tenemos todavía una serie de tiempo y que apenas la vamos a formar. Entonces el pronóstico se tendrá que realizar hasta que contemos con los “n” datos y no antes, así que no podremos pronosticar para los períodos Y_0, Y_1, \dots, Y_{10} , lo cual es una de las desventajas de este

método. Una vez que contemos con estos 10 datos podremos proceder a realizar el primer pronóstico, que será la media de los 10 primeros valores. Aunque esto es un poco contraintuitivo, porque se supone que se establece el modelo en base a la serie y no se crea el modelo sin serie. Pero como estos métodos se utilizan para procesos que necesitan un pronóstico no muy preciso ni costoso, como en el caso de una fábrica de 10,000 tipos de productos plásticos diferentes, el esperar a los datos puede resultar en una desventaja.

$$\hat{y}_{11} = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 + y_{10}}{10}$$

Así, nuestro segundo pronóstico a una etapa será

$$\hat{y}_{12} = \frac{y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 + y_{10} + y_{11}}{10}$$

y así sucesivamente. Podemos escribir nuestro modelo

$$\hat{y}_{t+1} = \frac{y_t + y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4} + y_{t-5} + y_{t-6} + y_{t-7} + y_{t-8} + y_{t-9}}{10}$$

que bien puede crear pronósticos **dentro** y **fuera de la muestra**.

Otra vez este método sólo hace uso de unos cuantos "n" valores y está desperdiciando demasiada información. Para ser exactos está dejando "m-n" valores sin utilizar en el momento t, que es el momento del pronóstico.

Además podemos ver que la observación más reciente Y_t se está multiplicando por el mismo factor 1 que el dato Y_{t-9} más rezagado. Esto, si lo razonamos un poquito, no tiene mucho sentido, ya que se esperaría y es bastante lógico, que lo ocurrido en el pasado reciente tenga mayor importancia para el pronóstico que lo que sucedió hace ya un tiempo.

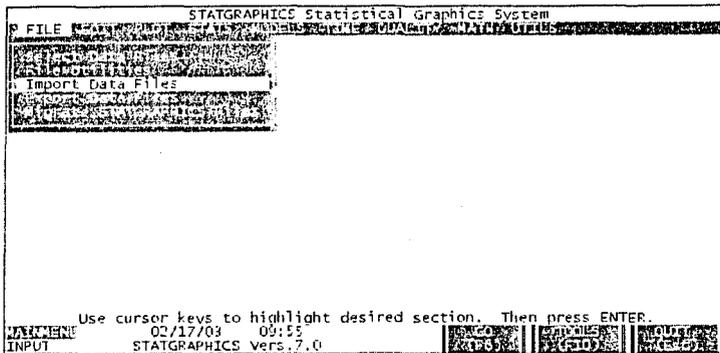
Por supuesto este método no pronostica a más de una etapa, no funciona para series con estacionalidad ni con tendencia y mucho menos con ciclicidad.

Uno se podrá empezar a imaginar desde ahora cuál sería un método ideal: uno que utilice todos los datos del pasado, que pondere las observaciones más recientes con un mayor peso que las que se encuentran más lejanas en el tiempo y que obviamente sea capaz de captar y reproducir los sube-y-bajas de las innovaciones estacionales así como las cíclicas.

El software que se estará utilizando es el de Statgraphics. Se puede utilizar cualquier otro, pero se escoge este por ser el más didáctico.

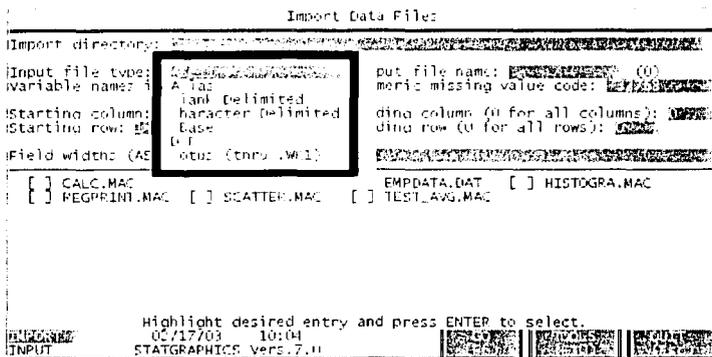
APLICANDO EL MÉTODO PROMEDIO MÓVIL CON STATGRAPHICS

Lo primero que debemos hacer cuando queramos utilizar este software es tener los archivos de datos disponibles en Statgraphics, ya que este programa utiliza un formato especial. Los datos que vayamos a utilizar deben estar guardados en el disco duro de nuestra computadora. Es necesario que los importemos desde el módulo de *FILE-Import Data Files*.



En este módulo se pueden importar cualquier tipo de archivos como DBase, Lotus, separado por espacios, separado por comas, ASCII, etc. Si tenemos los datos en Excel, los podemos guardar con la extensión WK1 de Lotus y después importarlas desde este módulo, seleccionando la respectiva extensión.

TESIS CON
 FALLA DE ORIGEN



Si en el archivo que estamos importando los datos empiezan desde el primer renglón, es decir no están etiquetados, seleccionamos *No* en el campo de *Variable names in first row*. Una vez que seleccionamos el archivo de la lista que nos despliega presionamos F6 (ésta es la tecla de acción de Statgraphics) y nos avisa del nombre que tendrá el archivo dentro de los archivos seleccionables de Statgraphics.

TESIS CON
FALLA DE ORIGEN

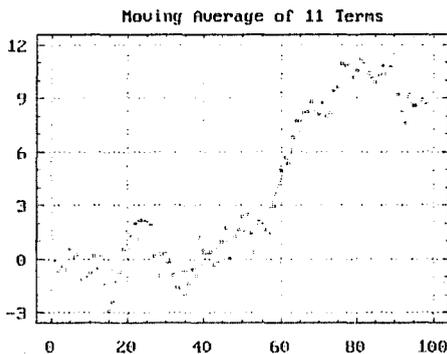
Simple Moving Average

Time series: AR05
Number of points on either side of target: 3

| | | |
|---------|-------|----|
| AR05A | .time | 11 |
| AR05A | .type | |
| AR | .arma | |
| AR05 | .var1 | |
| AR100 | .var1 | |
| AR101 | .var1 | |
| AR102 | .var1 | |
| AR1091 | .var1 | |
| AR1121 | .var1 | |
| AR1 | .var1 | |
| AR112 | .var1 | |
| ARMA11 | .var1 | |
| ARMA9E | .var1 | |
| ARRR | .var1 | |
| ARSAR11 | .var1 | |
| ARRRR | .var1 | 11 |

Highlight desired entry and press ENTER to select.
02/17/03 11:38
STATGRAPHICS Vers. 7.0

Finalmente presionamos F6 con lo que tenemos que seleccionar entre guardar los resultados o mostrar la gráfica de los datos promediados. Seleccionamos ésta última y se nos muestra una gráfica como la siguiente:



Se puede apreciar que la suavización fue algo fuerte ya que no reacciona mucho con los datos y más bien se mantiene en un término medio.

Ahora pasemos a ver el método de **promedio móvil doble**, que no es más que el método de promedio móvil pero adecuado para pronosticar series con tendencia lineal.

PROMEDIO MÓVIL DOBLE

El método consta de 5 ecuaciones, cuatro de las cuales se tienen que calcular una sola vez para pronósticos fuera de la muestra y la última se evaluará tantas veces como etapas se quieran pronosticar.

$$1) \hat{y}_{t+1} = \frac{y_t + y_{t-1} + \dots + y_{t-n+1}}{n}$$

$$2) M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \dots + M_{t-n+1}}{n} \text{ donde } M_t = \hat{y}_{t+1}, \text{ es decir, el pronóstico con promedio}$$

móvil a una etapa.

$$3) a_t = 2M_t - M'_t$$

$$4) b_t = \frac{2(M_t - M'_{t-1})}{n-1}$$

5) $\hat{y}_{t+p} = a_t + b_t(p)$ donde "p" es el número de etapas

Como podemos ver, este método sirve para pronosticar a más de una etapa y a diferencia del anterior, éste puede manejar series con tendencia lineal.

Lo que hace la primera ecuación es pronosticar con los "n" valores de la serie alisando igual que antes. La segunda ecuación repite este suavizamiento, pero esta vez lo hace para todos los pronósticos dentro de la muestra más nuestro pronóstico más reciente. Estima a_t que es donde comenzarán todos los nuevos pronósticos fuera de la muestra y seguidamente estima b_t que describe la pendiente de la tendencia que ha asimilado. Por último crea un modelo que si recordamos es bastante parecido a la función de una recta ($a + bx$).

Nuevamente este modelo tiene muchas de las deficiencias que el anterior de promedios móviles: las restricciones para el tamaño de n son las mismas, considera las observaciones más recientes tan importantes como las remotas. A pesar de que sí detecta la tendencia, no es capaz de modelar una tendencia no-lineal y mucho menos una serie con estacionalidad.

Con esto ya se puede pasar a ver el primer método de suavizamiento: el método de suavizamiento exponencial.

11.1 Suavizamiento exponencial

Para explicar el presente método primero recordemos la forma en que trabajaba el método de promedios móviles. Nuestro promedio iba recorriendo la serie con un ancho fijo "n" e iba pronosticando para la siguiente etapa y_{t+1} . Luego una vez que esta y_{t+1} sucedía o era observada, se volvía un valor observado y pasaba a formar parte del modelo, con lo que la primera observación y_{t-n} (la observación más lejana en el tiempo) se restaba del conjunto de información de tal forma que el modelo volvía a contar con "n" datos.

El suavizamiento exponencial funciona de una manera similar pero con una sencilla e importante diferencia: deja de desplazar los últimos datos para luego desecharlos. En lugar de esto va ampliando su horizonte de entendimiento del proceso con cada etapa, no permitiendo así la fuga de información. En otras palabras, no resta del modelo el término más antiguo, sino que reemplaza éste con un estimado: el pronóstico realizado para la presente etapa, es decir \hat{y}_t .

Tomemos la ecuación de promedios móviles

$$\hat{y}_{t+1} = \frac{y_t}{n} + \hat{y}_t - \frac{y_{t-n}}{n}$$

y en lugar de restar y_{t-n} restemos una aproximación de esto, \hat{y}_t

$$\hat{y}_{t+1} = \frac{y_t}{n} + \hat{y}_t - \frac{\hat{y}_t}{n}$$

reagrupando

$$\hat{y}_{t+1} = \frac{y_t}{n} + \left(1 - \frac{1}{n}\right) \hat{y}_t$$

y sustituyendo $1/n$ por α , la ecuación para el método queda:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t$$

De esta manera hemos logrado armar una fórmula mucho más fácil de calcular dinámicamente. En esta fórmula, cuanto mayor sea alfa (que implicaría tener una menor n) mayor importancia se le dará a las innovaciones y menos confiará el modelo en su propia información. Y por el contrario si tenemos una n grande, conllevará a que nuestra alfa sea pequeña y por ende que el modelo no se mueva tanto hacia la innovación observada y presente un comportamiento más alisado.

$$\alpha = \frac{1}{n}$$

Si $n \rightarrow \infty$ entonces $\alpha \rightarrow 0$

Si $n \rightarrow 1$ entonces $\alpha \rightarrow 1$

Con lo que corroboramos que alfa solo puede tomar valores entre [0,1]. El nombre de este modelo contiene "exponencial". ¿Por qué?. La respuesta es sencilla.

Tomemos nuestra última ecuación y retrasémosla una unidad de tiempo. En otras palabras, cada vez que veamos una t pondremos una t-1, cada vez que veamos una t-1 pongamos una t-2 y así sucesivamente, de tal forma que nos quede

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha) \hat{y}_{t-1} \quad (2.1)$$

Que sería el pronóstico para el tiempo t. ¿Qué sucede si sustituimos esto en el modelo original?

$$\begin{aligned} \hat{y}_{t+1} &= \alpha y_t + (1-\alpha) \{ \alpha y_{t-1} + (1-\alpha) \hat{y}_{t-1} \} \\ \hat{y}_{t+1} &= \alpha y_t + \alpha(1-\alpha) y_{t-1} + (1-\alpha)^2 \hat{y}_{t-1} \end{aligned} \quad (2.2)$$

Antes de continuar, no debemos olvidar que alfa tiene un valor entre uno y cero. Ahora retrasemos una etapa más la ecuación (2.1):

$$\hat{y}_{t-1} = \alpha y_{t-2} + (1-\alpha) \hat{y}_{t-2}$$

Y volvamos a sustituirla en la ecuación (2.2) de manera que nos quede:

$$\begin{aligned} \hat{y}_{t+1} &= \alpha y_t + \alpha(1-\alpha) y_{t-1} + (1-\alpha)^2 \{ \alpha y_{t-2} + (1-\alpha) \hat{y}_{t-2} \} \\ &= \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + (1-\alpha)^3 \hat{y}_{t-2} \end{aligned}$$

si continuamos de este modo, nos encontraremos con que

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \alpha(1-\alpha)^4 y_{t-4} + \dots$$

y advertimos patentemente que las observaciones van teniendo un peso cada vez menor. *exponencialmente* menor, conforme se van alejando hacia el pasado. Esto explica el nombre de suavizamiento exponencial.

Se trata de un modelo que aprende. Dado que el modelo comete errores de pronóstico constantemente, este modelo aprende de los errores. Pronostica conjuntando el pronóstico que realizó para la etapa anterior y el error que cometió.

Ahora vale la pena enumerar las ventajas que nos trae este modelo así como sus desventajas. Tiene una mayor capacidad de asimilación de la forma en que se comporta la serie original, ya que va renovando su fuente de datos conforme avanza en el tiempo.

porque no elimina el pasado por completo como lo hacían los primeros dos modelos que estudiamos, simplemente le va dando menor importancia. Pero a pesar de las mejoras, este modelo sigue teniendo sus debilidades, como son las series estacionales y con tendencia.

Establézcase un nuevo concepto: el de error de pronóstico.

ERROR DE PRONÓSTICO

El error de pronóstico es la diferencia entre nuestro pronóstico para el momento t y la observación en ese mismo momento.

$$e_t = y_t - \hat{y}_t$$

Si modificamos la fórmula del modelo de suavizamiento exponencial para que quede

$$\begin{aligned} \hat{y}_{t+1} &= \alpha(y_t - \hat{y}_t) + \hat{y}_t \\ &= \alpha(y_t - F_t) + F_t \\ &= \alpha e_t + F_t \end{aligned}$$

donde F_t es el pronóstico. Entonces podemos decir que

$$\hat{y}_{t+1} = \hat{y}_t + \alpha e_t$$

EL ERROR DE PRONÓSTICO A UNA ETAPA

Si existen n pronósticos y n valores observados, entonces existirán n términos e_t .

$$1) ME = \frac{1}{n} \sum_{i=1}^n e_i \quad \text{error promedio}$$

$$2) MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad \text{error promedio absoluto}$$

$$3) MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad \text{error cuadrático promedio}$$

No olvidemos que el pronóstico se realiza para el momento $t+1$ con todos los datos recabados hasta el momento t . La primera fórmula ME es bastante intuitiva, pero no es muy buena para indicar el ajuste, ya que permite que se cancelan los valores positivos y negativos, pero puede servir para señalar los sesgos: la sobre o subestimación.

El MAE y el MSE funcionan con un principio en común, ya que ambos cuantifican evitando las cancelaciones. De hecho el MAE tiene la ventaja de ser fácilmente interpretable, aunque el valor absoluto complica su manejo matemático. Esto deja al MSE como el predilecto para la optimización estadística.

El MSE, al igual que los otros dos estadísticos, sigue midiendo la bondad de ajuste pero de datos **históricos**. Dicho ajuste no necesariamente genera un buen pronóstico. Se puede obtener un MSE igual a cero si utilizamos una cantidad lo suficientemente grande de parámetros. A esto se llama sobreajuste.

El sobreajuste de un modelo a los datos históricos quiere decir que se está incluyendo a la aleatoriedad dentro del proceso generador, lo cual es tan inútil y peligroso como fallar en identificar el patrón sistemático de los datos.

Para poder cerciorarnos de que tenemos un modelo que verdaderamente se está ajustando al futuro, partimos nuestra serie en dos periodos. Tomamos la primer parte como nuestra *muestra*, con la que formulamos el modelo. La segunda parte será nuestro pseudo-futuro, con el que podremos comprobar la habilidad del modelo para pronosticar el futuro. Así no tendremos que esperar a que las observaciones sucedan para verificar la eficiencia de un modelo.

CRITERIO DE SELECCIÓN DE MODELOS PARA LOS MÉTODOS DE PROMEDIOS MÓVILES Y SUAVIZAMIENTO

Para escoger entre todos los posibles modelos es necesario contar con un cierto conjunto de características ideales que esperaríamos que contenga un modelo. Lo que buscaremos son dos cosas básicamente: una, que el modelo se ajuste lo mejor posible a los datos, pero no demasiado, ya que abría sobreajuste. Esto significa, en lo cuantitativo, **que el MSE sea mínimo**. La otra característica que buscamos es que **pronostique mejor que cualquier otro**. Es decir, seguimos el procedimiento antes mencionado de partir la serie de tiempo en dos.

A continuación veamos un nuevo método. Éste tendrá las ventajas del modelo anterior, pero además será capaz de manejar series con tendencia lineal.

11.2 Suavizamiento exponencial doble

El suavizamiento exponencial doble o suavizamiento de Brown*, como su nombre lo dice, suaviza dos veces los datos de la serie.

Primero crea una segunda serie A_t que es una representación suavizada de la serie original. Seguidamente crea una serie A'_t que es una nueva representación suavizada, pero esta vez de A_t , no de y_t . Después calcula la diferencia dato por dato entre estas dos series, de manera que pueda localizar la tendencia. Supongamos que el valor en el momento t en la serie A_t aumentó con respecto a su valor anterior $t-1$. Si la serie doblemente suavizada A'_t también lo hizo querrá decir que se trata de una tendencia, porque no olvidemos que el suavizamiento deshace la aleatoriedad, así que si después de eliminar el factor de azar todavía queda un cambio significativo entre dos valores consecutivos, se debe captar este cambio dado que es un cambio que está tomando lugar en el patrón de fondo de la serie.

Este método consta de 5 pasos:

$$1) A_t = \alpha y_t + (1 - \alpha)A_{t-1}$$

$$2) A'_t = \alpha A_t + (1 - \alpha)A'_{t-1}$$

* Robert Brown (1773-1858) descubrió el movimiento que lleva su nombre al realizar estudios con el polen de las flores. Publicó en 1828 un panfleto en el que explica que al estar realizando experimentos con granos de polen de *Clarkia pulchella* notó que tanto los granos sanos como los muertos presentaban un movimiento similar. Extendió el experimento a otras plantas y registró lo mismo. Descubrió con esto un movimiento que es propio de la materia en esas condiciones. El movimiento Browniano es un movimiento continuo e incesante de partículas microscópicas —cuando se encuentran suspendidas en líquido o gas— resultante de golpeteos aleatorios de las moléculas circundantes de líquido o gas.

$$3) a_t = 2A_t - A'_t$$

$$4) b_t = \frac{\alpha}{(1-\alpha)} (At - A'_t)$$

$$5) \hat{y}_{t+p} = a_{t+p} + (p)b_t$$

En donde p es la cantidad de etapas a las que se está pronosticando. Como podemos ver, este método tiene una estructura muy similar a la de promedios móviles dobles.

APLICANDO EL MÉTODO DE BROWN CON SOFTWARE

De la misma lista de métodos en la que aparece el método de promedios móviles simples, seleccionamos el de Brown, que por cierto es el primero de la lista.

```

STATGRAPHICS Statistical Graphics System
TIME 02/16/03 20:26
F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12
-----
Use cursor keys to highlight desired section. Then press ENTER.
INPUT 02/16/03 20:26
STATGRAPHICS Vers. 7.0
    
```

Una vez dentro de la pantalla del método, seleccionamos el archivo de datos importado y asignamos un valor a α .

Para poder escoger el mejor valor del parámetro podemos hacer diversos intentos. En el siguiente ejemplo se seleccionó un valor de $\alpha=0.5$. El programa nos muestra diversos estadísticos ME, MSE y MAE y nos presenta la comparación entre los diferentes valores de α que vayamos asignando.

En este ejemplo estaremos usando la serie de exportación de petróleo crudo de México de enero de 1980 a diciembre del 2002, tomando los datos hasta diciembre del 2001 para crear el modelo y lo que resta para verificar el pronosticamiento. (Fuente: INEGI, Petróleos Mexicanos, Indicadores Petroleros)

Como ya se dijo, uno de los criterios de selección de modelos es que el MSE sea mínimo.

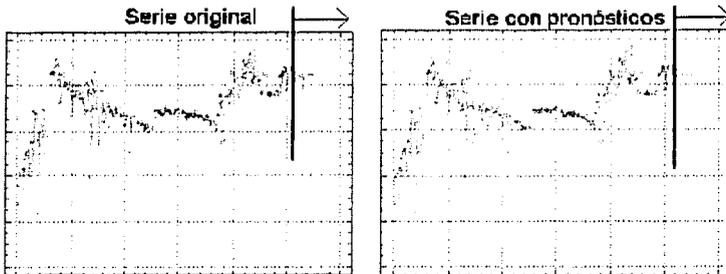
TESIS CON
FALLA DE ORIGEN

**TESIS CON
FALLA DE ORIGEN**

| Data: EXPORT1.var1 | | Percent: 100 | | | | |
|--------------------|---------|--------------|---------|----------|----------|--|
| Forecast summary | | | | | | |
| | M.E. | M.S.E. | M.A.E. | M.A.P.E. | M.P.E. | |
| Simple: 0.1 | 34.9259 | 23937.2 | 99.2532 | 7.42626 | 1.57674 | |
| Simple: 0.5 | 7.07673 | 16298.4 | 77.5061 | 6.15960 | -0.44601 | |
| Simple: 0.9 | 3.93171 | 18537.9 | 79.5305 | 6.21551 | -0.48348 | |
| Simple: 0.2 | 17.5605 | 17985.9 | 84.2907 | 6.50456 | 0.38232 | |
| Simple: 0.6 | 5.89662 | 16577.1 | 77.2029 | 6.14976 | -0.41761 | |
| Simple: 0.4 | 8.83802 | 16253.2 | 78.3775 | 6.20058 | -0.23554 | |
| Simple: 0.3 | 11.7565 | 16602.5 | 80.2049 | 6.29257 | -0.03227 | |
| Simple: 0.8 | 4.42360 | 17664.3 | 78.1390 | 6.21952 | -0.47165 | |
| Simple: 0.7 | 5.05593 | 17019.2 | 77.2763 | 6.15797 | -0.45030 | |

Press ESC, Cursor keys or Page Number: Page 1.1 of 1.2
 03/02/03 19:20
 INPUT STATGRAPHICS vers.7.0

Podemos ver que el modelo que seleccionamos es superado por el de $\alpha=0.4$. Ahora hacemos la prueba de eficiencia cotejándolo contra el futuro real.



Se puede apreciar que los pronósticos generados en esta gráfica caen extremadamente cerca de las observaciones reales, ya que ambas indican cierta estabilización.

SUAVIZAMIENTO POR EL MÉTODO DE HOLT

Este método, también llamado método de dos parámetros de Holt, suaviza la tendencia directamente a través de un parámetro. Este parámetro β es introducido al método, de tal forma que sea capaz de ajustar el peso de la tendencia. Los 3 pasos para este procedimiento son:

- 1) $A_t = \alpha Y_t + (1 - \alpha) (A_{t-1} + T_{t-1})$
- 2) $T_t = \beta (A_t - A_{t-1}) + (1 - \beta) T_{t-1}$
- 3) $Y_{t+p} = A_t + T_t(p)$

Observemos la primera ecuación. Es idéntica a la de suavizamiento exponencial, pero con una T_{t-1} sumada. lo que quiere decir que esta suavizando de la misma manera, pero añadiendo la tendencia para no quedarse rezagada –en caso de que ésta existiera–.

Para comprender la segunda ecuación tomemos en cuenta primero que la T_{t-1} de la extrema derecha es igual a cero al comenzar. Así que solamente nos queda el producto $\beta(A_t - A_{t-1})$. Si la serie no contiene tendencia, la diferencia $A_t - A_{t-1}$ debe algunas veces resultar negativa y otras veces positiva, de manera que después de que $(A_t - A_{t-1})$ haya recorrido toda la serie de principio a fin la sumatoria de esta diferencia debe dar aproximadamente cero. Pero si la serie efectivamente contiene una tendencia considerable

el resultado de la sumatoria lo debe indicar por medio de un valor positivo si la tendencia es positiva y un valor negativo si ocurre lo contrario.

De este modo podemos observar que ésta ecuación trabaja para la tendencia de la misma forma que el suavizamiento exponencial trabaja para las series sin tendencia, siendo A_{t-1} el nuevo valor observado de tendencia y T_{t-1} el pronóstico de tendencia realizado desde la etapa anterior.

Por último la tercera ecuación trabaja de la misma forma que la ecuación de una línea recta, donde p es el número de etapas en el futuro.

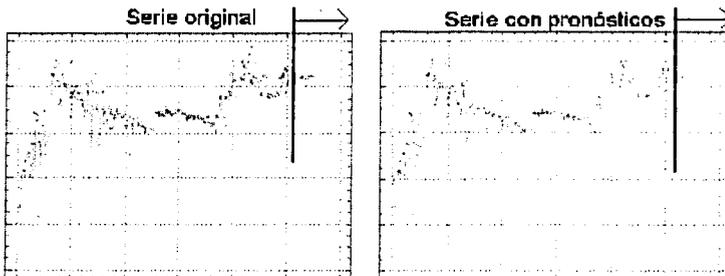
APLICANDO EL MÉTODO DE HOLT CON SOFTWARE

Para este método podemos seleccionar los valores de α y β . Aunque se pueden hacer tantas combinaciones como se deseen, sólo se muestran tres para fines didácticos, de las cuales la pareja (0.9,0.1) muestran el menor MSE.

| Forecast summary | M.E. | M.S.E. | M.A.E. | M.A.P.E. | M.P.E. |
|------------------|---------|---------|---------|----------|----------|
| Holt: 0.5 0.5 | 2.26851 | 23515.8 | 93.6360 | 7.45863 | -0.56782 |
| Holt: 0.9 0.1 | 0.42468 | 21039.0 | 85.1298 | 6.76095 | -0.02400 |
| Holt: 0.1 0.9 | 7.06324 | 39525.7 | 127.959 | 10.1525 | -0.05866 |

Press Esc, Cursor keys or Page Number: Page 1.1 of 1.4
 03/02/03 20:28
 INPUT STATGRAPHICS Vers. 7.0

Recurriendo a la misma serie de tiempo de exportaciones de petróleo crudo que se utilizó en el método de Holt, realizamos la comparación contra los datos fuera de la muestra de la misma manera.



TESIS CON FALLA DE ORIGEN

Podemos apreciar que los pronósticos están bastante cerca de los valores reales, lo cual resulta muy práctico, ya que no olvidemos que se trata del pronóstico para un año.

11.3 Suavizamiento exponencial triple

También llamado TRIX en el ámbito bursátil, es un método que realiza tres promedios consecutivos; la media móvil de la media móvil de la media móvil. Es utilizado para filtrar ciclos insignificantes pudiendo resaltar así los movimientos de mayor peso.

Está compuesto por tres fórmulas:

$$1) A_t = \alpha y_t + (1 - \alpha) A_{t-1}$$

$$2) A''_t = \alpha A_t + (1 - \alpha) A''_{t-1}$$

$$3) A'''_t = \alpha A''_t + (1 - \alpha) A'''_{t-1}$$

Cada ecuación suaviza a la ecuación anterior. Con esto se logra un suavizamiento máximo que lo único que permite que se muestre sean los ciclos, más que cualquier otro comportamiento, como el estacional.

Se acostumbra utilizar para dar una idea más macro de lo que está sucediendo en el proceso, por ejemplo, si se está esperando que aparezca pronto una cima o un valle en el comportamiento de la serie de tiempo, y se quiere analizar el pasado de una manera barata y práctica (todos los métodos de suavizamiento utilizan pocos recursos en relación con la capacidad de las computadoras actuales)

DIAGRAMA CONCEPTUAL DEL FUNCIONAMIENTO DE LO PROMEDIOS MÓVILES Y SUAVIZAMIENTO EXPONENCIAL

En la figura 1 y 2 podemos ver del lado izquierdo una columna con un plano cartesiano, donde las abscisas representan el tiempo, desplazándose hacia el futuro en dirección hacia la derecha. Las ordenadas representan el peso que van teniendo los datos observados en el tiempo. Como se observa en los promedios móviles el peso que tienen los datos conforme se retrocede en el tiempo permanece inmutado, en cambio para el suavizamiento exponencial tiene un decrecimiento exponencial en la importancia de los datos conforme se alejan éstos hacia el pasado.

Del lado derecho se encuentra una representación de los datos recabados que se toman en cuenta en el método y de los que se desechan y desperdician.

FIG. 2.1 PROMEDIOS MÓVILES

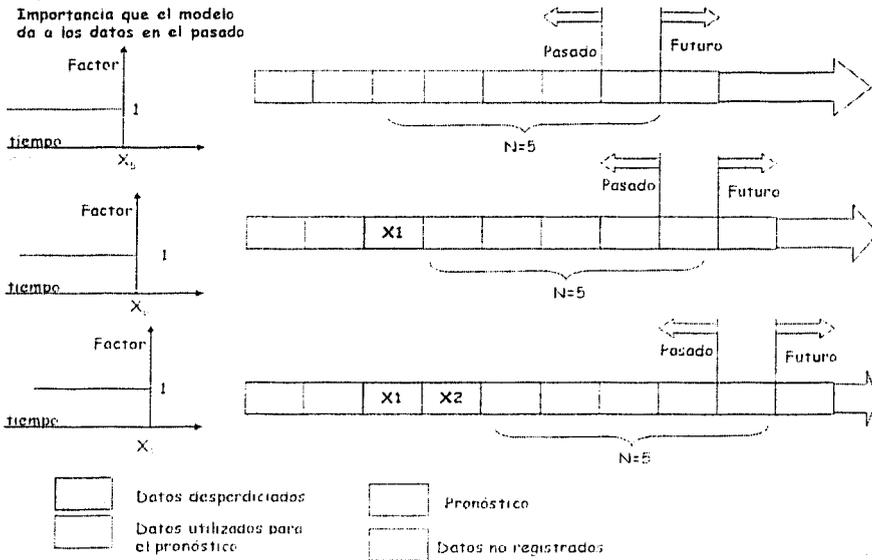
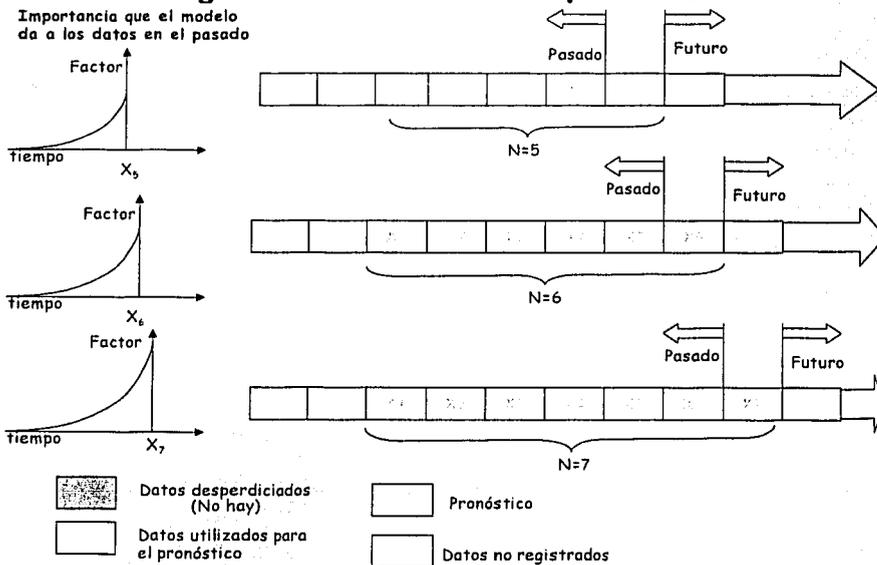


Fig. 2.2 Suavizamiento exponencial



TESIS CON FALLA DE ORIGEN

Capítulo 3. Pronósticos de series de tiempo con variación estacional

III.1 Descomposición multiplicativa.

Hemos visto ya varios métodos que nos ayudan a pronosticar de una manera satisfactoria procesos reales por medio del suavizamiento. Pero estos métodos de suavizamiento en realidad **esbozan** únicamente el comportamiento del proceso, porque tratan de ser "cautelosos" y buscan siempre un término medio, a fin de minimizar su posibilidad de equivocación.

Uno pensará que no es conveniente ser demasiado osado en premoniciones. Pero sí se puede ser más analítico y tratar de comprender mejor el proceso. Una manera de hacer esto es por medio de la **descomposición**. La descomposición separa la serie de tiempo en los cuatro componentes que ya habíamos mencionado antes: estacionalidad, tendencia, ciclicidad y aleatoriedad. Analiza cada uno por separado. Una vez que asimila el patrón de cada uno de los tres primeros componentes por separado, los conjunta para formar el pronóstico y supone que el cuarto componente aleatorio será la diferencia entre el pronóstico y lo real. Esta diferencia será, idealmente, desdeñable.

Existe la descomposición multiplicativa y la aditiva. La primera multiplica los componentes.

$$y_t = S_t \times T_t \times C_t \times R_t$$

Donde:

y_t = Valor real de la serie en el período t

S_t = Componente estacionario en el período t (por sus siglas en inglés: *Seasonality*)

R_t = Componente aleatorio en el período t (por sus siglas en inglés: *Randomness*)

T_t = Componente de tendencia en el período t

C_t = Componente de ciclicidad en el período t

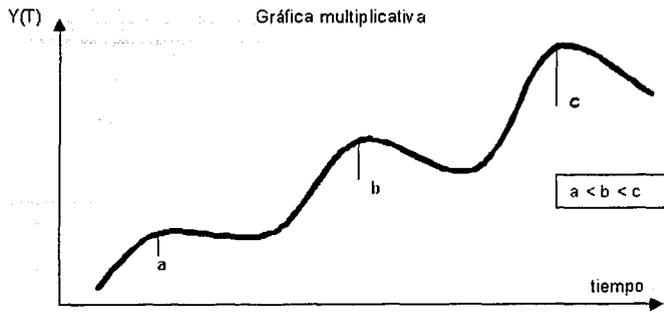
Y la descomposición aditiva los suma:

$$y_t = S_t + T_t + C_t + R_t$$

Comenzaremos por estudiar la descomposición multiplicativa y después veremos la aditiva.

ANÁLISIS VISUAL DE LA DESCOMPOSICIÓN

Una nota importante. Si multiplicamos los efectos entonces algo que no habíamos planeado sucede con la ciclicidad y la tendencia que no sucede cuando sumamos los componentes. La ciclicidad y la estacionalidad se verán afectados por el crecimiento de la tendencia, así que la gráfica de una serie donde el enfoque multiplicativo es el más apropiado se ve como la siguiente gráfica:



Y sin en cambio, una serie donde los efectos tiene una relación aditiva se parece más a la siguiente serie.



Fuente: (Kirkpatrick, 1996)

Como vemos, en la segunda serie la estacionalidad o la ciclicidad permanecen constantes, aunque exista una tendencia. Si una serie no contiene alguno de los componentes ¿cuánto debería valer el factor de ese componente?. Tiene que ser cero para la descomposición aditiva y uno para la multiplicativa.

ERROR Y ALEATORIEDAD

Supongamos que ya tenemos nuestro modelo conformado con el método que se quiera (descomposición, suavizamiento) que obviamente formamos a través de la información proporcionada por los datos de la serie de tiempo –que no es más que una muestra–. Una vez que tenemos nuestro modelo listo para pronosticar *al futuro*, lo primero que tenemos que hacer es pronosticarlo *en el pasado*. Realizamos el pronóstico dentro de la muestra, comparando los datos que nos va arrojando el modelo con los datos de la misma muestra con la que se formó. De esta manera se obtendrá un modelo que puede pronosticar, por lo menos, para el pasado, que es cuando tenemos datos registrados.

Al estar haciendo nuestra comparación dentro de la muestra entre los valores observados y los valores pronosticados, la diferencia entre de estos dos valores se llama residual o error de pronóstico dentro de la muestra. Este error debe ser mínimo para tener un modelo confiable, así que se fuerza a la fórmula pronosticadora a que genere los residuales lo más pequeños posibles. Este concepto de minimizar el error será el principio con el que se encuentren los *modelos óptimos* de pronóstico como veremos más adelante.

Además de pequeño, el error debe ser completamente impronosticable es decir sin patrón de comportamiento identificable, por que si lo pudiéramos pronosticar podríamos mejorar el modelo.

El ruido o componente aleatorio tiene una forma de actuar verdaderamente extraña. Tan extraña que si la tratáramos de pronosticar con el método que fuera estamos seguros que nunca acertaríamos. Carece de patrón, o sea que es completamente impronosticable. De eso estaremos siempre seguros.

Y es en esta certeza donde radica la importancia de la aleatoriedad; aunque no podemos entenderla, sabemos como *se ve*. Y no es de sorprender que la aleatoriedad sea la columna vertebral de todos los pronósticos formales, dado que es lo único que sabemos de antemano como es *en apariencia*.

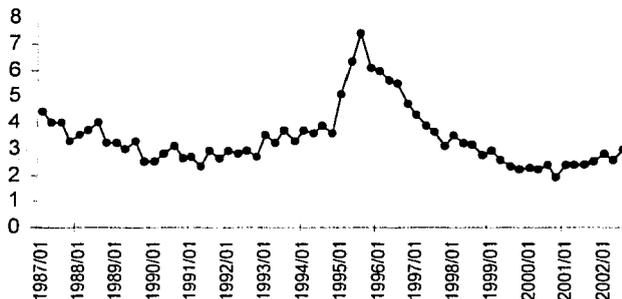
Ya para este punto el estudiante debe estarse imaginando que estas dos cosas, el error y la aleatoriedad, deben estar ligadas de alguna forma, por que la primera debe tratar de ser impronosticable y la segunda ya lo es por formación, además de contar con algo con lo que no habíamos mencionado hasta ahora: la certidumbre, la certidumbre de ser completamente impronosticable.

Por tanto, podemos concluir que el error debe parecerse a la aleatoriedad lo más que se pueda.

Y con esto tenemos dos condiciones para el error en nuestro modelo: debe ser mínimo y actuar como si fuera aleatoriedad. Ahora, si lo que quiere hacer la descomposición es precisamente descomponer la serie, ¿cómo planea hacerlo?. Bueno, por un lado, hay que tener bien claro cómo trabajan las fluctuaciones aleatorias: "golpean" los datos. Azarosamente hacen que aumenten los valores o que decrezcan. Por tanto, si sumamos muchos de estos valores "golpeados", estos "golpes" deberán cancelarse y desaparecer, porque aleatoriamente son golpes positivos o negativos.

Tenemos así que por lo menos en teoría debería ser muy fácil separar la aleatoriedad de la serie. Pero no olvidemos que ahí andan mezclados los otros tres factores.

Para poder entender cada paso del método, nos auxiliaremos con un ejemplo. La serie de datos que estaremos utilizando cuenta con los cuatro factores que deseamos identificar y extraer. Los datos que estaremos usando representan la tasa general de desempleo abierto en México por trimestres de Enero de 1987 a Abril del 2002.



Fuente: INEGI. Encuesta Nacional de Empleo Urbano

CÓMO SEPARAR EL COMPONENTE ESTACIONALIDAD

El componente de estacionalidad es fácil de identificar, ya que se encuentra dentro de períodos pequeños de tiempo, por lo que sólo necesitamos pocos datos. Lo mismo no se puede decir para la ciclicidad ya que es necesario contar con más datos, pero ya atacaremos este tema más adelante.

Al observar la serie advertimos la estacionalidad inmediatamente. Podemos ver como se repite cada cuatro trimestres. Es posible comprender por qué la serie se comporta así: los meses de bonanza económica son los mismos año con año.

Para escoger el tamaño de estacionalidad, que para el ejemplo estamos suponiendo de cuatro, debemos conocer y entender plenamente de qué se trata el proceso cuya serie estamos analizando. Si estamos hablando de desempleo sabemos que puede ser influenciada de alguna forma, de tal manera que cada año se repita el mismo torcimiento en los datos.

Supongamos que estuviéramos tratando con datos semanales de desempleo en vez de mensuales. Si en esta serie apareciera un comportamiento que nos hiciera suponer que existe un tamaño de estacionalidad de tres, es decir, que existe un comportamiento que se está repitiendo cada tres semanas, deberíamos analizarlo más afondo. Nos deberíamos preguntar ¿que está sucediendo cada tres semanas que afecte la forma en que se ve afectado el desempleo?. Sin ser unos expertos en el área, podríamos respondernos que quizá no es nada y que simplemente sea una coincidencia. Si investigamos al respecto y aún así no encontramos algo que nos ayude, podemos concluir que lo más probable es que tengamos razón. Quizá el tamaño de la serie (la cantidad de datos) no es lo suficientemente descriptivo y sólo estamos mirando una mala racha que nos confunde.

Promediamos los cuatro primeros trimestres de la serie

$$MA_1 = \frac{(y_1 + y_2 + y_3 + y_4)}{4}$$

Una vez esto ¿qué tal si hacemos lo mismo pero desde el segundo trimestre hasta el quinto?.

$$MA_2 = \frac{(y_2 + y_3 + y_4 + y_5)}{4}$$

Luego repetimos: cesemos y_2 y añadamos y_6 , abandonemos y_3 e incluyamos y_7 y así sucesivamente hasta abarcar todos la serie. Lo que estamos haciendo es obteniendo el promedio móvil que estudiamos en el tema anterior. Es por eso que hemos escrito las siglas MA en el lado izquierdo de la ecuación, por las siglas de promedio móvil en inglés (Moving Averages).

Si estamos suponiendo que los datos trimestrales contienen fluctuaciones aleatorias, entonces con esta suma de trimestres se está cancelando la aleatoriedad como lo habíamos planteado antes y podemos suponer que estos promedios ya no la incluyen.

Pero también observemos otra cosa. Estamos pasando de tener cuatro datos (trimestres) a tener un solo dato. Es decir, un dato por año. Y así resulta imposible que podamos tener

estacionalidad en estos promedios, dada la definición de ésta. Hay que recalcar que escogimos promediar cuatro datos expresamente por esta misma razón.

Podemos concluir entonces que hemos extraído de nuestros datos originales la estacionalidad y la aleatoriedad, por lo que los datos promediados contienen únicamente los factores de ciclicidad y tendencia.

$$MA = T_1 \times C_1 \tag{3.1}$$

Si acomodamos esto en columnas, queda de la siguiente manera

| (1) Cuarto t | (2) Observación Y | (3) Promedio móvil $T \times C$ | (4) Razón $S \times R \times 100$ |
|----------------------|---------------------------|---------------------------------------|---|
| 1 | 4.4 | — | — |
| 2 | 4 | — | — |
| 3 | 4 | 3.925 | 101.91 |
| 4 | 3.3 | 3.7 | 89.18 |
| 5 | 3.5 | 3.625 | 96.55 |
| 6 | 3.7 | 3.625 | 102.07 |
| 7 | 4 | 3.6 | 111.11 |
| 8 | 3.2 | 3.525 | 90.78 |
| 9 | 3.2 | 3.35 | 95.95 |
| 10 | 3 | — | — |

Sólo utilizamos los primeros diez datos para fines didácticos.

Tomemos nuestros datos observados y dividámoslos entre la tercera columna $T \times C$, ¿qué nos queda?

$$\frac{y_t}{MA} = \frac{S_t \times T_1 \times C_1 \times R_t}{T_1 \times C_1} = S_t \times R_t \tag{3.2}$$

Nos da como resultado el producto de estacionalidad con aleatoriedad. Así que hemos logrado separar en dos grupos los cuatro factores que contiene nuestra serie: uno con C y T y el otro con S y R .

Pero la cuarta columna está etiquetada " $S_t \times R_t \times 100$ ". No olvidemos que siempre debemos hacer que nuestros resultados sean fáciles de interpretar a simple vista. Así que al multiplicar por 100 lo único que se está logrando es precisamente eso, porque si vemos en esta columna un valor mayor a 100 comprenderemos inmediatamente que la estacionalidad y la aleatoriedad tienen un peso mayor para este período que en el resto del año. Y lo contrario, si la proporción $S_t \times R_t \times 100$ es menor que 100 querrá decir que $S_t \times R_t$ tienen un peso menor al promedio anual en ese período.

Pero, aunque hemos hecho un gran avance al desmenuzar parcialmente nuestros datos, no hemos todavía aislado por completo los cuatro componentes. Si retomamos la

mecánica de las cancelaciones de las fluctuaciones aleatorias por medio del promedio, es lógico que la podamos volver a aplicar en donde R_t se encuentre.

Para hacer esto primero acomodemos nuestras proporciones $S_t \times R_t \times 100$ de todos los trimestres en otra tabla y de manera diferente. Pongamos en nuestra primer columna el año correspondiente. En nuestra segunda columna irán las $S_t \times R_t \times 100$ correspondientes a todos los primeros trimestres. En la tercera, las correspondientes a los segundos trimestres, en la cuarta la de los terceros y en la quinta la de los cuartos trimestres.

Luego promediamos las cuatro columnas de la derecha, de manera que nos quede un valor para cada una, o sea cuatro valores únicamente. Estos valores no pueden contener aleatoriedad dado que al hacer la suma por columna la hemos cancelado. Y como resultado, estos cuatro valores contienen única y exclusivamente el factor de la estacionalidad promedio de cada uno de los cuatro trimestres.

$$\overline{S_t \times R_t} = S_t \quad (3.3)$$

Donde la barra horizontal superior denota promedio

| Año | Trimestre | | | | |
|------------|-----------|----------|----------|----------|-----------------|
| | Primer | Segundo | Tercero | Cuarto | |
| 1987 | - | - | 101.9108 | 89.18919 | |
| 1988 | 96.55172 | 102.069 | 111.1111 | 90.78014 | |
| 1989 | 95.52239 | 94.48819 | 110 | 88.49558 | |
| 1990 | 90.09009 | 102.7523 | 112.7273 | 92.85714 | |
| 1991 | 100.9346 | 87.61905 | 110.4762 | 97.19626 | |
| 1992 | 103.5714 | 100 | 102.6549 | 90.7563 | |
| 1993 | 113.8211 | 97.70992 | 108.0292 | 94.96403 | |
| 1994 | 103.4965 | 99.31034 | 105.4054 | 88.88889 | |
| 1995 | 107.9365 | 112.5 | 118.8755 | 94.57364 | |
| 1996 | 95.61753 | 96.55172 | 100.9174 | 93.57889 | |
| 1997 | 93.51499 | 93.46247 | 98.32552 | 87.97737 | |
| 1998 | 103.8576 | 98.61325 | 100.3165 | 91.95021 | |
| 1999 | 101.9264 | 97.16446 | 93.10689 | 93.71672 | |
| 2000 | 101.2155 | 98.23789 | 107.2727 | 86.3892 | |
| 2001 | 105.3097 | 105.0773 | 98.44881 | 100.2973 | |
| 2002 | 108.9494 | 94.39853 | 110.434 | - | |
| S | 101.49 | 98.67 | 105.63 | 92.11 | $\Sigma=397.89$ |
| S ajustada | 102.03 | 99.19 | 106.20 | 92.59 | $\Sigma=400$ |

Pero hay un detalle. Si sumamos los cuatro valores de S debería de dar igual a 400 (es decir $L \times 100$, donde L es el tamaño de estacionalidad), pero no lo hace. En aras de un entendimiento automático, debemos hacer que la suma nos dé igual a 400. Si dividimos 400 entre el resultado de la suma de los cuatro términos, tendremos 1.005317. Si después multiplicamos cada uno de los términos por este cociente tendremos unos valores normalizados que se entenderán a simple vista y que además sumarán 400.

Si el dato 1 correspondiente al primer trimestre presenta un factor de estacionalidad de 101.49, querrá decir que el primer trimestre tiene un peso de estacionalidad del 1.5% mayor que el resto del año.

Otro pequeño problema que podemos identificar fácilmente en nuestras tablas, es el del acomodamiento de los datos. Si observamos la primera tabla, veremos que los datos para los promedios móviles comienzan en la fila 3 y terminan una fila antes, o sea que nos están faltando 3 valores. La razón para esto es porque al promediar, como ya se dijo antes, se pierden datos. Así que si tenemos cuatro trimestres y al promediarlos nos queda un solo valor promedio. Pero la cuestión problema es: ¿dónde acomodamos el promedio?. En realidad, dado que representa el término medio de los datos, el promedio debería ir justamente a la mitad del año, en el trimestre 2.5. Pero desgraciadamente esto no es posible ya que no contamos con datos de medios trimestres. La solución a este inconveniente no es nada complicado: si obtenemos los dos primeros promedios, es decir el promedio de y_1 hasta y_4 y el de y_2 hasta y_6 , tendremos que estos dos promedios corresponden a los trimestres 2.5 y 3.5 respectivamente. Por lo tanto, si promediamos estos promedios tendremos que el resultado corresponde al trimestre número 3 exactamente. A este promedio se le llama Promedio Móvil Centrado (CMA por sus siglas en inglés: *Centered Moving Average*). Resulta obvio que si tenemos una cantidad impar de datos por año no es necesario centrar los promedios.

| (1) Cuarto T | (2) Promedio móvil T x C | (3) Promedio Móvil Centrado |
|--------------------|-----------------------------------|-----------------------------------|
| 1 | — | — |
| 2 | — | — |
| 3 | 3.925 | 3.81 |
| 4 | 3.7 | 3.66 |
| 5 | 3.625 | 3.62 |
| 6 | 3.625 | 3.61 |
| 7 | 3.6 | 3.56 |
| 8 | 3.525 | 3.44 |
| 9 | 3.35 | 3.26 |
| 10 | — | — |

CÓMO APLICAR LA DESCOMPOSICIÓN MULTIPLICATIVA CON STATGRAPHICS

Aunque ya se vio como separar la tendencia, Statgraphics tiene un módulo especial para la descomposición —*Seasonal Decomposition*— ya sea multiplicativa o aditiva. Prosiguiendo con el mismo ejemplo, se pueden introducirle al software los datos y seleccionar el módulo de descomposición.

```

Seasonal Decomposition
Time series: 1982:1RE703
Length of seasonality: 12
Method: ADDITIVE

```

```

Plot seasonal component
Plot residual component
Plot adjusted data
Save moving averages
Save seasonal indices
Save residuals
Save adjusted data

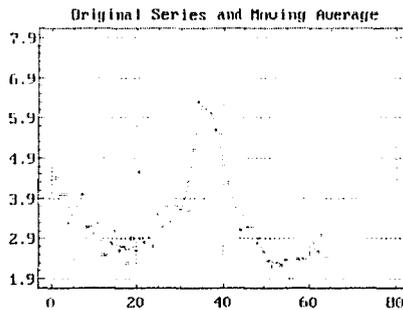
```

```

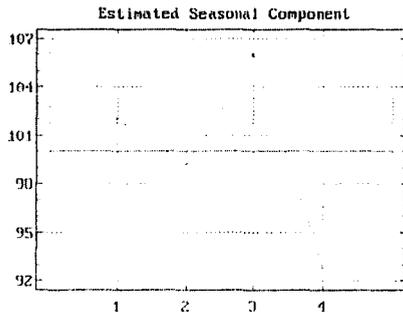
Highlight desired entry and press ENTER to select.
02/18/03 18:33
STATGRAPHICS Vers.7.0

```

Hemos cambiado el tamaño de la estacionalidad a 4 porque estamos trabajando con trimestres (4 trimestres en un año) y dejamos el *Method* en multiplicativo. Una vez hecho esto seleccionamos *Plot Moving Averages*, con lo cual nos muestra una gráfica de los promedios móviles, los cuales contienen sólo **tendencia** y **ciclicidad**.



Desde este módulo también se puede graficar el componente de estacionalidad. Al hacer esto se ubica gráficamente el peso que tiene cada uno de los componentes.



TESIS CON FALLA DE ORIGEN

Los resultados de las gráficas cuadran con nuestros resultados anteriores de las tablas, aunque en una gráfica los datos se tornan más comprensibles. El tercer trimestre adopta valores muy por encima del resto del año. Nos confirma además que efectivamente hay estacionalidad; cuando no exista el patrón debería ser menos alejado de la media y prácticamente una línea horizontal.

CÓMO SEPARAR EL COMPONENTE DE TENDENCIA

Dado que ya aislamos la estacionalidad, ahora se pueden dividir las observaciones entre ella para desestacionalizar la serie.

$$\frac{y_t}{S_t} = \frac{S_t \times T_t \times C_t \times R_t}{S_t} = T_t \times C_t \times R_t$$

Si aplicamos regresión lineal a nuestros datos desestacionalizados obtendremos los valores de a y b correspondientes a la fórmula de la recta $a+bt+ct^2$, donde t representa al tiempo. Al hacer esto estamos tomando por sentado que la línea que mejor se ajusta a nuestros datos es una línea polinómica de segundo grado. Pero para cada caso será diferente el tipo de línea que mejor se ajuste: recta, cuadrática, exponencial o curva en forma de S. Supongamos que la curva que mejor se ajusta es una cuadrática. Los valores de los coeficientes obtenidos con Statgraphics son

$$a=2.75$$

$$b=0.079$$

$$c=-0.00137$$

Con lo que se puede calcular la tendencia

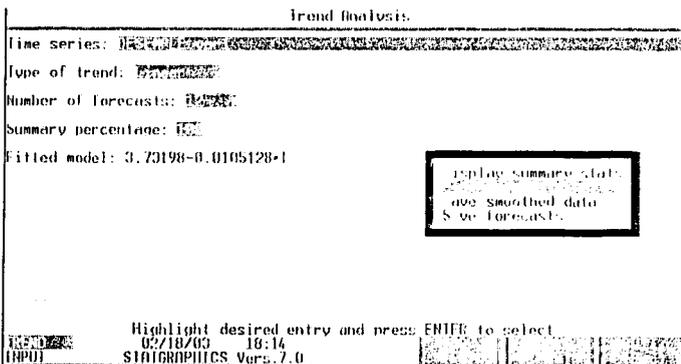
$$T_t = a + bt + ct^2 = 2.75 + 0.079(t) - 0.00137(t^2) \quad (3.4)$$

Una vez que se cuenta con los parámetros a, b y c se puede obtener el coeficiente de tendencia para cualquier etapa del tiempo. Tomemos por ejemplo el trimestre 15, es decir el primer trimestre del quinto año. Su coeficiente de tendencia será:

$$T_{15} = a + bt + ct^2 = 2.75 + 0.079(15) - 0.00137(15^2) = 3.627$$

CÓMO OBTENER LA TENDENCIA CON STATGRAPHICS

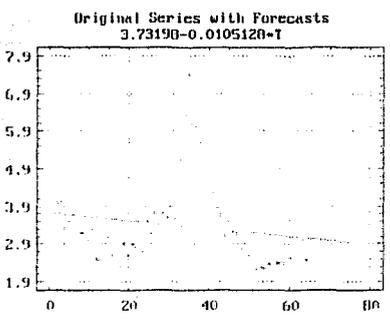
El análisis de tendencia se encuentra en el mismo submenú que los anteriores métodos. La siguiente pantalla nos muestra los valores solución a y b para una tendencia lineal.



TESIS CON FALLA DE ORIGEN

En este caso el valor estimado para la constante es de 3.73198 y la pendiente b es de 0.0105128. Si deseamos ver el pronóstico con sólo la tendencia seleccionamos *Plot forecasts*.

TESIS CON FALLA DE ORIGEN

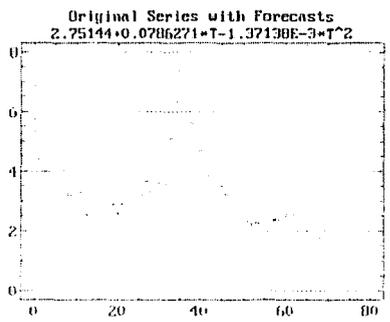


Pero en las opciones de tipo de tendencia "*Type of trend*" existen adicionalmente las tres opciones antes mencionadas: tendencia cuadrática, exponencial o en forma de S. Como se puede apreciar en la siguiente gráfica, el que menor MSE genera es el cuadrático con valores de $a=2.75$, $b=0.79$ y $c=-0.00137$.

| Data: DESEMPL.var1 | M.E. | M.S.E. | N.A.E. | M.A.P.E. | Percent: 100 |
|----------------------------------|---------|---------|---------|----------|--------------|
| Forecast summary | | | | | M.P.E. |
| $3.73198-0.0105128*T$ | 0.00000 | 1.19590 | 0.78690 | 22.7218 | -8.13151 |
| $2.75144+0.0786271*T-1.37138E-3$ | 0.00000 | 1.02082 | 0.76823 | 22.3236 | -6.87331 |
| $EXP(1.30318-3.90631E-3*T)$ | 0.13974 | 1.22330 | 0.75470 | 20.7187 | -3.56919 |
| $EXP(1.14805+0.380206/T)$ | 0.14306 | 1.23284 | 0.74889 | 20.8259 | -3.70176 |

Press Esc, Cursor keys or Page Number: Page 1.1 of 1.4
 TREN 02/19/03 07:29
 INPUT STATGRAPHICS Vers. 7.0

A la hora de pronosticar se genera una curva de tendencia que realmente no es muy convincente, al igual que la lineal, y nos puede hacer pensar incluso que probablemente esta serie no necesite que se le extraiga la tendencia.



CÓMO SEPARAR EL COMPONENTE DE CICLICIDAD

Dado que ya separamos el componente de tendencia, podemos extraer fácilmente el de ciclicidad, dividiendo la ecuación (3.1) de Promedios Móviles entre la ecuación (3.4) de tendencia

$$\frac{MA}{T} = \frac{T \times C}{T} = C$$

Así tenemos que para el dato 15, el componente de ciclicidad tiene un peso de:

$$C_{15} = \frac{MA_{15}}{T_{15}} = \frac{2.675}{3.5725} = 0.75$$

Siendo fieles a nuestro principio de tener nuestros datos interpretables a simple vista, necesitamos multiplicar los coeficientes de ciclicidad por 100, de tal manera que si el producto nos resulta mayor a 100 querrá decir que la ciclicidad en ese trimestre es mayor con respecto al promedio de la serie.

El único componente que nos falta por separar es el de la aleatoriedad. Aunque es también relativamente sencillo de obtener, veremos que el hacerlo nos llevará a conclusiones que no nos figurábamos.

CÓMO SEPARAR EL COMPONENTE DE ALEATORIEDAD

Nuestros datos originales contienen las cuatro componentes

$$y_t = S \times R \times C \times T$$

De la ecuación (3.3) ya obtuvimos la estacionalidad S y de la ecuación (3.1) logramos separar C x T, con lo que solamente necesitamos dividir y_t entre el producto de éstas tres para obtener R

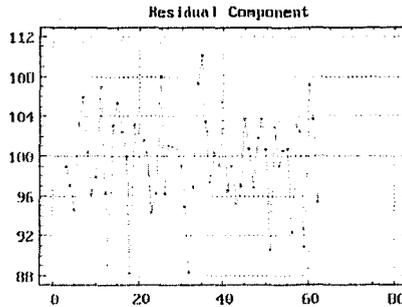
$$\frac{S \times R \times C \times T}{T \times S \times C} = R$$

Dado que por construcción el error o aleatoriedad R es impronosticable, aislarlo no nos es de mucha ayuda para formar nuestro pronóstico, aunque nos puede ayudar a seleccionar entre modelos, ya que comparándolo podemos seleccionar el modelo óptimo. El error dentro de nuestros datos en general nos habla mucho acerca de como será el error de pronóstico en el futuro.

CÓMO SEPARAR LA ALEATORIEDAD CON STATGRAPHICS

En el mismo módulo de Statgraphics con el que se obtuvo la estacionalidad se encuentra la función de graficación de los residuales *-Plot residual component-*.

TESIS CON
FALLA DE ORIGEN



En cuanto al residual podemos ver que toma valores aleatoriamente arriba o debajo de la media, lo cual es un buen indicador de que todos los demás componentes fueron captados de la mejor manera posible.

III.2 Método de Winter

Ahora que conocemos la metodología que sigue la descomposición podemos comprender mejor la forma de trabajar del Método de Winters.

El método de Winter se puede considerar una extensión del Método de dos parámetros de Holt, con la innovación que el de Winter utiliza un parámetro adicional γ (gamma) para suavizar la estacionalidad, lo que conlleva a que debamos resolver para 3 variables independientes: los parámetros α , β y γ , utilizando una ecuación para suavizar cada uno.

La primera ecuación se encargará de eliminar la estacionalidad de la manera que lo hace la descomposición multiplicativa, para dejar solamente los otros tres componentes (C x R x T).

$$A_t = \frac{y_t}{S_t}$$

$$\frac{C_t \times R_t \times S_t \times T_t}{S_t} = C_t \times R_t \times T_t$$

Como este método no puede modelar la ciclicidad, simplemente diremos que y_t incluye $R_t \times T_t$.

Si lo razonamos un momento, el subíndice t en S_t no es del todo adecuado. Recordemos que la estacionalidad es un comportamiento que se repite dentro del lapso de un año pero que obviamente no se presenta en todas las " t ". Así que lo correcto es que el subíndice de S sea $t-L$, donde L es el tamaño de estacionalidad.

$$\frac{y_t}{S_{t-L}}$$

Así, si existe un comportamiento que se repita cada 7 días, en el caso que trabajáramos con datos diarios, L deberá ser igual a 7, de manera que S_{t-L} funcione como una bandera que "señale" que en ese período se encuentra susodicho comportamiento.

Hace falta suavizar, esto se hace como ya vimos, multiplicando el cociente anterior, que es nuestro pronóstico sin suavizar, por un parámetro alfa α . Y al pronóstico de la anterior unidad de tiempo multiplicarla por $(1-\alpha)$.

$$A_t = \alpha y_t + \frac{(1-\alpha)A_{t-1}}{S_t}$$

La segunda ecuación del método es:

$$T_t = \beta(A_t - A_{t-1}) + (1-\beta) T_{t-1}$$

Si se restan dos valores consecutivos que sólo contienen tendencia y aleatoriedad, encontraremos la tendencia y aleatoriedad para ese momento t . Al multiplicarla por α y sumar el segundo término se elimina el azar, como se hacía el suavizamiento exponencial, aislando así la tendencia.

La tercera ecuación es la que aísla la estacionalidad:

$$S_t = \frac{\gamma(y_t)}{A_t} + (1-\gamma)S_{t-L}$$

La cuarta ecuación es la que conjunta los tres componentes para realizar finalmente el pronóstico para m etapas en el futuro.

$$1) A_t = \alpha \frac{x_t}{S_t} + (1-\alpha)A_{t-1}$$

$$2) T_t = \beta(A_t - A_{t-1}) + (1-\beta) T_{t-1}$$

$$3) S_t = \gamma \frac{x_t}{A_t} + (1-\gamma)S_{t-L}$$

$$4) \hat{y}_{t+m} = (S_t + T_{t+m}) A_{t-L+m}$$

Al igual que con el método de Holt, se seleccionan los valores de los tres parámetros y por prueba y error se comparan los valores que menor MSE nos proporcionen. Ya que este es un método que trabaja con series estacionales, utilicemos una serie con comportamiento estacional: la de tasa de desempleo general abierto trimestral.

Winter's Seasonal Smoothing

Time series:

Number of forecasts:

Summary percentage:

Smoothing constant alpha: 0.4

Smoothing constant beta:

Smoothing constant gamma:

Length of seasonality:

TESIS CON FALLA DE ORIGEN

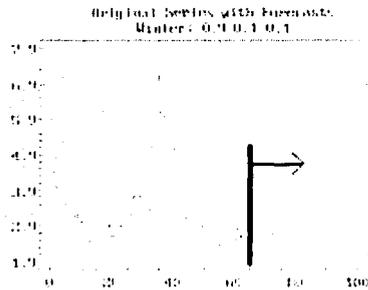
Se puede percibir a continuación que después de varios intentos el modelo que menor MSE genera es con los valores ($\alpha=0.9, \beta=0.1, \gamma=0.1$).

TESIS CON FALLA DE ORIGEN

| Data: DESEMPL.var1 | | | | Percent: 100 | | | | |
|--------------------|-----|-----|-----|--------------|---------|---------|----------|----------|
| Forecast summary | | | | M.C. | M.S.E. | M.A.C. | M.A.P.E. | M.P.E. |
| Winter: | 0.4 | 0.4 | 0.4 | 0.03714 | 0.45083 | 0.48004 | 12.8702 | 1.88486 |
| Winter: | 0.4 | 0.4 | 0.4 | 0.03714 | 0.45083 | 0.48004 | 12.8702 | 1.88486 |
| Winter: | 0.1 | 0.1 | 0.1 | 0.04027 | 2.00493 | 1.17806 | 35.5807 | -4.92974 |
| Winter: | 0.8 | 0.7 | 0.6 | 0.00335 | 0.20840 | 0.42991 | 12.5435 | 0.54443 |
| Winter: | 0.6 | 0.5 | 0.4 | 0.00990 | 0.34167 | 0.41552 | 11.7033 | 0.74372 |
| Winter: | 0.5 | 0.9 | 0.4 | -0.00336 | 0.46438 | 0.48029 | 13.8096 | 0.11076 |
| Winter: | 0.5 | 0.4 | 0.3 | 0.02307 | 0.31674 | 0.43561 | 11.8528 | 1.32868 |
| Winter: | 0.5 | 0.5 | 0.5 | 0.01887 | 0.39692 | 0.44345 | 12.2502 | 1.02791 |
| Winter: | 0.7 | 0.5 | 0.4 | 0.00684 | 0.30195 | 0.41325 | 11.8403 | 0.66664 |
| Winter: | 0.7 | 0.4 | 0.3 | 0.01070 | 0.29150 | 0.39616 | 11.2606 | 0.86503 |
| Winter: | 0.7 | 0.3 | 0.2 | 0.01549 | 0.28626 | 0.38231 | 10.7993 | 1.10956 |
| Winter: | 0.8 | 0.2 | 0.1 | 0.01786 | 0.25874 | 0.36538 | 10.4425 | 1.21581 |
| Winter: | 0.9 | 0.1 | 0.1 | 0.02368 | 0.23845 | 0.34777 | 10.0544 | 1.19621 |
| Winter: | 0.9 | 0.1 | 0.1 | 0.02368 | 0.23845 | 0.34777 | 10.0544 | 1.19621 |

Press Esc. Cursor keys or Page Number: ** ** Page 1.1 of 1.4
 02/18/03 22:12
 INPUT STATGRAPHICS Vers. 7.0

Y como se ve en los pronósticos, el método considera que la serie se estabilizará después de la última caída.



III.3.- Series con variación aditiva

Antes de proseguir, se debe aclarar donde estamos parados con todos estos métodos. La descomposición, a pesar de ser muy fácil de entender y manipular, es un enfoque que es principalmente de carácter intuitivo por lo que no es fuerte en sus fundamentos teóricos, pero aun así da resultados extraordinarios.

Una de sus principales puntos débiles es el modelado de la ciclicidad. Pero era de esperarse. Simplemente imaginemos un comportamiento cíclico de 11 años, como las explosiones solares. Para poder modelar el ciclo necesitaríamos unos 3 o 4 ciclos para identificar el patrón, es decir, necesitaríamos al menos ¡30 o 40 años de datos registrados para poder hacer el modelo!. Por esta razón es por la que la mayoría de los métodos cuantitativos fallan en captar la ciclicidad. En situaciones reales se trata de intuir la ciclicidad por medio de métodos no cuantitativos que resultan más eficaces para este propósito.

La descomposición aditiva es más apropiada cuando la serie de tiempo cuenta con los componentes de tendencia, estacionalidad y error aleatorio exclusivamente. El error

debe ser completamente aleatorio y el comportamiento estacional debe ser el mismo para cada "estación" o período.

Este tipo de descomposición funciona de manera similar que el multiplicativo, con algunas diferencias en las operaciones, obviamente, pero los pasos son básicamente los mismos. Todo el procedimiento que se siguió con la descomposición multiplicativa se puede aplicar a la aditiva reemplazando las multiplicaciones de los componentes con las sumas y las divisiones con las restas.

**TESIS CON
FALLA DE ORIGEN**

Capítulo IV. Métodos cualitativos

*"A red sky at night, sailors' delight
a red sky in the morning, sailors take warning"*

*"Un cielo rojo en la noche, delicia de los marineros
un cielo rojo en la mañana, los marineros se ponen alertas"*

PROVERBIO DEL FOLKLOR INGLÉS PARA PREDECIR EL CLIMA

El método más primitivo de pronóstico es adivinar. El resultado puede ser calificado de aceptable si la persona que hace la adivinación es un experto en el asunto. La adivinación es el único método que puede hacer uso del conocimiento implícito que el especialista no ha sido capaz de expresar en palabras o cifras exactas.

La razón por la cual no nos limitamos a los métodos cuantitativos en la vida real es porque la mente humana parece tener capacidades que todavía no se terminan de entender. Es capaz de manejar una gran complejidad sin que nos percatemos de ello. Relaciona las cosas y encuentra patrones por sí sola, lo que nos permite presentir comportamientos futuros de los que se puede sacar provecho para la toma de decisiones.

Hay casos que se estudian con vehemencia ya que hacen gala del potencial de la mente humana, y que la ciencia, en su afán de entenderlo para poder imitarlo, ha hecho su mejor esfuerzo. Por ejemplo los pronosticadores exitosos y los jugadores de ajedrez; tienen mucho en común cuando se habla acerca de toma de decisiones.

Los campeones mundiales de ajedrez tienen la facilidad de *intuir* la mejor jugada, pero a la vez dependen, al igual que cualquier otro jugador de ajedrez, de años de entrenamiento. Es decir, por un lado está el presentimiento: conocimiento inefable y que por tanto ellos mismos no pueden explicar, y por el otro está la experiencia.

Se han puesto cámaras debajo de los tableros de ajedrez para seguir el movimiento ocular de los campeones de ajedrez e increíblemente los ojos empiezan 3 de 4 veces con la mejor jugada. Esto quiere decir que incluso antes de analizar conscientemente el tablero, ¡ya saben cuál es la jugada ganadora!. Más impresionante aún, los campeones de ajedrez no pueden hacer esto en tableros donde las piezas de ajedrez han sido colocadas aleatoriamente, lo que indicaría que su mente está identificando automáticamente (¿e inconscientemente?) si se trata de una partida real o no (Wright and Goodwin, 1990).

Otro interesante ejemplo es el de personas con aparentes limitaciones mentales que son capaces de proezas sobrehumanas. Oliver Sacks, neurofisiólogo autor del libro "*Despertares*" habla acerca del caso de dos hermanos gemelos John y Michael que podían "ver" el día de la semana en que caería cualquier fecha que se les pidiera de aquí a 40,000 años, respondiendo en unos cuantos segundos. Esto aunado a que los gemelos no podían realizar la más sencilla de la suma o restas y ni siquiera entendían lo que era una multiplicación. Gauss, uno de los más grandes matemáticos y famosos calculadores se encontró con grandes dificultades al tratar de encontrar un algoritmo similar. Suponiendo que el algoritmo lo conocen los gemelos ¿cómo pueden utilizarlo si no saben sumar o restar?.

Un hecho más puede ayudarnos a responder esta pregunta. Un día en el que el neurofisiólogo se encontraba con los gemelos, cayó accidentalmente una caja de cerillos de una mesa esparciendo el contenido por todo el piso. En un instante y al mismo tiempo los gemelos dijeron "111" y después John murmuró "37". Después Michael repitió "37" y John volvió a susurrar "37". El doctor contó los cerillos, lo cual le tomó un tiempo, y comprobó que efectivamente eran 111. Cuando se les preguntó cómo pudieron contarlos tan rápido respondieron "no los contamos, *vimos* el 111". Sin saber sumar ni restar encontraron los factores primos (37,37,37) de 111 inmediatamente. (Revista "Ciencia y desarrollo" Septiembre-Octubre 1985 p. 15-28).

Éstos ejemplos que se muestran bástennos para darnos una idea de lo que pasa dentro de nuestras mentes y de cómo resultaría un desperdicio desde un punto de vista práctico limitarnos a trabajar con métodos cuantitativos que, aunque fueron creados por nosotros mismos, se cierran muchas puertas debido a la mecanización irrestricta e indiscriminante que desconoce el contexto del problema.

MÉTODOS CUALITATIVOS EN LA PRÁCTICA

El mejor método para obtener tal pronóstico del experto es la entrevista. El método de la entrevista nos permite inquirir sobre las razones y explicaciones para el pronóstico presentado, que podría optar por criticar y así intentar llegar a un pronóstico mejorado. Cuando entrevistamos a un experto puede que también aprendamos algo que más tarde podamos usar si preferimos construir nuestros propios pronósticos con otros métodos. Podemos en ocasiones conseguir nombres y direcciones de expertos que vivan lejos y a los que sería difícil entrevistar. Para consultar a tales expertos, podemos recurrir a un cuestionario en lugar de la entrevista. Si deseamos preguntar a varias personas simultáneamente, podríamos considerar el uso del método Delphi, que se verá más adelante.

Los sondeos se pueden ver como una extensión de la entrevista. La cualidad del sondeo es que puede tomar una foto instantánea en un cierto momento. Con ella se obtiene una idea de la percepción que está teniendo la gente en un momento dado. Esto influencia las decisiones que se deben tomar, porque podríamos saber si debemos o no implementar algo para lo que la gente no está lista todavía. La televisión por ejemplo se basa principalmente en lo que a la gente le gustaría ver más, con lo que deciden a qué actores, productores y guionistas contratar para satisfacer esta demanda. O por el contrario, en base a los sondeos, las televisoras concentran sus esfuerzos en alguna actitud que les gustaría cambiar de la gente.

Veamos, antes de proseguir a ver los métodos, algunas aplicaciones de los métodos cualitativos que han resultado en prognosis acertadas.

El siguiente es un extracto de un artículo "Las Comunicaciones" escrito en 1965 por John R. Pierce, en donde se adelanta a predecir la invención de las actuales computadoras de bolsillo, como la PALM, así como de los teléfonos celulares:

"La tecnología de los circuitos integrados hará posible unos adelantos fundamentales. Tales circuitos se emplean ya en los laboratorios, pero aún no en el hogar ni en la oficina: mas cuando se conviertan en productos comerciales en escala mucho mayor será posible tener aparatos tan complicados como una computadora eléctrica en el teléfono, el coche o incluso en el bolsillo: serán, además, baratos.

necesitarán poca energía, durarán decenios, y cuando se estropeen se los tirará, en lugar de repararlos.

¿Qué adelantos prometen tales aparatos? Si se dispone del número suficiente de radiofrecuencias, podrán permitir que haya teléfonos en todos los coches, y acaso que los llevemos en el bolsillo. Pero, a su vez, el teléfono rebasará sus antiguas funciones, ya se encuentre en la oficina, en el hogar o el automóvil; y una de las cosas que hará es ponernos en contacto con computadoras electrónicas”.

Pero para no dar una idea errónea al lector, también leamos lo que dijo esta misma persona acerca del destino del teléfono con pantalla, que nunca se logró comercializar a gran escala*:

“En los Estados Unidos, la compañía Bell está trabajando en esta dirección con un servicio de ‘ver mientras se habla’ llamado PICTUREPHONE, del cual tengo en mi oficina un terminal correspondiente a uno de los primeros diseños: me enlaza con unas 30 personas situadas en el mismo Murray Hill de Nueva Jersey en que trabajo y en Holmdel, a unos 50 km de distancia. Una vez que se hayan resuelto los problemas de puesta a punto, se dispondrá comercialmente de tal servicio”.

Las cualidades de los métodos cualitativos se usan para pronosticar a **largo plazo** principalmente y no se espera que apliquen los patrones históricos**.

IV.1 Métodos exploratorios

Los métodos exploratorios enfocan el problema del pronóstico con una extrapolación a partir de la situación del momento, con la que se intenta ver de qué modo los conocimientos y realizaciones del presente se desarrollarán, probablemente, en el futuro.

GRUPOS DE ENFOQUE

Este es el equivalente al pronóstico intuitivo de los métodos cualitativos. Un grupo de gente se reúne para discutir un tema en especial y sólo ese tema, durante 1.5 horas aproximadamente. El mediador lo único que hace es introducir al tópico de interés. No hay agenda ni lista de preguntas. Los integrantes ya se les ha hecho saber con anterioridad el tema de discusión, y este grupo de integrantes está conformada por personas con diversos grados de experiencia en el tema. Todos se escuchan y discuten, para que si alguien escucha algo que no sabía o no se le había ocurrido, puede hacer comentarios adicionales al respecto.

La ventaja es que los integrantes siguen desmenuzando el problema hasta que se sienten satisfechos y llegan a un cierto consenso.

Si se utilizan 2 o 3 grupos diferentes de personas para estudiar el mismo tema se obtienen diferentes enfoques, se explota al máximo el conocimiento de los expertos, ya que casi siempre todos tratan de sacar a relucir todo lo que conocen acerca del tema. Se logra así una captación de la dimensión, y un gran entendimiento, del problema.

Este método ha dado buenos resultados en el pasado y se viene utilizando desde hace mucho tiempo aunque no se le había asignado un nombre.

* Ver el Anexo 5 para ejemplos de pronósticos erróneos famosos.

** Fuente: Karen Popovich, especialista en pronósticos de la Universidad Estatal de Cleveland, tomado de la página: <http://grail.cba.csuohio.edu/~popovich/oms335>

MÉTODO DELPHI

El método Delphi es probablemente el método no cuantitativo más utilizado. Fue inventado por Olaf Helmer en la Corporación RAND, uno de los *think-tanks* californianos.

El método Delphi pronostica con la ayuda de un grupo de expertos en el área que tratan de llegar a consensos, con la característica de que elimina algunos de los problemas que se presentan en los métodos en los que los pronosticadores tienen que socializar. Los expertos no pueden estar en contacto ni comunicarse entre ellos a través de todo el proceso, de manera que las opiniones que haga cada uno no puedan ser influenciadas por la opinión del resto. Está diseñado para la obtención sistemática de opinión experta. Existen 3 características que lo distinguen con la interacción en grupo: el anonimato, retroalimentación iterativa controlada y respuesta de grupo estadísticamente manejada.

Desgraciadamente la exactitud de los pronósticos obtenidos está seriamente limitada por la calidad de las opiniones proporcionada por los expertos, pero a pesar de sus limitantes está patente su efectividad y se utiliza en una proporción mucho mayor a los demás métodos cualitativos.

La forma en que trabaja es como sigue. Les son enviadas a los expertos unas formas con los asuntos que se quieren tratar y se les piden que escriban sus opiniones. Una vez que las formas son contestadas y devueltas, se establecen un conjunto de resultados fragmentando las respuestas de los expertos en conglomerados más pequeños, cada uno con el consenso en la idea "promedio". Después se les envían las formas nuevamente a los participantes, con especial atención a aquellos que se alejaron demasiado del consenso general y se les pide que expliquen la razón de sus respuestas tan alejadas de lo *normal*. Así se les fuerza a los expertos a que razonen nuevamente sus opiniones: si tiene una buena razón la respaldarán y, si no, los conglomerados se harán menos. Este proceso iterativo se repite hasta que no exista un cambio significativo en los pronósticos.

Este método es utilizado generalmente para pronósticos de largo alcance, buscando ideas primero a través de un proceso llamado lluvia de ideas* (*brainstorming* en inglés) y luego proponiendo las más populares al mismo grupo para que las evalúe. El método se utiliza en los casos en que, por ejemplo, se desea conocer cuándo un producto va a tener aceptación a gran escala o qué nuevos inventos científicos van a aparecer.

Las principales desventajas que presenta son que las respuestas son demasiado sensibles a la ambigüedad de las preguntas, las respuestas son diferentes cuando se utiliza un diferente grupo de expertos y la imposibilidad de predecir lo inesperado.

Este método ha sido sometido a prueba las fuerzas aéreas de los E.U., la organización sueca para investigaciones militares y la compañía norteamericana espacial TRW Systems; posiblemente su aplicación más importante se halle en campos en los que el pensar intuitivo continúe siendo una de las fuentes principales de datos de entrada y en los que el consenso tenga gran importancia, como sucede en la determinación de objetivos a perseguir.

La siguiente figura demuestra un fragmento del resultado de las predicciones haciendo uso del método de Delphi, realizado en los años sesenta bajo el auspicio de la

* Ver el glosario para una definición formal.

corporación RAND. Se aplicó a un grupo de veinte personas destacadas en varios campos de la ciencia.

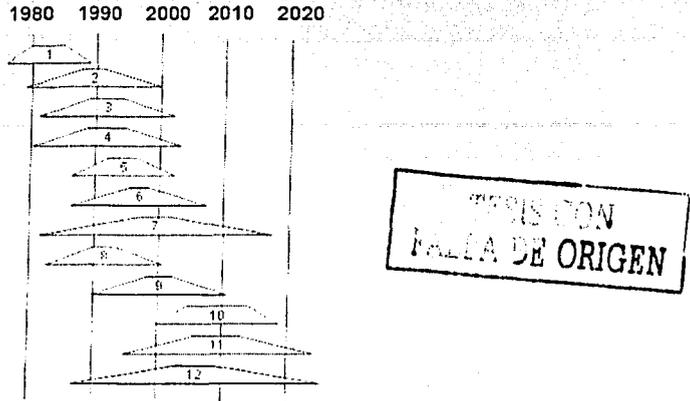


Fig. 4.1 En el diagrama se ven las medianas de las respuestas finales y el intervalo comprendido por el 50% de ellas. (Fuente: Jantsch, 1965, p. 49-50).

Los temas sobre los que se inquirió fueron los siguientes:

1. Emisión estimulada (láseres) en la región de los rayos X y gamma.
2. Regimiento de la energía termonuclear.
3. Vida artificial primitiva (al menos en forma de moléculas que se automultipliquen).
4. Minería económica del fondo de los océanos.
5. Factibilidad de la regulación económica regional del tiempo metereológico.
6. Factibilidad económica de la síntesis de las proteínas para la alimentación.
7. Multiplicación por diez del número de caos de psicosis reducibles mediante terapia física o química.
8. Inmunización bioquímica general frente a enfermedades bacterianas y virales.
9. Factibilidad de la ingeniería genética destinada a domeñar defectos hereditarios.
10. Cultivo económico del océano de suerte que proporcione, por lo menos, el 20 por ciento de los alimentos mundiales.
11. Productos bioquímicos estimulantes de la formación de nuevos órganos y miembros
12. Factibilidad del empleo de fármacos para aumentar la inteligencia.

Es notable que la mayoría de estos pronósticos hayan fallado tanto. El número 3, por ejemplo. En la actualidad se habla de computadoras que pueden arreglar por sí mismas pequeños errores de software, pero el hardware sigue siendo el dominio del ser humano. Así que si todavía no resolvemos el problema de la auto-compostura en las computadoras a nivel macro y por tanto distamos todavía mucho de la autorreproducción de las mismas a nivel micro.

Se puede apreciar la predisposición de los pronosticadores en algunos temas, como es el caso del cultivo económico del Océano. En la actualidad se habla cada vez más de la escasez de alimentos y la ONU ha vaticinado de hecho que el siglo XXI será el siglo de las

migraciones, debido en parte a este factor. Así que a lo mejor en su momento no *creían* seriamente que fuera posible cultivar el mar pronto, pero sí lo *querían*.

A grosso modo es posible concluir que resulta difícil anticipar todos los factores que darán lugar a nuevas invenciones, ya que algunos pueden pasar inapercibidos al principio, pero pueden funcionar como detonadores para un cambio radical en la complejidad del mundo.

Un ejemplo es el de la píldora conceptiva. Cuando se inventó, no fue posible prever todos los cambios sociales que induciría. Con la invención de la píldora las mujeres pudieron entrar de lleno al campo laboral, ya que adquirieron un mayor control sobre sus vidas y su cuerpo. Las disputas en el campo laboral por los empleos se tornaron mucho más equitativas. Se dio paso a la revolución feminista, donde las mujeres reclaman derechos que desde siempre les habían sido negados. Con ello se re-dibujó el esquema demográfico del planeta, sobre todo de los países del primer mundo, ya que las mujeres adquirieron una mayor responsabilidad por los hijos que tenían.

PROCESO DE JERARQUÍA ANALÍTICA (AHP)

AHP por sus siglas en inglés Analytical Hierarchy Process, es un método de toma de decisiones que encuentra la mejor decisión por medio de un análisis introspectivo del tomador o grupo de tomadores de decisiones. Es decir, lo hace reflexionar metódicamente forzándolo a establecer sus prioridades y luego a cuantificarlas por medio de valores numéricos asignados a comparaciones uno a uno entre éstas. Fue inventado por Thomas L. Saaty*. Existe un software que facilita el uso de este método llamado "Expert Choice", creado también por el Dr. Saaty.

Si alguien nos pide que comparemos la importancia que tiene para nosotros el llegar temprano a la escuela contra el desayunar bien y seguidamente nos pide que asignemos un valor numérico que represente la prioridad entre estas dos situaciones, podríamos entonces establecer un valor de 80/100 para el levantarnos temprano, por ejemplo, y el valor complemento a cien de 20/100 a desayunar bien en el caso de que si nos importe tener algo en el estómago al llegar a clase, pero estemos dispuestos a sacrificar esto por llegar en punto a la escuela.

El método parte de la decisión objetivo y luego trabaja hacia atrás: dado que la decisión objetivo traerá consecuencias que conocemos y que serán diferentes dependiendo de la decisión que tomemos, debemos subdividir y ramificar las consecuencias de manera que podamos revisar si efectivamente estamos dispuestos a asumir la responsabilidad de

* En 1970 el gobierno de Egipto le pide a Tom Saaty, un matemático pionero en su área, que resuelva el conflicto en el Medio Oriente con Rusia. Lo logra a través de Teoría de Juegos y obtiene resultados satisfactorios, pero él no se siente satisfecho y comienza a investigar. Descubre el AHP, un método que logra capturar la intensidad del sentimiento humano y asignarle valores numéricos. Él mismo crea años más tarde el software que realiza este mismo proceso y que a la vez lo hace más eficiente. El software es llamado Expert Choice y es considerada una de las herramientas más eficaces en la toma de decisiones hoy en día: es el método de soporte de toma de decisiones más usado en el mundo. El Dr. Saaty es actualmente profesor en la Universidad de Pittsburg. Obtuvo su Doctorado de Matemáticas en la Universidad de Yale y realizó trabajo posdoctoral en la Universidad de París. Ha trabajado para el departamento de defensa de los Estados Unidos y ha sido consultor para toma de decisiones de diversas compañías y agencias, como son la fuerza aérea de los Estados Unidos, Xerox, Boeing, Merck, General Electric, Conoco Oil, Rockwell International y los gobiernos de los Estados Unidos, Canadá, Sudáfrica, Singapur, Indonesia, China, Irán, Argentina, Egipto y Kuwait.

cada una de estas decisiones. Una vez que hayamos ramificado hacemos un chequeo por parejas de manera explícita de éstas contra nuestros objetivos y nuestra filosofía como empresa. Y ya que hemos establecido qué nos importa más y qué menos, podemos tomar la decisión de una manera responsable, segura y con una visión más clara de que lo que nos depara.

Se cuantifican numéricamente las decisiones a causa de que desafortunadamente no siempre podemos dividir todo en blanco y negro: o prefiero esto o prefiero esto otro, sino que debido a que estaremos "pesando en la balanza" situaciones que nos pudieran parecer sin importancia, que ni siquiera habíamos pensado o que preferiríamos no tomar, resolveremos éste problema asignando valores numéricos del 1 al 100.

Ventajas del método: su aplicabilidad es única, ya que es un proceso que se resuelve de la manera en que resolvemos los problemas de la vida diaria. Es muy humano, ya que logra un alto grado de involucramiento en el problema por parte del usuario. Logra crear un balance entre el precio que sí se está dispuesto a pagar y el que no para llegar a cumplir un objetivo.

Desventajas del método: puede llegar a ser un proceso muy exhaustivo, incluso utilizando el software ya que las ramificaciones de decisiones pueden hacerse muchísimas.

EXTRAPOLACIÓN DE CURVAS DE TENDENCIA DE LARGA DURACIÓN

La extrapolación de tendencias es un enfoque puramente exploratorio que funciona solamente cuando los datos se están comportando como uno *espera*, lo que de entrada lo hace bastante limitado. Su precisión disminuye conforme las interacciones se vuelven más complejas y más influyen el pensamiento normativo al proceso.

El objetivo de este método es encontrar las tendencias de largo alcance, que existen además de, o en conjunto, con la ciclicidad. Podríamos decir que buscamos la tendencia *verdadera*. Es como si ampliáramos nuestro enfoque y tratáramos de ver a la ciclicidad como si fuera estacionalidad y observáramos la tendencia como lo hacemos normalmente. El problema es que si no se cuentan con por lo menos 25 años de registros, no se pueden encontrar las tendencias de largo alcance y por tanto este método es inaplicable.

Así que queda en las manos de los expertos analizar los datos cuidadosamente y decidir de qué tipo de tendencia de largo alcance se trata: lineal, exponencial o en forma de S.

La tendencia en forma de S es muy común de encontrar y se puede confundir fácilmente con una tendencia lineal. La forma de S surge a partir de los límites naturales de algunos procesos.

Por ejemplo las ventas de televisores blanco y negro, a color y computadoras personales han seguido un patrón en forma de 'S' con respecto al tiempo. Y suena lógico que así sea, porque al principio las ventas no son muchas, debido a que se trata de un artículo de novedad y sus precios no son muy accesibles. Después, conforme la tecnología se hace más barata, los precios disminuyen lo que hace que el artículo en cuestión se vuelva accesible a las masas y las ventas y producción se disparan. Una vez que todo el mundo tiene uno y que ha pasado la euforia ya no existe una necesidad tan exacerbada y las ventas se estabilizan, aunque no desaparecen ni mucho menos disminuyen.

Este enfoque se denomina de "curva de crecimiento" ya que la curva presenta un comportamiento similar a la forma en que algunos seres crecen en la naturaleza. Por ejemplo nosotros los seres humanos crecemos a una razón que gráficamente tiene forma de S, en donde la curva superior "de estabilización" de la S representaría nuestros años como adultos jóvenes.

Algunos autores⁷ advierten sobre sus desventajas. Presentan como ejemplo el caso de los televisores en los E.U. En los años 50's y 60's los tamaños de las televisiones en los hogares norteamericanos no cesaron de crecer, lo que las compañías electrónicas tomaron como una clara tendencia de que en los años subsecuentes la tendencia continuaría. Pero no fue así, las compañías japonesas se percataron de esto y concentraron sus esfuerzos en televisores más portátiles y con ello encontraron un nicho de mercado en el que a la fecha se encuentran todavía.

Una de las grandes ventajas que tiene este método que es muy útil para pronosticar las características de los avances futuros –por ejemplo sobre la base de los límites naturales– incluso antes que la realización técnica siquiera se haga visible. Es muy utilizada en medios militares.

TESIS CON FALLA DE ORIGEN

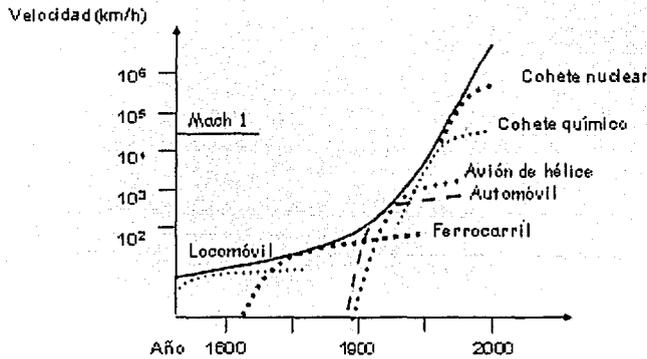
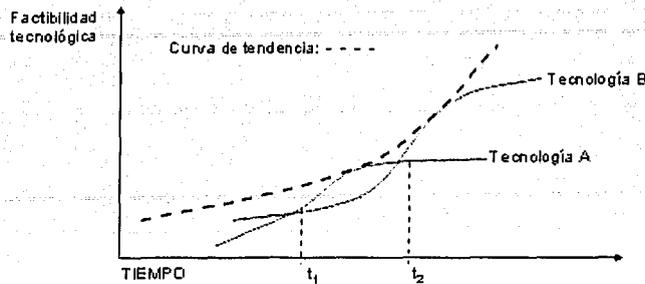


Figura 4.2. La extrapolación de curvas es un importante medio de pronósticos de factibilidades tecnológicas. La velocidad máxima lograda por el hombre aumenta de acuerdo con una curva en forma de S, y cada una de las tecnologías tiende a estabilizarse a su respectivo nivel de 'saturación'. El pronóstico por extrapolación de la curva de tendencia significa, esencialmente, pronosticar el efecto del avance decisivo siguiente, sin definir con qué tecnología se llevará a cabo. (Fuente: Jantsch, 1965, p.27)

⁷ Ascher, W., 1978. "Forecasting: An appraisal for policymakers and planners" Johns Hopkins University Press, E.U.; Schnaars, S., 1989. "Megamistakes: Forecasting and the myth of rapid technological change" The Free Press, E.U.; ambos tomados de <http://www.hfac.uh.edu/MediaFutures/forecasting.html>



TESIS CON FALLA DE ORIGEN

Figura 4.3. El gráfico hace visibles las ventajas y los riesgos de la extrapolación de curvas de tendencia. Para una compañía que explote la tecnología A puede ser muy ventajoso advertir y poner a punto la B antes de que aquella comience a estabilizarse: por otra parte una comparación directa de las tendencias respectivas de A y B en el instante t_1 o poco tiempo después podría llevar a conclusiones a largo plazo erróneas en caso de que no se les sitúe en el contexto de la gráfica de la tendencia (pocas firmas preverían la situación t_2 en el instante t_1). (Fuente: Jantsch, 1965, p. 29)

ANALOGÍAS

Cuando se cuenta con información acerca de un cierto proceso, la suficiente como para poder pronosticar a largo plazo, es posible realizar pronósticos acerca de otro proceso que le sea similar o que se le parezca. Aunque no es necesario que se comporten cuantitativamente igual, sino al menos proporcionalmente. Por ejemplo, si conocemos la manera en que se desarrolló la Revolución Industrial, dado que se cuentan con 200 años de datos, y conocemos poco acerca de la Revolución de la Información, que empezó en 1946, se puede percibir claramente que, en apariencia, lo que le tomó a la Revolución Industrial 50 años, a la de la Información sólo le ha tomado 12.5 años. Contamos con una idea de la proporción que va a tener el crecimiento científico en el área de la tecnología de la información en el futuro.

También se pueden hacer analogías con objetos de la naturaleza. Por ejemplo, en el libro de Alvin Toffler. "The third wave" ("La tercera ola") la analogía de la ola es usada para describir la evolución desde la sociedad agrícola a la industrial, y más tarde a la sociedad de la información. Oswald Spengler, en el libro "Untergang des Abendlandes" ("La decadencia de Occidente"), explica cómo las culturas antiguas de Egipto, Roma y muchos otros países se desarrollaron como plantas o animales que nacieron, crecieron, florecieron, declinaron y murieron; Spengler hizo la predicción de que la cultura occidental, de forma análoga, seguiría el mismo patrón.

Pero pueden surgir errores fatales si no se hacen las analogías correctas. Por ejemplo se ponderaron las ventas del Laser Disc y se hizo una analogía con el crecimiento que tuvo televisión a color en el mercado. La popularidad del Laser Disc finalmente no tuvo el comportamiento tan parecido que se esperaba y resultó en grandes pérdidas económicas para los que confiaron en el pronóstico.

Así que la importancia en el manejo de este método recae sobre la analogía que se está tomando como precedente: el método funciona, pero si se escogen sesudamente las comparaciones.

MATRICES DE IMPACTO CRUZADO

Parecido al Delphi, este método (Cross Impact Matrices, su nombre en inglés) trata de obtener la factibilidad de que un determinado grupo de eventos (avances tecnológicos) suceda. Una matriz de cross-impact maneja dos conjuntos de valores: el primero es la probabilidad de que un cierto evento suceda en un período determinado de tiempo. El segundo es la probabilidad de que, una vez que haya sucedido el evento, éste suceso afecte o influencie en la factibilidad de ocurrencia de cualquiera de los otros. La información de la matriz puede ser asignada por un método como el Delphi.

El fin del análisis de impacto-cruzado es el de refinar y cernir las diferentes probabilidades de ocurrencia de futuros avances tecnológicos y su interacción con otros avances, para poder así realizar planes en el presente tomando en cuenta los avances que se esperan en el futuro.

Para cada elemento de la matriz se tiene que hacer la siguiente pregunta: "¿Si el evento renglón sucediera, cómo afectaría la ocurrencia del evento columna?"

Las matrices en principio sólo se llenan con flechas, indicando hacia arriba si la ocurrencia del evento-renglón aumentará la factibilidad de ocurrencia del evento-columna. La flecha debe apuntar hacia abajo cuando se considera que disminuirá la probabilidad de ocurrencia, y no se pone ninguna flecha si se considera que el renglón y la columna no tienen ningún tipo de interacción. Después se necesita la opinión de los expertos o algún tipo de método subjetivo para poder asignar valores numéricos en vez de flechas.

Es posible crear fórmulas matemáticas que calculen las probabilidades de los eventos conforme algunos eventos se vayan volviendo realidad.

El equipo encargado de ponderar los valores de las matrices debe estar formado de 5 a 10 participantes. De preferencia los gerentes con capacidad de decisión dentro de la empresa o incluso consultores externos. El moderador debe tener habilidades tales como: establecer una relación de comunicación con el grupo, establecer las reglas de interacción del grupo, establecer los objetivos, provocar una discusión intensa en las áreas relevantes y realizar las conclusiones de respuesta del grupo.

Las ventajas de las matrices de impacto cruzado sobre los otros métodos son que toma los cambios sociales en cuenta y elimina la interdependencia entre preguntas.

Una de las desventajas es que normalmente la ocurrencia de un evento tiene como influencia varios eventos y no solamente uno, lo cuál hace más difícil y complejo el uso del método. Pero a pesar de esto, ha encontrado una gran aceptación y ha sido de gran ayuda en aplicaciones de actualidad tanto en el sector privado como en el público.

| Impacto en el evento : Evento : | E ₁ | E ₂ | E ₃ | E ₄ |
|------------------------------------|---|---|---|---|
| E ₁ | _____ |  (+) | _____ |  (+) |
| E ₂ | _____ | _____ |  (-) | _____ |
| E ₃ |  (-) | _____ | _____ | _____ |
| E ₄ | _____ | _____ |  (-) | _____ |

TESIS CON FALLA DE ORIGEN

Las flechas indican hacia el evento que está siendo afectado y el signo indica si está reforzando o impidiendo su ocurrencia.

TRABAJANDO DESDE ABAJO

Looking from Bottom to Top como se le conoce en inglés, es un método que consiste en simplemente mirar alrededor y observar lo que la gente está haciendo. En un libro de John Naisbitt llamado "Megatrends" se ejemplifica perfectamente*. Consiste en buscar en diversos periódicos, temas noticiosos en común, para después hacer un análisis de cuántas líneas (el autor los mide en pulgadas) se les dedica en cada ejemplar y hacer un análisis de cómo va cambiando esta medida con el tiempo, con lo que se pueden detectar tempranamente tendencias.

Se cita a continuación un ejemplo de la aplicación de este método, tomado de La Jornada, de uno de los analistas políticos más reconocidos a nivel mundial, Robert Fisk.

En Enero pasado [Enero 2002] Enron tenía 1.137 menciones en New York Times, Washington Post y Los Angeles Times, mientras se mencionaba a Irak sólo 200 veces. Los reportes sobre Irak se incrementaron casi 100 por ciento a principios de la primavera, cuando las menciones a Enron disminuyeron 50 por ciento, hasta llegar a sólo 618.

Tras una baja a principios de verano, las menciones a Irak se dispararon hasta llegar a mil 529, y Enron recibía sólo 310 menciones.

(Fuente: Periódico La Jornada, 26 Enero 2003, "En la búsqueda de petróleo", Robert Fisk.)

Con los cambios que ya se detectaban en la primavera del 2002, un analista que hubiera seguido este método podría haberse percatado de que se *apuntaba* a la situación actual en dónde es de todos sabido que está a punto de estallar la guerra, pero el saber esto hace un año hubiera dado una ventaja en la toma de decisiones.

* Publicado en 1982, este libro es uno de los éxitos más grandes de la historia de la publicación de libros. Es el resultado de 10 años de investigación. Ha sido publicado en 57 países y ha vendido más de 8 millones de copias. John Naisbitt se ha vuelto el filósofo global entre los futuristas.

IV.2 Métodos normativos

El enfoque normativo consiste en partir del futuro y retroceder hasta el presente; dicho de otro modo, lo que pretende es aislar algún objetivo deseable (por ejemplo, la tecnología de la producción agrícola de alimentos) y, retrocediendo en el tiempo, intentar descubrir en qué medida tendrán que desarrollarse los conocimientos y técnicas existentes para que pueda alcanzarse aquel objetivo.

PATTERN

El método PATTERN se puede considerar como el ejemplo perfecto de la metodología normativa.

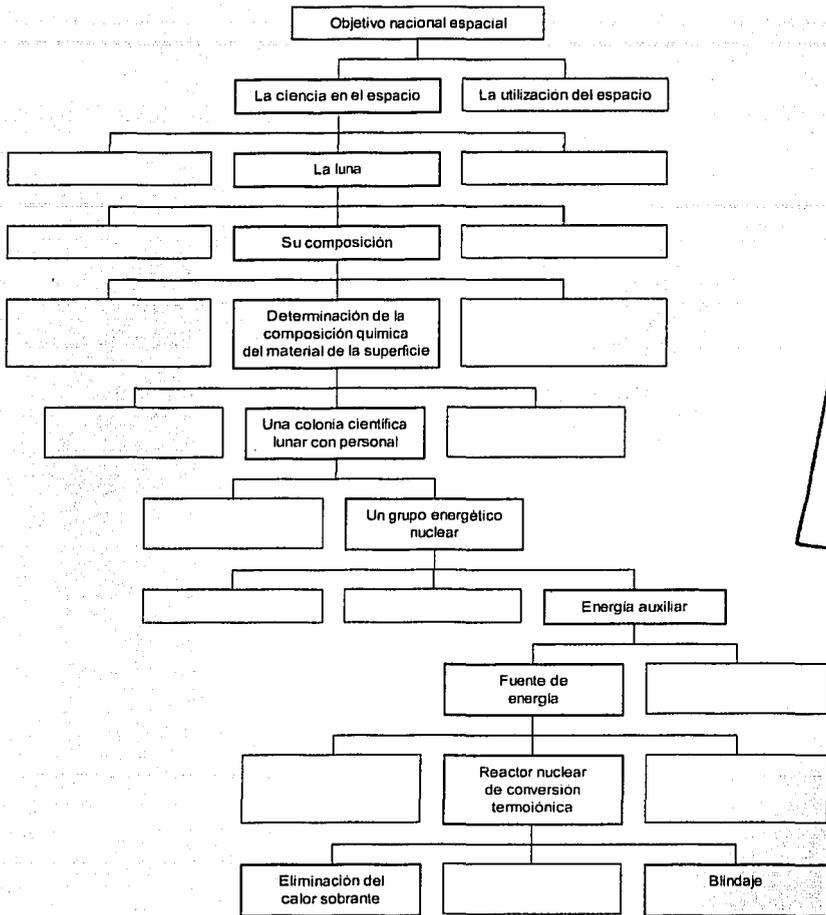
Primero se deben establecer los objetivos deseados. Después se van desglosando los avances tecnológicos necesarios para poder lograr los objetivos, que serán los sub-objetivos. Una vez esto repetimos lo mismo tomando los sub-objetivos como si fueran objetivos. Iterativamente vamos a llegar a los avances científicos y tecnológicos que se necesitan alcanzar en la actualidad. De esta manera podemos saber en el presente en qué áreas debemos canalizar más nuestros esfuerzos para poder lograr lo que queremos.

Es común que se utilicen diagramas con ramificaciones de las alternativas que se van generando en cada iteración, a los que se les llama árboles de decisión.

Una vez que se tiene armado el árbol, los expertos votan para asignar valores que representen la relevancia o importancia de cada evento del árbol (por ejemplo se puede acordar que las "calificaciones" vayan del uno al cien). En esta etapa los expertos se pueden reunir para discutir las calificaciones.

Teniendo los objetivos en la parte superior y las sub-metas en la parte inferior, se procede a reasignar valores de peso, mediante un proceso en el que se multiplican las probabilidades de cada hoja y rama, hasta llegar a los tallos en el tronco. De tal manera que el evento-tallo cuyos eventos-ramas y eventos-hojas hayan tenido mayor calificación que las demás, presentará el valor principal y por tanto es el avance tecnológico de mayor preponderancia y que necesita mayor atención en la actualidad.

Se ha aplicado a las actividades aeroespaciales y de electrónica médica de la compañía Honeywell, a la determinación del programa Apolo de la NASA (en el que se empleó para especificar no menos de 2,329 deficiencias tecnológicas que habían que ser superadas) y a diversos problemas de las fuerzas aéreas de los E.U. En la actualidad, la armada de ese país y el Battelle Memorial Institute están aplicando otras versiones de éste mismo método.



TESIS CON FALLA DE ORIGEN

Fig. 4.4. La NASA aplicó el árbol esquemático PATTERN a la evaluación de la carga útil del programa Apolo. (Fuente: Jantsch, 1965, p. 18)

FUTURIBLES

La Prospective, como también se le conoce a este método, es un método de pronósticos de largo alcance que funciona de una manera radicalmente diferente a las que hemos visto hasta el momento: el futuro es la razón de ser del presente.

Difiere en la forma de pensar de los métodos anglo-sajones, porque considera que el futuro no es algo que ya está escrito y que necesita descubrirse, sino que el futuro se va armando con cada acción que tomamos. No todos tenemos la misma influencia en el futuro y no existe una sola persona de la que dependa por completo el futuro, así que existen una diversidad de posibles futuros, y se debe trabajar para lograr el que queremos. Pero para lograr lo que queremos es necesario establecer normas o reglas.

Cuando comenzamos un negocio, lo hacemos porque consideramos que el tipo de negocio que estamos por comenzar va a ser próspero en el futuro. O sea que estamos tomando una decisión basándonos en nuestras expectativas (pronósticos), pero a la vez nuestra decisión está afectando el futuro.

ADOPCIÓN DE ROLES

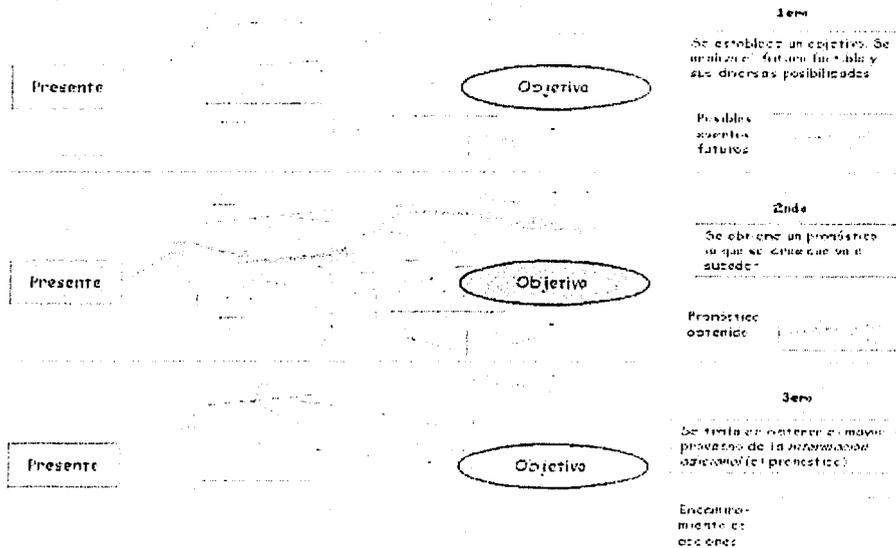
A veces el realizar un pronóstico, o simplemente una observación, como acabamos de ver, puede alterar el futuro. Esto se da en la física (principio de incertidumbre de Heisenberg) y en las ciencias sociales. Afortunadamente para éstas últimas existe el *Role Playing*, como se conoce este método en inglés. Consiste, como su nombre lo indica, en actuar o personificar una cierta situación de manera que, a través de su representación con actores y escenarios ficticios, ambos lo más parecidos a la situación real, podamos prever cómo sería un posible resultado de la situación que nos interesa pronosticar. De esta manera se encuentran pronósticos, que aunque no resultan muy económicos, han probado ser muy eficientes y justifican con creces la inversión hecha en la Adopción de roles.

Es un método que pretende ser abarcador y totalizador, porque se pueden incluir muchos factores influyentes diferentes. Al final los participantes no llegan necesariamente a una predicción, si no a una gama de ellas, cada una con una cierta probabilidad de ocurrencia que los mismo actores asignan.

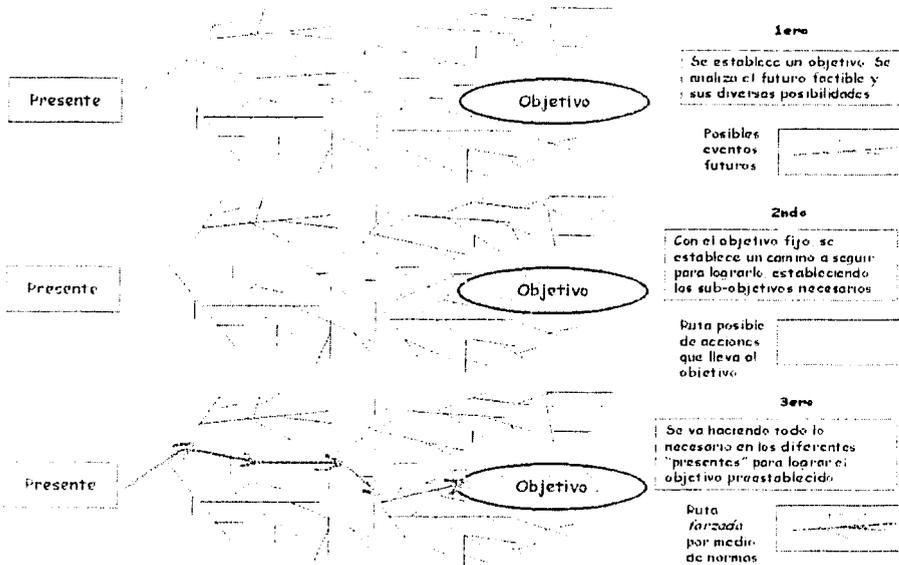
Una de las ventajas más importantes que tiene es que se pueden simular situaciones de "si se *hiciera* esto", se pueden jugar con las interrelaciones intercambiándolas, y no se tiene que modelar las situaciones complejas.

Por ejemplo, se puede hacer un intento por augurar en qué va a terminar el conflicto palestino-israelí. Se preparan a ambos grupos de actores, unos representando a los líderes políticos palestinos y otros a sus contrapartes israelíes y se les explica cuáles son los puntos que deben defender y cuáles son las creencias de ambas partes. Una vez hecho esto, se procede a realizar pequeñas sesiones de menos de una hora, en la que se debe discutir sin perder nunca el rol que se ha adoptado. Se deben realizar cuantas sesiones sean necesarias para llegar a algún tipo de consenso. Y es este consenso el que se utilizará como pronóstico. Pero de no llegarse a uno, son los participantes los que deben dar su opinión acerca de cómo creen ellos que terminaría el conflicto.

FIG.4.5 ENFOQUE DE LOS MÉTODOS EXPLORATORIOS



4.6 ENFOQUE DE LOS MÉTODOS NORMATIVOS



TESIS CON FALLA DE ORIGEN

**TESIS CON
FALLA DE ORIGEN**

Capítulo V. Pronósticos y planeación

"¡Sapere aude!"

"¡Ten coraje para usar tu propia razón!"

LEMA DEL SIGLO DE LAS LUCES

"El sueño de la razón produce monstruos"

FRANCISCO DE GOYA

Uno de los problemas más comunes al momento de aterrizar los pronósticos en casos reales, como en el caso de empresas que verdaderamente utilizan y necesitan de los servicios de un pronosticador, es que los pronosticadores normalmente se aprenden dos o tres métodos bien y después sólo andan buscando a ver en qué los aplican. Cuando en verdad se debería trabajar al revés, primero analizando y comprendiendo plenamente el problema y la necesidad del pronóstico, y después buscando, entre toda la gama de métodos que hemos visto, el más adecuado.

Es la experiencia de la gente que ha trabajado durante muchos años con pronósticos y pronosticadores que es necesario tener a dos grupos de personas como parte del equipo de pronosticadores. El primero debe ser experto en métodos. Es decir, se dedica exclusivamente a aplicar correctamente el método que se le pide y, obviamente, conoce todos los métodos que existen. El otro grupo de personas (que al igual que el anterior bien podría constar de una sola persona) debe estar conformado por gente que comprenda plenamente el ambiente administrativo y sus problemáticas, así como la forma en que funcionan diversos métodos para pronosticar en la administración, para que sea a este grupo de personas a quien se le reporten los resultados del primero. Ambos grupos deben estar bien al tanto de las limitaciones y alcances de cada uno de los métodos.

Es un error muy común tratar de buscar mayor exactitud y precisión en los pronósticos por medio del escalamiento a métodos más matemáticos, complejos, costosos y difíciles de entender. Es un hecho comprobado empíricamente que un método más complejo no necesariamente va a funcionar mejor que uno sencillo. Además, el escalamiento de los métodos sólo conlleva a una cadena de problemas: los administradores dejan de entender qué es lo que hace el método, es necesario contratar pronosticadores más especializados y los que están tienen que tomar cursos para poder manejar los nuevos métodos.

No se puede procurar ser precisos desde el principio. Cuando se pretende implantar metodologías de pronóstico donde nunca las había habido, es necesario comenzar prácticamente con cualquier cosa, que funcione de manera aceptable, para que así, tanto los administradores como los pronosticadores, con el tiempo y poco a poco, comiencen a comprender bien los métodos y puedan más tarde, si es necesario, cambiar de método.

Además, en la gran mayoría de los casos, el problema es atacado desde una perspectiva incorrecta, ya que la mayor cantidad de imprecisión de los métodos proviene de datos inadecuados o incompletos, la utilización de los métodos erróneos, la falta de pronosticadores bien entrenados y la falta también de una sistematización que permita pronosticar en casos semejantes o repetitivos.

En resumen, al comienzo es preferible manejar métodos sencillos y fáciles de entender para los administrativos, en lugar de métodos a los que se les ve como "cajas negras" aunque éstos produzcan mejores resultados.

DATOS ADECUADOS

Al momento de escoger el método o modelo es indispensable que se establezcan claramente las limitantes de la información disponible. Si lo que se requiere son resultados para planear a largo plazo, es imposible de realizarse si solamente contamos con pocos datos. Es inútil pensar que el largo plazo se va a comportar como el corto plazo. Los pronósticos de largo alcance dependen en gran parte de la tendencia "global", en cuanto que los de corto plazo son muy sensibles a la estacionalidad y los efectos cíclicos. Así que se debe prestar especial atención a la calidad de la información a ser equiparada con la calidad esperada del método de pronóstico y sus alcances.

CARACTERÍSTICAS DE UN PRONOSTICADOR EN JEFE

Primero, el gerente encargado del pronóstico debe entender la situación para la cual se está haciendo el pronóstico y saber qué es lo necesario para hacer las decisiones correctas en su área. Segundo, el gerente debe estar verdaderamente preocupado por mejorar las tomas de decisiones. De lo contrario no se involucrará lo suficiente como para hacerlo trabajar, ya que existen muchos detalles implícitos en el área de pronósticos, como los son una actitud humilde pero valerosa, que nadie puede enseñar y que sólo se logran cuando se está verdaderamente interesado en sacar adelante y ayudar en las tomas de decisiones. Asombrosamente esto también está respaldado empíricamente por gente con mucha experiencia. Y por último, lo que ya habíamos mencionado antes, el gerente debe estar familiarizado con la forma en que trabaja el método, para que le pueda tener fe y confianza, y básicamente para que sepa qué es lo que él está haciendo.

Se tiene que mantener a la gerencia de áreas necesitadas o que hacen uso de los pronósticos, al corriente y entrenada en la metodología de los mismos. También resulta muy útil hacer del conocimiento de toda la compañía los aciertos que han tenido en cuanto a pronósticos.

V.1 Planeación y decisiones

Planear para el futuro significa que estamos seguros que desde ahora mismo ya está escrito cómo va a suceder todo, solamente que no tenemos esa información. En este caso al pronosticar estamos adivinando algo que está fuera de control y al tomar las decisiones estamos preparándonos o resguardándonos para lo que no podemos controlar. Por el contrario, si planeamos el futuro o lo inventamos, creemos íntegramente que somos capaces de transformar y dar forma al porvenir a pesar que estamos seguros que sucederán cosas impredecibles. En este último caso, al tomar las decisiones concienzudamente, estamos creando nuestro propio futuro de una manera responsable. Desde cierto punto de vista se podría considerar a este último enfoque como uno más valiente*.

Irremisiblemente los temas de planeación y decisiones nos llevan al tema de la condición humana y su naturaleza, porque todos en algún momento de nuestras vidas

* Ver "casualidad" y "causalidad" en el glosario.

debemos tomar alguna decisión y en esta decisión nadie nos puede realmente decir cuál es la "verdadera" elección, es decir, la correcta.

De entrada es posible preguntarse ¿qué se quiere decir por "correcta"?, es acaso ¿lo más ético?, ¿lo mejor para mí?, ¿lo mejor para la persona que más quiero?, ¿lo mejor para el Universo?, ¿lo mejor para un ser Supremo celestial que me ha conferido alguna misión en este mundo y espera que la cumpla? Y en caso que supiera a quién o a qué es lo que debo beneficiar, ¿cómo llego a saber cómo es que debo beneficiarlo?. Estas cuestiones se pueden tornar –aunque muy interesantes– muy filosóficas, demasiado para ser analizadas en esta tesis, pero no es un tema que se deba descuidar sobre todo para aquellos que tenemos la fortuna de poder estudiar y reflexionar sobre más cosas que lo necesario para sobrevivir.

En el ambiente empresarial, donde se deben tomar muchas decisiones diariamente, la cuestión se vuelve muy común y la mayoría de las veces bastante sencilla. Pero cuando aumenta el "peso" de la decisión, es decir, los bienes (económicos, humanos, etc.) que están en juego, es necesario ser más concienzudos. Pero ser concienzudos cuando se tienen que considerar muchos y muy diversos factores es un trabajo en el que una computadora pareciera ser más apta que un ser humano.

Más adelante, en el tema de "Problemas en los Pronósticos de Juicio" se ve a fondo el tema, pero anticipémonos al plantearnos esta pregunta: ¿Por qué todos (en general) tomamos nota en nuestra agenda de nuestros compromisos futuros, para que no se nos olviden, pero no apuntamos nuestros sesgos mentales que alguien, o nosotros mismos, nos ha ayudado a descubrir? Por ejemplo: "tiendo a ser más agresivo con las mujeres que con los hombres; Cuando es cuestión de mucho dinero me acobardo. Recordar: no acobardarme a la primera, pensarlo bastante; Antes de las 10 de la mañana estoy de un humor de los mil diablos. Evitar al máximo cerrar tratos a esta hora.", etc.

Así que de tomar una decisión importante no es cosa fácil: es cuestión de valor e inteligencia (se consideran el instinto y el "conocimiento inexplicable" dentro de la inteligencia, para fines de simplicidad). Cuando falta alguna de estas dos (o ambas) se recurre ya sea a que alguien más tome la decisión por nosotros o a "echarnos un volado". En cualquier caso estamos relegando una responsabilidad y fallando en formar parte de este mundo, pasando a ser sólo espectadores. Pero tomar una decisión sin fundamentos no es mejor que tomar la decisión de no tomar la decisión por el momento. No es preferible la valentía a la inteligencia como regla, ni lo contrario, lo que se necesita una mezcla *inteligente* de ambas. La responsabilidad acompaña intrínsecamente al tomador de decisiones, ya que no se vale pretender que no se sabe algo que sí se sabe, o tomar una decisión cuando hay el tiempo suficiente de aprender algo que puede ayudarnos a tomar una mejor decisión.

V.2 Comparación y selección de métodos

Pese a que puede resultar sobrecogedor el tener una opción tan grande de métodos, afortunadamente no existe razón para preocuparse, ya que en realidad uno debe de mantener las cosas sencillas. Existe un principio muy útil que ha sido cerciorado empíricamente que es el de **parsimonia**. Este principio establece que en igualdad de condiciones, un método sencillo es preferible a un método complejo.

Existen razones de peso para que los métodos más pequeños y sencillos resulten más atractivos. En primera, se pueden estimar con más precisión los parámetros de los modelos parsimoniosos. En segunda, dada la simplicidad de los modelos, son más fáciles de entender y de revisar, por lo que con más facilidad se transmiten para efecto de ser utilizados en la toma de decisiones. Y un tercer punto, pero más importante que los anteriores, es que los modelos sencillos resultan ser más realistas y menos propensos a modelar o imitar los caprichos del pasado, lo que puede resultar ser un verdadero problema ya que el que el modelo se ajuste perfectamente al pasado no necesariamente quiere decir que va a ser el mejor modelo para pronosticar, ya que se está ajustando a las discrepancias del proceso y con esto echando a perder el modelo que se debería preocupar más por el futuro, que todavía no sucede, que por el pasado. Así que los modelos parsimoniosos eliminan este problema.

V.3 Pronósticos basados en el juicio personal

Lo relevante de estos pronósticos es la manera en que se obtiene la información y se extrae de sus diferentes fuentes: clientes, trabajadores de las empresas o de los directivos.

Aunque el juicio se utiliza al momento de tomar la decisión basados en esta información, nos concentramos primero en la obtención de la información y después en los problemas que se presentan comúnmente a la hora de utilizar el juicio para pronosticar.

1) OPINIÓN EJECUTIVA

Este método es el más sencillo y más utilizado. Consiste en que los ejecutivos más altos de las empresas, de áreas como finanzas, ventas o producción, se reúnen para discutir acerca de lo que consideran que sucederá en el futuro con respecto a una específica situación relacionada con la empresa.

Al utilizar ejecutivos de diferentes áreas se concentran diversos puntos de vista para nutrir así de información el proceso de la realización del pronóstico.

Una de las ventajas de este método es que es rápido y no se necesitan recabar datos ni elaborar complicados modelos estadísticos para después buscar la forma de explicárselos a los mismos ejecutivos.

Otra ventaja es que las personas que están formulando el pronóstico tienen la capacidad ellos mismos de influenciar en los resultados. Además en muchos casos este método es el único que se puede aplicar, sobre todo si no se cuentan con suficientes datos o la información no es la adecuada.

Pero la ventaja más importante, y esto va para todos los pronósticos que no son cuantitativos, es que los pronósticos de juicio son los únicos en los que se puede confiar totalmente en los casos en que estén sucediendo cambios radicales y de fondo en el comportamiento del proceso, como pueden ser en épocas de crisis, un súbito cambio de estrategia comercial o administrativo, la introducción al mercado de un nuevo producto, etc.

Una de las desventajas es que el tiempo de los ejecutivos es caro, ya que se está tratando con los de más alto rango. Además no se pueden realizar predicciones acerca de muchos productos a la vez y se presentan problemas de parcialidad propios del *groupthink*

o razonamiento de grupo, que se verá más adelante, en el tema "Problemas propios del razonamiento en grupo".

2) MÉTODOS DE FUERZA DE VENTAS

Estos métodos obtienen pronósticos de ventas, haciendo uso para esto de los gerentes y ejecutivos de ventas. Están divididos en tres que son los siguientes.

- a) El enfoque de raíz. Consiste en recolectar las estimaciones del personal encargado de ventas, es decir del personal de menor jerarquía en el área de ventas, acerca de diferentes productos. Los pronósticos se pueden obtener por medio de una entrevista personal o por medio de formas hechas para este fin. Una vez que se cuenta con las estimaciones individuales, se envían todas con los ejecutivos correspondientes para que puedan tomar la decisión final.
- b) Técnica de administración de ventas.- Funciona de manera similar a la de Opinión Ejecutiva, con la diferencia que no se utilizan los ejecutivos más altos, sino a los ejecutivos de ventas solamente. Lo que se quiere hacer con esto es minimizar la cantidad de personal que aporta sus proyecciones y maximizar la experiencia con la que cuenta el personal que las realiza. Se obtiene así un grupo reducido de personas que pueden aprender más rápido y mejor que con el método de Opinión Ejecutiva.

Una de las fallas que presenta el método es que la gente de ventas de nivel inferior, que está en contacto directo con las comercializaciones y tiene la capacidad de influenciar en los resultados, no se sentirá tan comprometida con cumplir el pronóstico realizado ya que ellos no formaron parte de la toma de decisiones.

- c) Enfoque de distribuidor.- Este método se crea a raíz que no todas las compañías venden sus productos directamente al cliente final, sino que algunas realizan sus ventas a distribuidores. Siguiendo la misma óptica que se ha seguido hasta ahora, podemos inferir de lo que se trata: se encuestan a las distribuidoras y se les pide que pronostiquen las ventas de cierto producto o productos que la compañía les provee. Para obtener mejores resultados es común que se les suministre información de realizaciones anteriores con sus respectivos pronósticos. Otras veces incluso la compañía misma les ayuda a las distribuidoras a realizar sus pronósticos, para así obtener mejores resultados propios.

Las ventajas de estos tres enfoques que acabamos de ver son que utilizan la experiencia de los empleados más cercanos al comercio real, poniendo en sus manos la responsabilidad de los pronósticos realizados, permitiéndoles influenciar en ellas y consiguiendo pronósticos específicos por zona geográfica, producto, etc.

En los casos que la empresa se encuentre en un estado de emergencia o que esté en medio de un cambio drástico sin precedencia, los métodos cuantitativos no nos sirven de mucho y se debe recurrir a los de juicio.

Las desventajas que presentan son similares a las que veremos en el siguiente método: el personal de ventas puede ser extremadamente optimista o pesimista y no tiene la visión del contexto económico en que se encuentra, lo cual puede limitar en gran parte sus estimaciones.

3) EVALUACIONES BASADAS EN ENCUESTAS E INVESTIGACIÓN DE MERCADO

Consiste de sondeos de opinión acerca de lo que la gente planea o no planea comprar de cierto producto. A diferencia de los anteriores métodos que sólo se preocupaban por la opinión de los expertos, éste se preocupa por lo que opina el cliente, ya que finalmente es el cliente el que moldeará la demanda en el mercado.

Los sondeos resultan de gran utilidad, porque aunque se cuente con datos históricos, éstos pueden resultar engañosos debido a la ciclicidad que muchas veces resulta imposible de modelar, como ya habíamos visto, incluso en el corto plazo. Algunos sondeos los realiza el gobierno y son publicados periódicamente, pero algunos son realizados por las compañías mismas ya sea por correo, teléfono o personalmente.

4) INVESTIGACIÓN DE MERCADO

El potencial de este método recae en la destreza del personal que realice el estudio de mercado, ya que si no se realiza de manera adecuada puede arrojar resultados un tanto confusos y engañosos que nos pueden hacer concluir aberraciones. Sin en cambio si se aplica conociendo y considerando las limitaciones propias de la estrategia, puede resultar una arma muy poderosa en la planeación y toma de decisiones en una compañía.

Consiste básicamente en la obtención de una amplia gama de información acerca del cliente: quién es, por qué o por qué no compra el producto, qué es lo que planea comprar y cómo es que utiliza el producto. Con esto se puede conocer y especular de una manera más confiable acerca del comportamiento de una parte del mercado y su potencial a futuro. Como se puede apreciar, es solamente una extensión del método anterior.

Un ejemplo: A mediados de los años 60 se proyectaba, de acuerdo a pronósticos que no habían sido respaldados por estudios de mercado, que el teléfono con pantalla iba a estar en pleno auge para el año de 1973, debido a que se consideraba que para esa fecha los servicios de telecomunicación y el producto mismo ya habrían bajado de precio lo suficiente para entrar en su etapa de producción a gran escala.

Se realizaron dos pruebas de mercado, una en Pittsburgh y otra en Nueva York y desgraciadamente para las personas que confiaron en el pronóstico, se observó que no se cumpliría, por lo menos no en el corto plazo, debido a que los precios eran excesivamente altos y que se necesitaban hacer ajustes a los métodos de transmisión de información para poder hacerlo más rentable. Se descubrió también que el mercado no sería el doméstico, como se creía, sino el de la telecomunicación entre empresas.

Así tenemos una clara mirada de los problemas que pueden surgir debido a los prejuicios o predeterminaciones que nos formamos todos los humanos, desgraciadamente, con gran facilidad.

5) EVALUACIONES PROBABILÍSTICAS SUBJETIVAS

Todos los métodos anteriores que hemos visto se enfocan a arrojar un solo valor estimado. Este método no busca un solo valor, sino una serie de valores posibles, cada una con su respectiva probabilidad de ocurrencia.

Por ejemplo, en cada vuelo por avión de alguna aerolínea comercial los pronósticos que se nos hacen llegar son acerca del tiempo que tardará el vuelo, pero sabemos que no es

la única posibilidad. Internamente se tiene un control meticuloso de las probabilidades de falla mecánica del motor y de cada una de las partes que forman la aeronave. Se tienen perfectamente calculadas las probabilidades de una posible colisión ya sea en el aire, aterrizando, despegando o cuando se encuentra estacionada. Así también se conocen con relativa exactitud las probabilidades que el avión sea secuestrado por su tripulación o pasajeros.

En medio de esta distribución de probabilidad con la posibilidad más alta de ocurrencia se encuentra la de "sin problema de consideración" que implica que el avión pueda despegar y aterrizar sin mayor contratiempo.

PROBLEMAS EN LOS PRONÓSTICOS DE JUICIO

A continuación veremos los problemas más serios y frecuentes con los que se han encontrado administradores, gerentes, tomadores de decisiones y cualquier grupo de trabajadores que han aplicado los métodos de juicio. Cabe mencionar que los problemas que aquí se señalan están respaldados por múltiples estudios de importancia realizados por psicólogos y estudiosos del área de la estadística aplicada a la administración, así como de las ciencias administrativas para la toma de decisiones.

El razonamiento es una de las capacidades que distinguen al ser humano del resto de los animales en este planeta. Es la actividad suprema por excelencia con la que el hombre cuenta y ha sido considerada desde siempre por los grandes pensadores como la función más elevada en este Universo.

El sesgo y prejuicio de nuestro razonamiento se encuentra en las cuatro etapas del proceso de realización del pronóstico por medio del juicio: Adquisición de información, procesamiento de la información, el arrojamiento de resultados y por último la retroalimentación, que no es más que la observación de lo sucedido y su comparación con lo que habíamos pronosticado.

1) En la adquisición de la información. La **fácil accesibilidad a los datos** nos hace más propensos a formar pronósticos basados en éstos datos y no a otros que podrían resultar de más difícil acceso pero de igual o mayor importancia que los anteriores. Los eventos recientes o a los que se les magnifica más quedan mejor grabados en nuestras mentes que los demás.

La **predisposición a un cierto resultado** nos hace ver los datos recabados en color de rosa, desechando así cualquier dato que consideramos adverso a un resultado que deseamos que suceda.

Cuando los **datos son de difícil entendimiento** las personas hacen menos caso a éstos y más a los que sean de fácil entendimiento y menos abstractos, como podría ser la opinión de un cliente (más concreta y directa) en contra de algún dato estadístico.

La **correlación ilusoria** se presenta cuando atacamos el problema con una actitud poco profesional de pensar que se puede concluir algo importante a partir de una pequeña muestra o de que se puede tener un control total del proceso realizando complicados métodos matemáticos y estadísticos. De esta manera, al estar realizando un análisis de correlación, se puede uno imaginar que dos variables están correlacionada cuando en verdad no lo están o inventar patrones de los cuales no entendemos su porqué. Pensemos por ejemplo en las veces que imaginamos animales o figuras en los contornos de las nubes.

De la misma manera jugando con la escala de las gráficas pueden aparecer y desaparecer ciertos patrones, haciéndonos llegar a conclusiones erróneas.

2) En el procesamiento de la información. La **inconsistencia** en nuestra forma de procesar la información puede ser un grave problema, ya que somos por naturaleza cambiantes. Dependiendo del humor en que estemos podemos dar increíblemente diferentes pronósticos. Está comprobado incluso que si se imita el modelo de toma de decisiones de una persona, este modelo extrapolado tiene un error mucho menor comparado con el que realice la persona.

Ser **conservador** y cerrado puede resultar un problema porque no se puede pretender mantener a una empresa al día en un mundo donde los gustos de los clientes, la ciencia y la tecnología mantienen en constante movimiento a los creadores de nuevas ideas y productos. Estos factores exigen que, por el contrario, se mantengan los empresarios unos pasos adelante de las tendencias actuales para que incluso sea la empresa misma la que imponga productos que la gente no conozca ni se haya imaginado antes pero que pueda llegar a desear.

La experiencia de muchísimos tomadores de decisiones ha confirmado que la mejor manera de seguir adelante y no desaparecer, como empresa, es mantener una mente abierta a cambios, como podrían ser realizar grandes reestructuraciones o modificaciones a la filosofía de una empresa, en casos extremos, si descubrimos que lo que teníamos simplemente no estaba funcionando.

Aunque el conservadurismo, o aferramiento a ideas "viejas" o inmediatas, es una arma de dos filos. Cuando se ha presentado en una gran parte del mercado, es decir, en los clientes, ha impulsado grandes avances de un mismo servicio o producto. Tomemos por ejemplo la creación de los automóviles con caja de velocidades automática. Si no hubiera sido por que mucha gente consideraba que manejar un automóvil estándar era demasiado difícil o "avanzado", no se hubiera inventado los coches automáticos. Otro ejemplo mucho más conocido es el del sistema operativo de computadoras personales con ventanas y que se puede operar por completo con el mouse, en vez del sistema operativo con la tétrica pantalla negra que se manejaba únicamente por comandos tecleados. El impacto en el mercado de un sistema operativo tan amigable fue tal que cambió la forma de ver la computadora incluso para los ingenieros en computación y los programadores mismos.

El **anclaje** a una cierta idea o valor puede nublar nuestra visión y angostar nuestro panorama de ideas. Se ha demostrado que cuando a un grupo de personas se le pide que pronostique con respecto a un proceso del cual existe una muestra de valores históricos y se les expone una estimación no necesariamente cierta como referencia, tienden a hacer sus estimaciones alrededor de este valor.

3) Arrojamiento de datos. La inclinación hacia un resultado deseado hace que exageremos nuestras estimaciones pesimista u optimistamente, debido a que **subestimamos la aleatoriedad del futuro**. Se debe recordar siempre que esto último es un producto del control imaginario que sentimos por el simple hecho de haber realizado un pronóstico. Para realizar pronósticos serios es nuestro deber tener una actitud humilde y tomar muy en cuenta al azar que comprende el futuro.

4) Retroalimentación. **Confiarnos de nuestros aciertos** nos puede hacer pensar que hemos encontrado una bola de cristal, perdiendo así la inercia necesitada de actualizarnos constantemente.

Lo contrario también es válido. Si el pronosticador cuenta con una racha de aciertos, es común que sienta la necesidad de prepararse para una serie de desaciertos, fundamentándose en concepciones erróneas de probabilidad. Esto se conoce como “**la falacia del apostador**”.

Inventar pretextos para pronósticos erróneos es una traba en el proceso de aprendizaje, ya que nos inhabilita a mejorar los pronósticos. Debemos primero aprender a vivir con el error por que nunca vamos a dejar de hacerlos y luego procurar entender el porqué de éstos para asimilarlos y poder así mejorar.

PROBLEMAS PROPIOS DEL RAZONAMIENTO EN GRUPO

Es un hecho que el ser humano toma decisiones diferentes cuando las toma por sí solo que cuando las toma en grupo.

Los grupos o equipos de amigos de trabajo son muy comunes en todas las instituciones de gobierno, en las empresas privadas y en el ejército, por lo que debemos entender como funcionan y aprender a trabajar dentro de ellos así como conocer cuáles son las ventajas y desventajas que presentan, para de esta manera mejorar su eficacia una vez que formemos parte de ellos.

El grupo o equipo de amigos de trabajo hace que sus integrantes cuenten con ciertas características que lo diferencian del trabajador que no forma parte de un grupo.

Los integrantes sienten un gran apego al equipo y no gustan de fricciones dentro del mismo. Son susceptibles de aceptar ideas del líder o ideas que cuenten con la aprobación general, lo cual implica que la discusión sea nula acerca del tema y por tanto el análisis de nuevas ideas ni siquiera se considere. Esto también conlleva a que con el tiempo el equipo se vaya hermetizando cada vez más y que sus integrantes filtren la información que consideran adversa o amenazante para la “ideología” del equipo o incluso aislen a los integrantes rebeldes que propongan ideas diferentes. De esta manera tenemos que la actividad pensativa del equipo es bastante baja, en vez de ser lo opuesto, como uno pensaría que debería de ser: “dos mentes funcionan mejor que una”.

Los elementos del equipo se envalentonan más en el grupo que cuando toman las decisiones por sí solos, ya que nadie realmente está aceptando las consecuencias por que la responsabilidad está distribuida.

Y por último, es otra característica del razonamiento en grupo que los integrantes se sientan moralmente correctos y que sus decisiones disten mucho de serlo.

Con esto podemos concluir que puede haber gente muy inteligente en éstos equipos y aún así realizar un pésimo trabajo pronosticando y tomando de decisiones. Claros ejemplos de casos así son las decisiones que se tomaron en la Invasión de Bahía de Cochinos, la guerra de Corea, la de Vietnam, el ataque a Pearl Harbor y la crisis energética de los 80, y más recientemente en las guerras del Golfo Pérsico: 1991 y 2003.

Para finalizar este capítulo, se presenta un estudio realizado con el fin de analizar la cantidad de uso que reciben algunos métodos, en los que se descubrió que los métodos de juicio personal siguen siendo preferidos sobre los cuantitativos**.

| Técnica | Ventas bajas* (<\$100 M) | Ventas altas (>\$500 M) |
|------------------------|--------------------------|-------------------------|
| Opinión Administrativa | 40.7% | 39.6% |
| Opinión Ejecutiva | 40.7% | 41.6% |
| Fuerza de Ventas | 29.6% | 35.4% |
| Cantidad de compañías | 27 | 48 |

* El porcentaje se refiere a la cantidad de compañías que utilizan el método

| Técnica | Ventas bajas (<\$100 M) | Ventas altas (>\$500 M) |
|---------------------------|-------------------------|-------------------------|
| Promedios móviles | 29.6% | 29.2% |
| Proyección en línea recta | 14.8% | 14.6% |
| Pronóstico ingenuo | 18.5% | 14.6% |
| Suavizamiento exponencial | 14.8% | 20.8% |
| Regresión | 22.2% | 27.1% |
| Simulación | 3.7% | 10.4% |
| Descomposición | 3.7% | 8.3% |
| Box-Jenkins | 3.7% | 6.3% |
| Cantidad de compañías | 27 | 48 |

CONCLUSIONES DE PRIMER PARTE DE LA TESIS

Para cerrar el tema de la importancia de los pronósticos y de que se haga un correcto uso de las diversas metodologías de pronósticos con series de tiempo, se cita a un investigador del Banco de México que resume muy bien la actual y a veces desesperada situación que se puede presentar al contar con investigadores con una actitud resignada ante el conocimiento que tomamos prestado de otros países para aplicarlo al nuestro propio sin hacer un análisis previo que nos mostraría que nuestro país tiene una sociedad que proviene de una cultura diferente, así como de que nuestra economía tiene una estructura completamente diferente a la de los otros países. Es de vital importancia *voltear a ver* nuestra cultura y nuestro entorno, con ojos críticos para saber qué es lo que hace falta basados en la sabiduría de saber de dónde venimos y a dónde vamos.

Es innegable que la mayoría de las metodologías existentes para analizar información se han originado en los países desarrollados, esto se debe, entre otras cosas, a que en tales países existe una gran preocupación por hacer uso eficiente de la enorme cantidad de datos que se recolectan

** Fuente: Nada Sanders and Karl Mandrodt (1994) "Practitioners Continue to Rely on Judgmental Forecasting Methods Instead of Quantitative Methods." Interfaces, vol. 24, no. 2, pp. 92-100.

continuamente. De igual manera, no hay duda de que el análisis de los aspectos económicos es de vital importancia para la buena planeación y control, así como para el alcance de las metas fijadas en los planes de los gobiernos; por este motivo, el análisis de series de tiempo económicas se ha visto favorecido con el apoyo de los gobiernos de los países ya desarrollados.

En los países en vías de desarrollo, como es el caso de México, el panorama es un tanto distinto ya que, cuando algunos tomadores de decisiones tienen conciencia de la utilidad que brinda el realizar un análisis profundo de la información económica, existe la creencia generalizada de que las metodologías utilizadas en otros países son completamente válidas también para el nuestro y que por lo tanto no se requiere diseñar una metodología específicamente adecuada a nuestra situación. Esto puede ser así en muchos casos, más no necesariamente en todos, y aunque así fuera siempre, el hecho de que existan dos o más metodologías para realizar el mismo tipo de análisis, requiere que se realice una discriminación para elegir la que más razonablemente se adapte al caso de México. En lo que respecta a la desestacionalización de series de tiempo económicas esto es particularmente cierto, ya que existe una gran variedad de metodologías que podría intentarse utilizar, pero es necesario conocer entonces los puntos favorables y desfavorables de cada una de ellas.

Fuente: www.banxico.org "DESESTACIONALIZACION DE SERIES DE TIEMPO ECONOMICAS: PARTE I UNA INTRODUCCION A LA METODOLOGIA VÍCTOR M. GUERRERO*/AGOSTO, 1983 .DOCUMENTO DE INVESTIGACION No. 54 .

**TESIS CON
FALLA DE ORIGEN**

Capítulo VI. Conceptos fundamentales

A veces sucede algo inesperado en nuestras vidas. A veces tratamos de explicarnos: ¿por qué sucedió?. Si somos gente creyente o religiosa pensamos, en el caso que se trate de algo malo "¡Ah! hoy es el primer martes que pasa desde que llegué tarde cinco minutos a misa, ¡es la última vez que llego tarde!", o si el evento inesperado es algo bueno, decimos "se me está regresando de la vez que ayudé a cruzar la calle a aquella viejita". Nuestra naturaleza es buscar patrones, quizá sin quererlo. A veces pueden ser 'modelos' que funcionan temporalmente o incluso por siempre. Imaginemos que en este mundo loco hay gente que toda su vida se la pasa pensando que si rompe un espejo tendrá que cumplir siete años de mala suerte. Supongamos que lo hace, rompe un espejo, y posiblemente no tenga más de buena suerte o suerte promedio durante los siete años siguientes, pero para ella, debido a su prejuicio que tiene, se imaginará que no tiene más que mala (el tema de prejuicios se maneja también en pronósticos). También es posible que efectivamente, por un hecho fortuito, tenga solo mala suerte. Eso sería cómico, por decirlo de alguna manera, o curioso, pero sí puede suceder, así como también sucede en el formulamiento de modelos de series de tiempo.

Como los modelos de series de tiempo, son modelos no explicativos, pueden a veces encontrar patrones que en realidad no tienen nada que ver con el proceso, pero corremos el riesgo de aceptarlo como verdadero, si éste funciona. Ése es el peligro. Los modelos pueden "funcionar" pero eso no quiere decir que en realidad estén captando el proceso. Si nos reducimos a ser muy pragmáticos nos podría bastar un modelo que *funcione*, pero tarde o temprano habrá necesidad de alterarlo. Así que si nos interesa comprender lo que está pasando, será mejor indagar hasta encontrarlo.

PROCESO ESTOCÁSTICO

Un proceso estocástico es un modelo matemático que sigue la trayectoria, en cada momento después del momento inicial, de un fenómeno aleatorio. Y por aleatorio queremos decir comportamiento no sistemático. Cada observación de un proceso estocástico es una variable aleatoria y la manera en que se comportan estas variables a través del tiempo sigue unas ciertas leyes probabilísticas.

El movimiento Browniano, usado a veces para describir los cambios de los precios de la bolsa, es un ejemplo de proceso estocástico. Una variable sujeta a movimiento Browniano geométrico tiene una distribución lognormal y tendrá siempre una media positiva. Una variable con movimiento Browniano aritmético tendrá una distribución normal. Movimiento Browniano Aritmético es también el nombre que se le da al movimiento irregular de los granos de polen suspendidos en agua. El fenómeno fue supuestamente observado por un botanista Robert Brown en 1828. Este movimiento aleatorio fue atribuido al efecto de las moléculas del agua que golpeaban al polen y los esparcían a través del agua. También llamado Proceso de Wiener o Wiener/Bachelier, el movimiento Browniano es un tipo de Proceso Markoviano*.

* Ver glosario para una definición formal de Proceso Markoviano.

Los modelos que vamos a ver lo que tratan de hacer es representar la forma en que se comportan las salidas de un proceso que es muy complejo y que por tanto lo vemos como estocástico, para así pronosticar.

Si el modelo estocástico está tratando de modelar la salida, pero no el interior del sistema, (que seguramente sí sea único aunque complejísimo) entonces no se puede decir que el modelo que representa las salidas del sistema pueda ser único, como lo es de hecho el sistema. Así que puede haber más de un modelo de proceso estocástico que se ajuste a las salidas. Es por esto que se deben hacer ciertas competencias entre modelos para ver cuál es el más preciso. Se deben seguir dos criterios:

1) El modelo que presente un error cuadrático promedio mínimo* debe ser elegido.

2) El principio de parsimonia. Éste nos dice que lo sencillo es lo mejor. Así que si tenemos un modelo que presenta iguales resultados que otro pero es más complejo (tiene mayor cantidad de parámetros a ser estimados), se debe escoger el que sea más sencillo. Además de que según este principio, el modelo más sencillo siempre funciona mejor.

El principio de parsimonia es muy evidente, porque cuanto mayor sea nuestro número de parámetros, mayor será la incertidumbre inherente a la obtención de estos, porque una de las fuentes de incertidumbre es que estamos *estimando* unos valores que *suponemos que existen y que son los reales*.

VI.1 Procesos estocásticos y series de tiempo

Existen procesos deterministas y procesos no deterministas. Los deterministas son aquellos cuyo comportamiento puede ser calculado de manera exacta. Por ejemplo el lanzamiento de una pelota en el espacio. Los no deterministas son los que pueden ser calculados con una cierta probabilidad de acierto o exactitud.

Si se trata de lanzar una pelota en tierra firme, factores como el viento o el efecto de la gravedad hacen del proceso uno estocástico, como también se les conoce a los procesos no deterministas. En realidad no existe un proceso que sea completamente determinista, ya que existen siempre factores que se nos escapan o que dejamos que se nos escapen, ya que de tratar de tomarlos en cuenta aumentarían la complejidad de nuestros cálculos al punto de hacerlos inmanejables.

Si pensamos en el ejercicio de la pelota en el espacio, siempre existen fuerzas que no se están considerando (viento solar, reminiscencias de las fuerzas gravitatorias de estrellas y planetas, etc.) que, aunque generan desviaciones, se sobrellevan por ser consideradas insignificantes o tolerables.

Así tenemos que los procesos deterministas pueden ser predichos de una manera exacta con una función o funciones matemáticas, donde las variables de las funciones pueden tomar uno y sólo un valor. Pero, vamos, realmente ¿qué proceso es completamente determinista?

Los procesos estocásticos utilizan variables aleatorias que pueden tomar **una serie de valores** en un determinado momento en el tiempo.

* Ver el tema "Error de pronóstico a una etapa" en el capítulo 2.

SERIES DE TIEMPO

Ya sabemos lo que es una serie de tiempo, un conjunto de valores observados de algo, de un proceso o como le queramos llamar. También se puede considerar una serie de observaciones irrepetibles de un proceso estocástico, también llamadas realizaciones.

El enfoque que estaremos tomando es precisamente el de procesos estocásticos, y estaremos dando por sentado que todas las series de tiempo están siendo generadas por un proceso estocástico, de manera que cada observación es una de tantas que podría haberse generado. Así que no es que las series de tiempo *sean* procesos estocásticos, sino que las vamos a tratar como si fueran.

Para poder analizar las series de tiempo es necesario manejar ciertas herramientas estadísticas, que se verán a continuación. Hablar de estas herramientas sin la computadora es inútil ya que involucran muchos cálculos. Éstas son las que nos ayudarán a proporcionarnos pistas de cómo es en realidad el proceso subyacente detrás de la serie de tiempo.

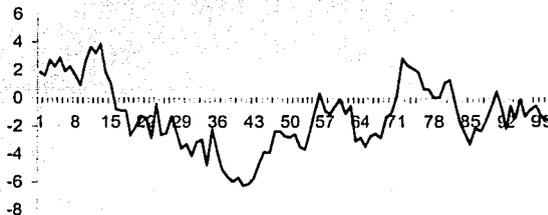
VI.2 Función de Autocorrelación (ACF)

De estadística básica sabemos que el coeficiente de correlación mide el nivel de relación entre dos variables. El coeficiente de autocorrelación ρ_k es el que se usa en series de tiempo y mide el nivel de relación que existe entre diferentes retrasos o rezagos en el tiempo de la misma serie.

Por ejemplo, ρ_1 mide la correlación entre y_t y y_{t-1} , la correlación entre y_{t-1} y y_{t-2} , la de y_{t-2} y y_{t-3} y así hasta abarcar toda la serie. La ρ_2 mide el nivel de correlación entre y_t y y_{t-2} , y_{t-1} y y_{t-3} , etc. El rezago más grande que puede haber es el que va de y_t a y_{t-m} donde m es el tamaño de la serie, así que la k en ρ_k puede ir desde 0 hasta m .

Ahora, ¿cómo es que mide éste coeficiente?, pues con un estimador, porque las series de tiempo no son más que muestras.

El coeficiente de autocorrelación va de -1 a 1 , al igual que el de correlación, lo cual lo hace muy cómodo y fácil de interpretar. Si por ejemplo $\hat{\rho}_1=0.9$ y todas las demás ρ_k son cero, quiere decir que existe una interacción muy grande entre lo que sucede en periodos adyacentes. Por ejemplo:



TESIS CON
FALLA DE ORIGEN

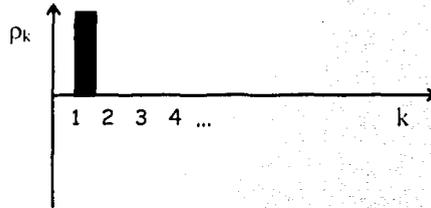
Si, en cambio, $\rho_6=0.9$ y las demás $\rho_k=0$, la influencia que sucede en un periodo con la que sucede seis periodos después es muy grande. Veamos:



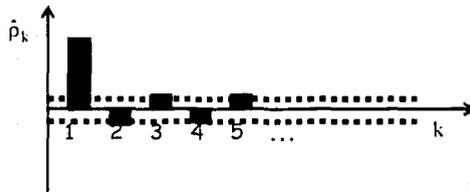
Este comportamiento podría ser el de tránsito de automóviles en carreteras, por ejemplo, en el caso de que los datos fueran mensuales, porque la gente viaja más en vacaciones de verano y en las de navidad.

Se acostumbra utilizar la gráfica de los coeficientes de autocorrelación, también llamado correlograma y por lo general se prefiere éste a los coeficientes expresados numéricamente.

TRABAJOS CON
FOLETA DE ORIGEN



Nótese que las ordenadas está etiquetadas con un ρ_k y no con un $\hat{\rho}_k$. Lo que sucede es que en los casos reales nunca vamos a encontrar unos ceros absolutos, como los teóricos, Si en la realidad vamos a usar muestras, lo que vamos a encontrar son ceros estadísticos.



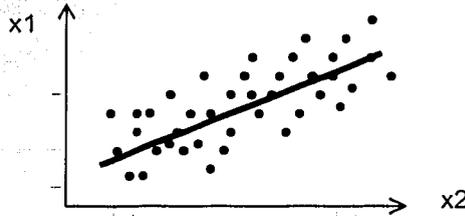
¿Cuál es el valor de ρ_0 para cualquier serie?. Dado que cualquier serie está perfectamente correlacionada consigo misma, siempre ρ_0 será igual a 1, así que éste no se acostumbra graficar, por eso el correlograma empieza en $k=1$.

Pero la trascendencia de este hecho es mayor y nos ayuda a entender de dónde se deriva el ρ_k : En estadística existe la función de covarianza, que no es más que la esperanza de los productos de dos variables. Nos ayuda a exactamente lo mismo que la correlación, sólo que la varianza está en las unidades de las variables y no está entre -1 y 1 como el de correlación.

COVARIANZA

$$\text{cov}(x,y)=E[(x-\mu)(y-\mu)]$$

Los diagramas de dispersión nos ayudan a visualizar, el nivel y el tipo (positiva, negativa o nula) de relación entre dos variables.

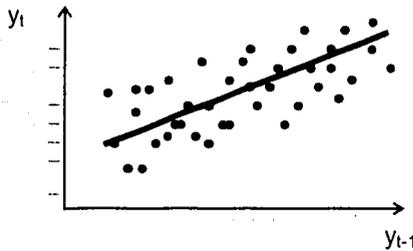


En la anterior ilustración se puede ver que existe una relación positiva entre lo que hace x_1 y lo que hace x_2 . (El término hace se utiliza porque estamos hablando de variables aleatorias, y como aparentemente las variables aleatorias más o menos hacen lo que quieren, se ha considerado propio utilizar este término.)

Si en vez de poner x_1 y x_2 , sustituimos con y_t y y_{t-1} obtenemos una gráfica de dispersión pero para la autocovarianza con $k=1$.

AUTOCOVARIANZA DE ORDEN K

La autocovarianza es la covarianza entre valores de una serie atrasados en k periodos de tiempo. Si la covarianza sirve para lo mismo que la correlación, entonces esperaríamos que la función de autocovarianza indique lo mismo que la función de autocorrelación. Si reetiquetamos la anterior gráfica como mencionamos, en vez de x_1 y x_2 escribimos y_t y y_{t-1} , a simple vista calcularíamos que ρ_k sea bastante alto, como de 0.9, porque si y_t sube y_{t-1} trata de subir, si aquella baja ésta también trata de hacerlo.

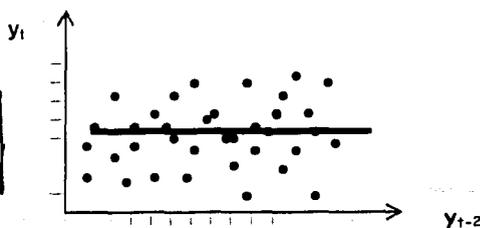


Recordemos otra vez que estamos hablando de procesos estocásticos y variables aleatorias. Obsérvese cómo para cualquier valor de y_{t-1} , le corresponden, no uno, sino varios valores de y_t .

Siguiendo el ejemplo que teníamos con $\rho_2=0$, al ajustar una curva en el diagrama de dispersión entre y_t y y_{t-2} , debemos obtener una línea horizontal como efectivamente sucede.

TESIS CON FALLA DE ORIGEN

TESIS CON
FOLIA DE ORIGEN



No olvidemos que estamos tratando de comprobar que $\rho_0=1$ para cualquier serie. Ya con las explicaciones que acabamos de ver es más fácil de hacer.

La fórmula de la covarianza está dada por:

$$\text{cov}(x,y)=E[(x-\mu)(y-\mu)]$$

La autocovarianza γ_k de una serie de tiempo se obtiene con:

$$\gamma_k = \text{autocov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$$

Si obtenemos la autocovarianza para $k=0$ obtenemos la varianza del proceso:

$$\gamma_0 = \text{cov}(y_t, y_t) = E[(y_t - \mu)(y_t - \mu)] = \sigma^2 = \text{var}(y_t)$$

La función de correlación es:

$$\text{corr}(x,y) = \frac{\text{cov}(x,y)}{\delta_x \delta_y}$$

Análogamente, la de autocorrelación:

$$\rho_k = \text{corr}(y_t, y_{t-k}) = \frac{\text{cov}(y_t, y_{t-k})}{\gamma_{y_t} \gamma_{y_{t-k}}} = \frac{\gamma_k}{\gamma_0}$$

Así para $k=0$:

$$\rho_0 = \text{corr}(y_t, y_t) = \frac{\text{cov}(y_t, y_t)}{\delta_{y_t} \delta_{y_t}} = \frac{\text{var}(y_t)}{\delta^2_{y_t}} = \frac{\gamma_0}{\gamma_0}$$

Con lo que queda comprobado el porqué del $\rho_0=1$.

VI.3 Función de Autocorrelación Parcial (PACF)

Con la autocorrelación parcial las cosas son diferentes. El coeficiente de correlación parcial que utilizamos en estadística mide el nivel de correlación que existe entre dos variables sin tomar en cuenta la influencia de otras variables.

Por ejemplo, si estamos haciendo una regresión de Y sobre x_1 , x_2 y x_3 , la correlación parcial entre Y y x_1 es la correlación entre estas dos que no está explicada a través de la correlación mutua que comparten con x_2 y x_3 .

De hecho, el coeficiente de correlación parcial se puede calcular obteniendo la raíz cuadrada de la reducción de la varianza obtenida al añadir x_1 a la regresión de Y sobre x_2 y x_3 .

Para dos valores de una serie que se encuentran a k unidades de tiempo de distancia, la autocorrelación parcial ρ_{kk} es la correlación entre estos dos valores que no es consecuencia de la propagación de los efectos de las autocorrelaciones de orden menor (de menor k).

Por ejemplo, si y_t y y_{t-1} están considerablemente correlacionadas, entonces y_{t-1} y y_{t-2} también lo estarán, ya que para ambas $k=1$. Por consiguiente y_t y y_{t-2} deben estar también correlacionadas porque ambas están interactuando con y_{t-1} . El coeficiente de correlación parcial es la correlación entre y_t y y_{t-2} menos los efectos de la correlación con $k=1$.

Al igual que la ACF, la PACF también se acostumbra graficar a partir de $k=1$.

VI.4 Periodograma

El periodograma está dado por el siguiente conjunto de fórmulas.

$$a_k = \begin{cases} \frac{1}{N} \sum_{i=1}^N y_i \cos\left(\frac{2tk\pi}{N}\right) & k = 0 \text{ y } k = \frac{N}{2} \text{ si } N \text{ par} \\ \frac{2}{N} \sum_{i=1}^N y_i \cos\left(\frac{2tk\pi}{N}\right) & k = 1, 2, \dots, \left[\frac{N-1}{2}\right] \end{cases}$$

$$b_k = \frac{2}{N} \sum_{i=1}^N y_i \sin\left(\frac{2tk\pi}{N}\right) \quad k = 1, 2, \dots, \left[\frac{N-1}{2}\right]$$

$$I(f_k) = \begin{cases} Na_0^2 & k = 0 \\ \frac{N}{2} (a_k^2 + b_k^2) & k = 1, \dots, \left[\frac{N-1}{2}\right] \\ Na_{\frac{N}{2}}^2 & k = \frac{N}{2} \text{ con } N \text{ par} \end{cases}$$

Resulta útil sobre todo para encontrar comportamientos estacionales. Correlaciona los valores de la serie con las funciones seno y coseno de diferentes frecuencias. Por ejemplo una serie con patrón estacional de período 12 o frecuencia $1/12 = 0.08$ registrará valores grandes para $(1/12)$, $(1/36)$, etc. Como resultado el periodograma mostrará saltos en estas frecuencias.

VI.5 Procesos de ruido blanco

Como se podrá recordar en la primera parte de la tesis (tema 1.3 "Series de Tiempo") se mencionó que era posible describir el comportamiento del error e_t , pero la verdad es que nunca lo hicimos.

Cuando estamos creando un modelo, obviamente tenemos que saber lo que estamos haciendo. no podemos simplemente decir "Mi modelo es éste. más (o por) un golpe

aleatorio, que no sé de cuanto es", como lo hicimos en la descomposición multiplicativa y aditiva. ¿Que tal si decimos que es de diez mil, o mejor de un millón?. ¿De que serviría este modelo? Se nos ocurrieron estos números al azar, pero se nos podrían ocurrir otros peores. Así que existe consenso general en que de entrada este e_t debe tener un cierto límite o rango. Luego ¿cómo se debe comportar dentro de ese rango? ¿a poco se comporta como una distribución uniforme, donde todos los valores dentro del rango tienen la misma probabilidad de ocurrencia?. Si estamos diciendo que el modelo que estamos armando es real y verdaderamente la 'ecuación' que está generando el proceso, esperaríamos que e_t fuera *de preferencia* cero. Pero nos resignamos y aceptamos que no vamos a acertar en todas, así que no siempre va a ser cero, pero va a *tender* a serlo.

La distribución normal tiene precisamente esta propiedad, donde los valores tienden a la media y cuanto más se alejan de ella, lo hacen con menos frecuencia.

Cuando e_t sigue una distribución normal con media cero y varianza constante, donde cada e_t es independiente de las demás e_t y además no está correlacionada en el tiempo, se dice que e_t es **ruido blanco gaussiano o ruido blanco normal**.

$$\begin{aligned} e_t &\sim N(0, \sigma^2) \\ \gamma_k &= E[e_t e_{t+k}] = 0 \text{ para } k \neq 0 \\ E(e_t) &= 0 \\ \text{var}(e_t) &= \sigma^2 \end{aligned}$$

y se dice que e_t es una secuencia de v.a.i.i.d. (variables aleatorias independientes idénticamente distribuidas)* con media cero y varianza σ^2 . Aunque existen otros tipos de ruido blanco, en adelante cuando se hable de ruido blanco se estará hablando de éste en particular.

El proceso de ruido blanco será de gran utilidad a través del desarrollo de los modelos, porque los residuales, es decir, el error del modelo, al medirse con la muestra (los errores de pronóstico muestrales) deben comportarse como ruido blanco, indicando que hicimos lo mejor que pudimos y que encontramos y modelamos cualquier rastro de correlación que hubiera en nuestros datos.

Suponer que los errores que se cometan al tratar de obtener un modelo van a ser Normales puede resultar ser una suposición peligrosa. Pero existe teoría que respalda a esta práctica. Se trata del Teorema Central Límite, que dice que la media de cualquier conjunto de variables de cualquier distribución con una varianza y media constantes, tiene una distribución Normal.

Podemos concluir también que la e_t se agrega al modelo para tener algo que el algoritmo de encontrar modelos tenga algo que minimizar.

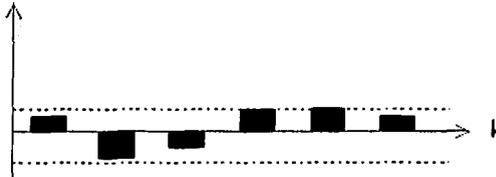
ACF Y PACF DE RUIDO BLANCO

Si calculamos la ACF y la PACF del modelo de ruido blanco

* No debemos confundir no existencia de autocorrelación con la independencia. Se dice que un proceso es independiente cuando la distribución conjunta de las variables que componen el proceso se puede obtener como producto de las distribuciones marginales. De forma intuitiva, la independencia implica la no existencia de ningún tipo de relación entre los valores que toma las variables del proceso en dos momentos del tiempo diferentes.

$$y_t = e_t$$

se debería ver así



media CON
FALLA DE ORIGEN

¿Por qué se ve así? porque no existe ninguna relación entre lo que sucede cada k períodos fijos de tiempo. aunque el estimador si encuentra "un poco" de relación, ésta no es significativa y es solamente circunstancial.

VI.6 Estimador del coeficiente de media, ACF y PACF.

El estimador de la **media** es $\hat{\mu}_t = \sum_{i=1}^T \frac{y_i}{T}$ donde T es el tamaño de la serie (muestra).

El estimador de la **ACF** es:

$$\hat{\rho}_k = \frac{\sum (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum (y_t - \bar{y})^2}$$

Donde

- y_t = datos de la serie
- y_{t+k} = valores k períodos adelante
- \bar{y} = media de la serie

El estimador de la **PACF** es:

$$r_{kk} = \frac{r_k - \sum (r_{k-l,j})r_{k-l}}{1 - \sum (r_{k-l,j})r_{k-l}}$$

donde

- r_k = coeficiente de autocorrelación para k unidades de tiempo de distancia.
- r_{k_l} = coeficiente de autocorrelación parcial para k unidades de tiempo de distancia con los efectos de los j rezagos intermedios ha sido retirado.

VI.7 Procesos lineales generales

La base de los modelos ARIMA que veremos en el siguiente tema es el Proceso Lineal General. El enfoque de PLG asevera que cualquier proceso estocástico puede ser explicado a través de sus choques aleatorios en el pasado.

El PLG lo que hace es tomar un conjunto de variables aleatorias e_t con distribución normal, media cero y varianza sigma cuadrada, es decir, ruido blanco, y formar una suma ponderada multiplicada por unos factores ψ .

PROCESO LINEAL GENERAL

$$y_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \psi_3 e_{t-3} + \dots \quad (6.1)$$

$$= \mu + \Psi(B)e_t \quad (6.2)$$

donde $\Psi(B) = \psi + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots$ es el operador lineal que transforma el ruido blanco en y_t . e_t es ruido blanco $WN(0, \sigma^2)$ y μ es la constante que indica el recorrimiento vertical de la serie.

La forma (6.2) está utilizando lo que también se denomina **polinomio de rezagos**.

Omitiendo el polinomio, el PLG también se representa:

$$y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$$

donde $\psi_0=1$. Ahora, la secuencia de operadores $\psi_0, \psi_1, \psi_2, \dots$ debe ser finita; o infinita y absolutamente sumable, es decir, **convergente**, en el sentido que

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

para que el proceso pueda tener un cierto equilibrio estadístico. Cuando se cuenta con este equilibrio, se dice que el proceso es **estacionario**.

Dado que las desviaciones con respecto a la media μ en un proceso estacionario son 0, por construcción, se puede eliminar este término para simplificar la nomenclatura. Suponer una media cero no resta generalidad y es un artificio comúnmente usado.

PLG ESTACIONARIO

$$y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} \cdot \psi_0=1$$

donde

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

Capítulo VII. Modelos estacionarios

Un ejemplo muy sencillo de modelo de un proceso estacionario es el de promedio móvil de primer orden.

$$y_t = e_t + \theta e_{t-1} \quad e_t \sim \text{WN}(0, \sigma^2)$$

A una determinada secuencia de e_t , $t=0,1,2,3,\dots,T$ le corresponde una determinada secuencia de observaciones y_t , $t=1,2,3,\dots,T$. Si tomamos otra secuencia diferente de e_t , obtenemos otra secuencia diferente de y_t . Si pudiéramos tomar un conjunto infinito de secuencias podríamos ver la forma de la distribución.

Con la distribución en mano es inmediato calcular la media, varianza y covarianza para cada tiempo t .

$$\mu_t = E(y_t) \quad t = 1,2,3,\dots,T \quad (7.1)$$

$$\text{var}(y_t) = E[(y_t - \mu_t)^2] \quad (7.2)$$

$$\text{cov}(y_t, y_{t-k}) = E[(y_t - \mu_t)(y_{t-k} - \mu_t)] \quad (7.3)$$

Si fuera posible tener varias observaciones del proceso para un determinado tiempo t , las anteriores cantidades 7.1, 7.2, 7.3 se podrían obtener con estimadores como el siguiente, que corresponde al de la media:

$$\mu_t = \frac{1}{m} \sum_{k=1}^m y_t^{(k)} \quad t = 1,2,3,\dots,T$$

donde m es el número de observaciones en el tiempo t . Nótese que no se puede inferir gran cosa a partir de una sola observación para cada tiempo t con los estimadores (7.1), (7.2) y (7.3).

Debido a que contamos, en cada tiempo t , con una muestra de tamaño uno de una población infinita, es necesario cambiar de enfoque.

ESTACIONARIDAD

En una serie de tiempo contamos con datos a través del tiempo, a lo largo. Pero a lo ancho sólo contamos con un solo valor. Si quisiéramos hacerlo más ancho, estaríamos condenados al fracaso, porque no se puede regresar en el tiempo y repetir la observación. Si una serie de tiempo sí puede incrementar su tamaño a lo largo entonces más vale aprovechar esta ventaja.

La única traba con la que nos topamos para poder utilizar de este enfoque es que el proceso estocástico, que presuntamente está generando la serie de tiempo, se tiene que someter a ciertas restricciones. Cuando el proceso cumple con estas restricciones se dice que el proceso es **estacionario**. Las restricciones son las siguientes:

$$E(y_t) = \mu \quad (7.4)$$

$$E[(y_t - \mu)^2] = \sigma^2 = \gamma_0 \quad (7.5)$$

$$E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \quad (7.6)$$

Por tanto, cualquier proceso real que satisfaga los tres requisitos anteriores se puede representar con el proceso lineal general estacionario.

¿Pero entonces, por qué establecemos estas condiciones si ya teníamos las condiciones de estacionaridad con el PLG estacionario, a saber, la de la suma convergente*?

Si tomamos una serie de tiempo y le tratamos de ver los $e_{i,j}$ no los vamos a encontrar, ¡imposible!. Pero en cambio, ¿se puede calcular la media, la varianza o la covarianza de una serie de datos? ¡en dos patadas!

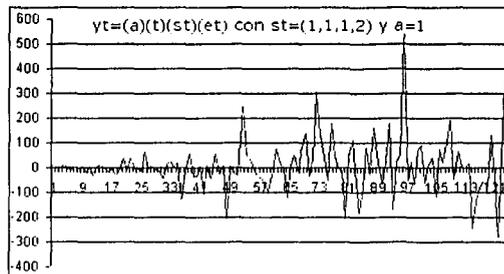
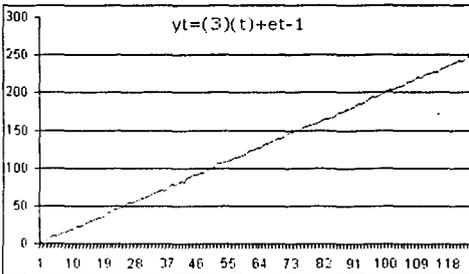
Tomemos otro ejemplo,

$$y_t = (a)(t) + e_{t-1} \quad e_t \sim WN(0, \sigma^2)$$

donde a es una constante y t es el tiempo (1,2,...). Este proceso claramente no puede ser estacionario, porque su media depende del tiempo, lo que la hace forzosamente cambiante. El siguiente es otro proceso no estacionario

$$y_t = (a)(t)(s_t)(e_t) \quad e_t \sim WN(0, \sigma^2)$$

donde s_t es el componente de estacionalidad. Éste con mucha más razón no va a poder ser estacionario, porque los efectos se están multiplicando y por tanto todos aumentan con el paso del tiempo. ¿Conclusión? Las fórmulas de los procesos de covarianza estacionaria no pueden contener al tiempo t como variable, porque ésta está aumentando constantemente lo cual implica una tendencia creciente. Véanse las siguientes gráficas como ejemplos de estos dos casos.



ANÁLISIS DE LAS CONDICIONES DE ESTACIONARIDAD

La primera condición (7.4) ¿qué implica?. Una constante como media implica que no puede haber tendencia. La segunda implica dos cosas. Por un lado que la varianza debe ser la misma para todo tiempo t y por otro lado si la varianza es una constante, implica que ésta debe ser finita. A la propiedad de tener la varianza constante se le llama **homoscedasticidad**.

La tercera es un poco más compleja. Quiere decir que la autocovarianza debe depender solamente de la separación k entre valores comparados y no del tiempo. Imaginemos que tenemos una serie y_t , que representa no importa qué proceso. Luego imaginemos que tenemos otra serie, llamémosle convenientemente y_{t+3} , pero no se preste atención al subíndice por un momento, solamente confórmese con saber que estamos hablando de dos series completamente diferentes. A continuación encontremos la

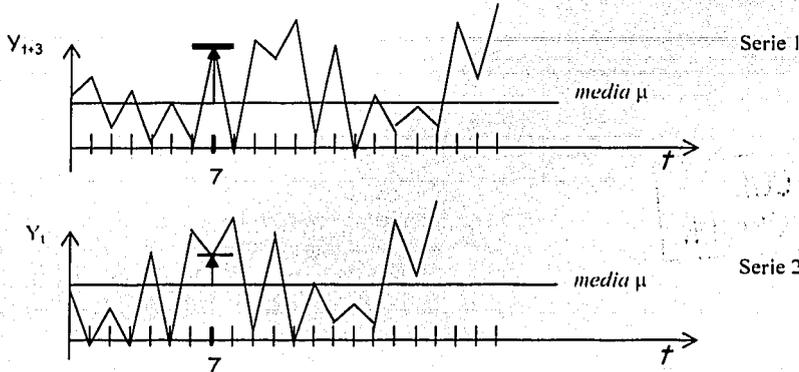
* Ver tema VI.7 "Procesos Lineales Generales" para ver a qué se refiere con "suma convergente".

covarianza entre estas dos series. Es en realidad muy sencillo. Lo único que hace el operador covarianza es responder a las preguntas:

Serie 1: "Yo varié tal cantidad 'a' con respecto mi media en el momento t , y tú?"

Serie 2: "Yo varié esta otra cantidad 'b' "

Entonces el producto de a y b es la **covarianza** entre estas dos series. Gráficamente:



En caso de que no se note patentemente, la primera serie es la misma que la segunda, con la diferencia que está *recorrida* tres unidades de tiempo hacia adelante. El fin de utilizar la misma serie de tiempo es para que se vea de manera más clara que la autocovarianza γ_k es la covarianza de una serie de tiempo con respecto de sí misma recorrida k etapas. Para las gráficas de arriba la k sería igual a tres.

Decíamos que la autocovarianza está dada por:

$$\gamma_k = \text{cov} [y_t, y_{t+k}] = E [(y_t - \mu)(y_{t+k} - \mu)] \quad (7.7)$$

Una covarianza constante garantiza una varianza constante, lo que a su vez garantiza que los valores de la serie no se van a poder alejar de la media. Si lo hacen, tiene que regresar. Así de sencillo.

Tomemos $\gamma_2 = 4$, por ejemplo. Si estamos viendo en la ecuación (7.7) que la autocovarianza es un producto, entonces si en cierto momento t la serie se alejó de la media 8 unidades (las unidades de la serie de tiempo), dos períodos de tiempo más tarde la serie tiene que valer 0.5 unidades, para poder satisfacer el $\gamma_2 = 4$.

Las tres restricciones, gráficamente, se resumen en que la serie de tiempo se mantiene horizontal y los datos no se alejan mucho de la media.

Para comprobar que efectivamente una serie de tiempo está siendo generada por un proceso estacionario, además de que la serie parezca horizontal, las gráficas de los ACF y PACF deben decrecer constante y rápidamente conforme k aumenta, con cortes bruscos o sin ellos. Sólo deberán mostrarse significativos los 4 o 5 primeros coeficientes, ya que en la realidad sólo existen procesos estacionarios con estas características.

TESIS CON
FALLA DE ORIGEN



Aunque aparenta ser un modelo muy sencillo, es de hecho el pilar fundamental de los modelos ARIMA que veremos a continuación. Los modelos ARIMA (Procesos Autorregresivos Integrados y de Medias Móviles) son simplemente ruido blanco con componentes de ciclicidad, o en otras palabras, ruido blanco *correlacionado*. Los modelos ARIMA están formados por una parte autorregresiva, una parte de medias móviles y pueden estar integradas o no. Se comienza por explicar la parte autorregresiva.

VII.1 Procesos Autorregresivos (AR)

Una de las aproximaciones al PLG estacionario (en adelante simplemente PLG) es el de la de suma infinita. Se dijo que el PLG podía tener:

- 1) Una cantidad infinita de términos con la condición de que la suma de los factores de cada término fuera una constante o
- 2) Un número finito de términos.

Los modelos AR toman el primero de los caminos. Un AR(p) se aproxima con p términos al PLG.

$$y_t = \sum_{k=1}^p \phi_k y_{t-k} + e_t$$

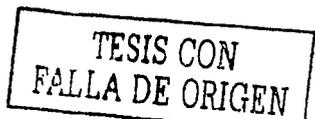
En apariencia no se parece en nada a la forma del PLG, porque la suma está compuesta por términos *finitos* no de infinitos, y de *valores* históricos no de *choques*. Pero desmenucémoslo y se verá que efectivamente sí representa al PLG.

$$\text{AR(1)} \quad y_t = \phi y_{t-1} + e_t \quad e_t \sim \text{WN}(0, \sigma^2) \quad (7.9)$$

Se dice AR(1) porque sólo contiene un término y_t . El otro término siempre forma parte de cualquier modelo ARIMA. Podemos ver que sencillamente lo que hace es tomar el valor observado anterior de la serie, lo multiplica por un ϕ y le añade el golpe aleatorio.

Si el valor de ϕ está más cercano al uno que al cero, el proceso estará haciendo más caso al valor que traía anteriormente, o sea maximizando el valor de y_{t-1} , que dijimos que podía tomar valores muy grandes. Sin en cambio, si ϕ tiene un valor pequeño, estará minimizando los efectos de y_{t-1} y será más propenso a los caprichos de los golpes aleatorios e_t .

Los valores colindantes de un AR(1) tienden a estar muy próximos sobre todo si el coeficiente ϕ es cercano a uno. Un AR(1) se parece mucho a la fórmula del suavizamiento exponencial, solo que el factor de ponderación lo tiene el pronóstico en vez del error.



Si le aplicamos el operador de rezago* varias veces, obtenemos

$$B[y_t] = y_{t-1} = \varphi y_{t-2} + e_{t-1}$$

$$B[y_{t-1}] = y_{t-2} = \varphi y_{t-3} + e_{t-2}$$

$$B[y_{t-2}] = y_{t-3} = \varphi y_{t-4} + e_{t-3}$$

Y al sustituir en 7.9 queda:

$$y_t = e_t + \varphi e_{t-1} + \varphi^2 e_{t-2} + \varphi^3 e_{t-3} + \varphi^4 e_{t-4} + \dots \quad (7.10)$$

Ahora sí, esto es idéntico al PLG, con la única diferencia que es el mismo factor φ en vez de ser una infinita cantidad de diferentes factores. Nótese que utilizamos φ en vez de ψ para diferenciar de cuando se está hablando de un PLG. No existe ninguna restricción que diga que un factor ψ_n no pueda ser el cuadrado de ψ_{n-2} .

Dado que se trata de una suma infinita, es necesario que el factor φ sea menor a uno para que y_t no se vuelva infinitamente grande. Esto es uno de los requerimientos para poder representar un PLG. que destaca la preponderancia de los valores más recientes sobre los más antiguos.

CONDICIÓN DE CONVERGENCIA PARA PROCESOS AR (1)

$$|\varphi| < 1$$

PROCESOS AR (P)

Un proceso AR(p) está denotado como:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + e_t$$

Para obtener el polinomio de rezagos es necesario despejar e_t :

$$y_t = [\varphi_1 B + \varphi_2 B^2 + \dots + \varphi_p B^p] y_t + e_t$$

$$y_t - [\varphi_1 B + \varphi_2 B^2 + \dots + \varphi_p B^p] y_t = e_t$$

$$y_t \{ 1 - [\varphi_1 B + \varphi_2 B^2 + \dots + \varphi_p B^p] \} = e_t$$

$$y_t [1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p] = e_t$$

Así que el polinomio de rezagos está dado por $[1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p]$ que llamaremos ϕ .

CONDICIONES DE CONVERGENCIA PARA PROCESOS AR(P)

Todas las inversas de las raíces del polinomio de rezagos deben estar dentro del círculo unitario, lo que equivale a decir que las raíces deben estar fuera del círculo unitario.

ACF DE UN PROCESO AR

Para obtener la Función de Autocorrelación de un proceso AR hacemos primero:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + \dots + y_{t-p} + e_t$$

* Ver capítulo primero en el tema 1.3 donde se explica qué es el operador rezago y cómo funciona

que no es más que la generalización de los procesos AR(p). Segundo, multiplicamos ambos lados por y_{t-k} :

$$y_{t-k} y_t = \varphi_1 y_{t-k} y_{t-1} + \varphi_2 y_{t-k} y_{t-2} + \varphi_3 y_{t-k} y_{t-3} + \dots + y_{t-k} y_{t-p} + y_{t-k} e_t$$

Obtenemos la esperanza matemática de ambos lados:

$$E[y_{t-k} y_t] = \varphi_1 E[y_{t-k} y_{t-1}] + \varphi_2 E[y_{t-k} y_{t-2}] + \varphi_3 E[y_{t-k} y_{t-3}] + \dots + E[y_{t-k} y_{t-p}] + E[y_{t-k} e_t]$$

Y si recordamos, la covarianza es la esperanza de los productos menos la media, (en nuestro caso estamos suponiendo una media igual a cero) así que:

$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \varphi_3 \gamma_{k-3} + \dots + \gamma_{k-p} + E[y_{t-k} e_t]$$

Y dividimos ambos lados por la varianza del proceso, γ_0 :

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \varphi_3 \rho_{k-3} + \dots + \rho_{k-p} + \frac{E[y_{t-k} e_t]}{\gamma_0}$$

La esperanza del último término sabemos que vale cero para $k > 0$, lo que nos es suficiente, porque sabemos que $\rho_0 = 1$.

Y concluimos:

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \varphi_3 \rho_{k-3} + \dots + \rho_{k-p} \quad \text{Para } k > 0 \quad (7.11)$$

La ACF de un proceso AR(1) está dada por:

$$\rho_k = \varphi_1 \rho_{k-1} \quad \text{Para } k > 0$$

Para poder inicializar la función de autocorrelación debemos conocer los primeros p valores. Conocemos que $\rho_0 = 1$, así que

$$\rho_1 = \varphi_1 \rho_0$$

$$\rho_1 = \varphi_1 (1)$$

$$\rho_2 = \varphi_1 (\varphi_1) = \varphi_1^2$$

$$\rho_3 = \varphi_1^2 (\varphi_1) = \varphi_1^3$$

$$\rho_4 = \varphi_1^3 (\varphi_1) = \varphi_1^4$$

...

$$\rho_k = \varphi_1^k$$

Se puede apreciar que los coeficientes de autocorrelación de k -ésimo orden dependen de sus coeficientes anteriores. Esto da cabida a que se pueda calcular la función de autocorrelación de un proceso AR de manera mucho más fácil, por medio de una función recursiva, llamada de Yule-Walker.

Esta función sería inservible si no se contara con el coeficiente de autocovarianza para $k=0$, pero éste es precisamente la varianza del proceso, que ya conocemos:

$$\gamma_0 = \delta^2 = \frac{\sigma^2}{1 - \varphi^2}$$

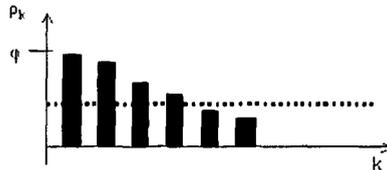
Con lo que se puede obtener recursivamente la ACF con la ecuación de Yule-Walker

ECUACIÓN DE YULE-WALKER

$$\gamma_k = \phi \gamma_{k-1}$$

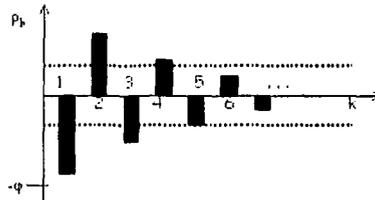
Sólo hace falta dividir cada autocovarianza entre la varianza para obtener la autocorrelación. Podemos ver que la función decrece exponencialmente. Si ϕ es positivo tendremos que la gráfica se debe ver como el siguiente ejemplo:

$$y_t = 0.9 y_{t-1} + e_t$$



Si ϕ es negativo se alternan los signos de las ρ_k

$$y_t = -0.8 y_{t-1} + e_t$$



TEMA CON
 FOLLA DE ORIGEN

Para un AR(2) las cosas cambian un poco. Haciendo uso de la ecuación (7.11) la ACF de un AR(2) queda:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

Ahora necesitamos los p valores iniciales. $\rho_0 = 1$ y es posible comprobar (Wei, 1994) que

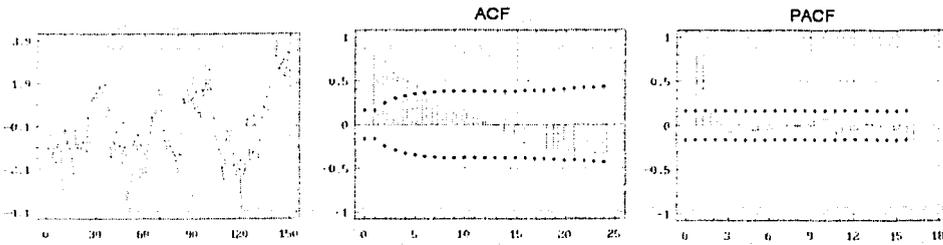
$$\rho_k = \frac{\phi_1}{1 - \phi_2}$$

con lo que se puede generar la ACF de un AR(2)

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

Ejemplo:

$$y_t = 0.6 y_{t-1} + 0.3 y_{t-2} + e_t$$



PACF

Si se trata de un modelo AR(1), la PACF lo indicará con un valor prominente en $k=1$ y los demás ($k>1$) valdrán cero. De hecho, el valor de la PACF con $k=1$ será precisamente el valor del coeficiente ϕ estimado, como sucede en la ACF.

Si se trata de un AR(2), los PACF sobresaldrán en $k=1$ y $k=2$. Si se trata de un AR(3) los PACF tendrán valores prominentes en $k=1, 2$ y 3 , y así sucesivamente.

FUNCIÓN DE AUTOCORRELACIÓN PARCIAL DE UN PROCESO AR(1)

$$r_{kk} = \begin{cases} \phi, & k = 1 \\ 0, & k > 1 \end{cases}$$

VII.2 Procesos de Medias Móviles (MA)

La otra aproximación al PLG lo toman los modelos MA (*Moving Averages* en Inglés). Un MA(q) se aproxima con q términos al PLG.

$$y_t = \sum_{k=1}^q \theta_k e_{t-k} + e_t$$

Empecemos por el modelo más sencillo, cuando $q=1$

MA (1)

$$y_t = e_t - \theta e_{t-1}$$



Si observamos con atención, este modelo está formando su predicción (θe_{t-1}) pero después está aceptando que no va a acertar completamente, así que introduce el e_t . Pero éste e_t después cobra mayor importancia, porque para el próximo período es con él, ahora llamado e_{t-1} , con el que se va a formar el pronóstico al ser multiplicado por θ . Pero no va a pasar de ahí. El efecto del choque e_t solo va a afectar para un período y después de ahí será olvidado. Es por eso que podemos decir que los procesos MA tienen poca memoria.

Veamos ahora un MA de mayor grado

MA(5)

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3} + \theta_4 e_{t-4} + \theta_5 e_{t-5}$$

Con el operador rezago queda:

$$y_t = \Theta(B)e_t$$

donde $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4 + \theta_5 B^5$.

Aunque éste sí cuenta con un mayor memoria, podemos ver que los choques sólo tendrán influencia durante "q" períodos ($q = 5$ en nuestro caso).

Con la única diferencia de que tipográficamente los símbolos ψ se están sustituyendo por los θ , los modelos MA cumplen con ser casos específicos del PLG, por tener una cantidad finita de términos: $\theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3} + \dots + \theta_q e_{t-q}$.

A pesar de que de acuerdo con el PLG los factores θ pueden valer cualquier cosa, al momento de hacer que los modelos MA representen procesos reales, surge una restricción adicional: un proceso debe poder ser representado a través de sus *valores* observables y no únicamente con sus *choques* no observables.

Veamos esto qué quiere decir. Despejemos el del proceso MA(1)

$$\begin{aligned} y_t &= e_t - \theta e_{t-1} \\ e_t &= y_t + \theta e_{t-1} \end{aligned} \tag{7.8}$$

Retrasemos (7.8) un período

$$B[e_t = y_t + \theta e_{t-1}] = e_{t-1} = y_{t-1} + \theta e_{t-2}$$

Repitamos este procedimiento varias veces

$$B[e_{t-1}] = e_{t-2} = y_{t-2} + \theta e_{t-3}$$

$$B[e_{t-2}] = e_{t-3} = y_{t-3} + \theta e_{t-4}$$

$$B[e_{t-3}] = e_{t-4} = y_{t-4} + \theta e_{t-5}$$

...

Ahora sustituyamos estos despejes en el modelo original MA (1), queda:

$$y_t = e_t - \theta y_{t-1} - \theta^2 y_{t-2} - \theta^3 y_{t-3} - \theta^4 y_{t-4} + \dots$$

A esta suma infinita se le llama **representación autorregresiva**.

Algo contraintuitivo ocurre si el parámetro θ vale más que uno. Sucede que el pasado lejano tiene una mayor importancia que el pasado más reciente, lo cual sería catastrófico y de hecho no puede funcionar así.

Añádase a esto el hecho que la representación autorregresiva está explicando el proceso en función de valores observables de la serie (y_t) y no en función de choques inobservables (e_t) como originalmente lo estaba haciendo el modelo. Es preferible algo observable a lo no observable ¡si con lo que pretendemos tratar es con procesos reales!. Así que éstas dos son las razones por las cuales el proceso deba ser **invertible** y ésta es la restricción de la que hablábamos.

Esta **condición** se llama **de invertibilidad** e implica que el modelo sea funcional:

$$|\theta| < 1$$

Esta condición aplica para los procesos MA de cualquier grado, pero se explicó con el de primero por ser más sencillo. Para los procesos MA de grado mayor a uno, las condiciones son las mismas, pero existe la posibilidad de obtener raíces complejas en el polinomio de rezagos $\Theta(B)$.

En otras palabras: en polinomios de rezago el proceso MA(1) se escribe $y_t = (1 + \theta B)e_t$. El polinomio de rezagos $\Theta(B) = (1 + \theta B)$ se iguala a cero para obtener la raíz. Debido a que el factor de ponderación θ es una constante ya establecida, la raíz de la

ecuación es simplemente $B = -1/\theta$. Uno se puede preguntar ¿pero esta raíz de qué nos sirve?. Pues desde el punto de vista práctico, no de mucho; pero en lo teórico, sí. Es otra forma de establecer la condición para una covarianza constante, porque podemos decir que la *inversa de la raíz del polinomio de rezagos tiene que ser menor a uno en valor absoluto*. Es lo mismo que pusimos como condición de invertibilidad, pero expresado de diferente manera.

$$B = -\frac{1}{\theta}$$

$$-\frac{1}{B} = \theta$$

Si $\theta \geq 1$ el proceso no es estacionario.

Ahora, en cuanto a procesos MA de ordena mayor concierne, las condiciones de invertibilidad se deben explicar con estas raíces de polinomios, porque no se puede de otra manera. Así que para procesos MA de orden mayor las condiciones son las mismas, *que todas las inversas de las raíces del polinomio de rezagos sean menor que uno en valor absoluto*. Pero por álgebra elemental sabemos que para una ecuación de orden diferente a uno existe la posibilidad de tener raíces complejas. Así que, tomando esto en cuenta, la restricción de invertibilidad se traduce a que *todos los inversos de las raíces del polinomio de rezagos deben estar dentro del círculo unitario, que equivale a decir que todas las raíces deben estar fuera del círculo unitario*.

Después del análisis que acabamos de hacer confirmamos lo que se había dicho anteriormente que un polinomio de rezagos inverso tenía una representación en sumas infinitas porque

$$y_t = e_t - \theta e_{t-1}$$

$$y_t = (1 - \theta B)e_t$$

$$\frac{1}{(1 - \theta B)} y_t = e_t$$

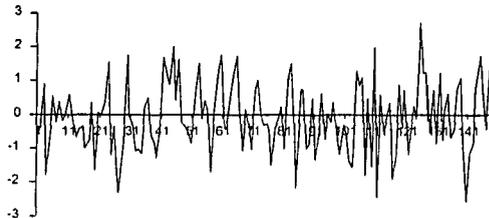
Así que

$$(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \dots) y_t = e_t$$

con lo que se comprueba lo igualdad entre la representación en polinomio rezago y sumas infinitas.

Ahora sí, simulemos procesos MA.

$$MA(1): y_t = 0.1 e_{t-1} + e_t$$



$$MA(1): y_t = 0.9 e_{t-1} + e_t$$

TESIS CON
FALLA DE ORIGEN

TESIS CON
FALLA DE ORIGEN



Es muy fácil hacer estas gráficas. Se pueden programar en algún lenguaje de programación o en alguna hoja de cálculo. El generador de ruido blanco es uno desarrollado por Box y Muller y es:

$$Z = \cos[2\pi(r_2)]\sqrt{-2\ln(r_1)}$$

Donde r_1 y r_2 son números pseudoaleatorios con distribución uniforme entre 0 y 1. Las dos gráficas utilizaron la misma serie de choques aleatorios.

El simular MA de orden 2 o de mayor orden en realidad no conlleva a resultados muy reveladores, por lo que nos limitamos a estos dos.

En las gráficas podemos ver que pese a que tienen parámetros muy distintos, las series en realidad no son tan diferentes entre sí como uno esperaría. Esto se debe a la poca memoria con que cuenta un proceso MA. La única diferencia es que el MA de 0.9 tiene mayor varianza que el de 0.1. La razón por la cual esto sucede, se puede ver si se calcula la varianza del proceso:

$$\text{var}(y_t) = \theta \text{var}(e_{t-1}) + \text{var}(e_t) = \sigma^2 (\theta^2 + 1)$$

Como podemos ver, si la σ^2 permanece constante, la varianza dependerá únicamente del valor del parámetro θ . Ahora veamos la función de autocorrelación.

Obtengamos primero la covarianza para después obtener la autocorrelación de un proceso MA (1):

$$\begin{aligned} \gamma_k &= E[y_t y_{t-k}] = E[(e_t + \theta e_{t-1})(e_{t-k} + \theta e_{t-k-1})] = \\ &= E(e_t e_{t-k}) + E(\theta e_{t-k} e_{t-1}) + E(\theta e_t e_{t-k-1}) + E(\theta^2 e_{t-1} e_{t-k-1}) = \begin{cases} \theta\delta^2 & \text{para } k = 1 \\ 0 & \text{para } k > 1 \end{cases} \end{aligned}$$

Recordemos que la covarianza, la esperanza de los productos entre dos variables, mide el nivel de relación que existe entre dos variables. Veamos entonces por separado cada uno de los términos que acabamos de obtener.

$E(e_t e_{t-k})$: Cuando $k=0$ sí existe una relación entre estas dos variables, que sería la varianza de e_t , o sea sigma cuadrada. Pero para $k=0$ ya dijimos que no nos interesa, porque toda serie está perfectamente correlacionada consigo misma. Cuando $k=1$ no puede existir una relación, porque estamos diciendo que cada golpe aleatorio es una variable aleatoria diferente, que no tiene nada que va con las demás variables, así que ésta, y las covarianza con $k>1$ valen cero. En suma, vale cero.

$E(\theta e_{t-k} e_{t-1})$: Existe una relación, σ^2 , cuando $k=1$. Pero para una k mayor todas las covarianzas son cero. O sea que es igual a $\theta\sigma^2$ para $k=1$.

$E(\theta^2 e_{t-1} e_{t-k-1})$; Cuando $k=0$, la covarianza vale cero, porque no existe relación entre e_t y e_{t-1} . Y con más razón cuando $k=1$, porque e_{t-2} está todavía más lejos en el tiempo a e_t . Vale cero.

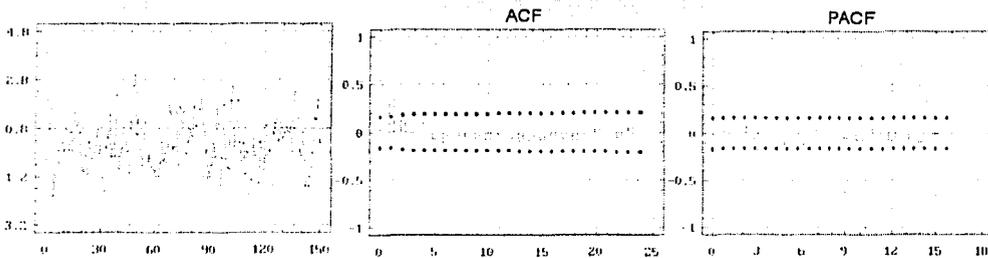
$E(\theta^2 e_{t-1} e_{t-k-1}) =$ Cuando $k=1$ queda $\theta^2 E(e_{t-1} e_{t-2})$ lo cual ya vimos que es igual a cero. Así que el único que no se volvió cero fue el segundo término, con lo que queda demostrado de dónde obtuvimos el resultado.

AUTOCORRELACIÓN DE UN PROCESO MA(1)

$$\gamma_0 = E(e_t e_t) + E(\theta e_t e_{t-1}) + E(\theta e_t e_{t-1}) + E(\theta^2 e_{t-1} e_{t-1}) = \sigma^2 + \sigma^2 \theta^2 = \sigma^2(\theta^2 + 1)$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\delta^2 \theta}{\delta^2(\theta^2 + 1)} = \frac{\theta}{\theta^2 + 1} & \text{para } k = 1 \\ 0 & \text{para } k > 1 \end{cases}$$

MA(2): $0.5e_{t-1} + 0.4e_{t-2} + e_t$



Los valores de PACF se obtienen de la siguiente manera: Tomamos un modelo que de alguna manera estemos seguros que es un MA(1) (por ejemplo si lo generamos nosotros mismos) y decimos 'No. Este no es un MA(1), es un AR(1) y calculamos el coeficiente ϕ_1 . Éste será el PACF para $k=1$. Luego repetimos 'No. Este no es un MA(1), es un AR(2)' y el coeficiente que obtengamos para ϕ_2 será el valor del PACF para $k=2$, y así sucesivamente, de tal manera que los valores del PACF serán los k -ésimos coeficientes estimados sobre una regresión a un AR de un orden k cada vez más grande.

Conclusiones: tenemos que todos los modelos MA de orden finito son aproximaciones válidas del PLG estacionario, sin importar cuánto puedan valer sus factores θ . Pero al quererlos aplicar a la vida real es necesario que cumplan con el requisito de ser capaces de representar un proceso real, es decir, de ponderar el pasado en menor medida que el presente. Por otro lado, todos los procesos AR tienen la posibilidad de ser aproximaciones válidas, siempre y cuando su representación PLG cuente con una suma convergente. Todos los procesos MA de orden finito ya son PLG estacionarios y sólo necesitan ser aplicables. En tanto que los AR solamente necesitan solamente ser PLG estacionarios, y con esto ya podrán ser aplicables. Podríamos llamarles 'aplicables' pero en adelante les llamaremos procesos AR y MA estacionarios.

TESIS CON FALLA DE ORIGEN

VII.3 Procesos mezclados

La única aproximación del PLG estacionario que queda, es en el que se combinan tanto una suma finita como una infinita, dando como resultado otra suma infinita.

$$y_t = \sum_{k=1}^p \phi_k y_{t-k} + \sum_{k=1}^p \theta_k e_{t-k} + e_t$$

de manera simplificada, con los polinomios de retraso:

$$\Phi(B)y_t = \Theta(B)e_t$$

Siendo un dado modelo AR y otro dado modelo MA estacionarios, se entiende que la suma de estos también lo sea. Porque si a una suma infinita convergente se le añade una suma con una cantidad finita de términos, la suma inicial si se va a hacer más grande, pero en definitiva no se va a hacer divergente ni infinita.

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3} + \theta_4 e_{t-4} + \theta_4^2 e_{t-5} + \theta_4^3 e_{t-5} + \theta_4^4 e_{t-5} + \dots$$



Parte MA

Los procesos ARMA para ser aplicables se topan con el mismo obstáculo que los MA. No les basta con ser una suma infinita convergente, ya que la parte MA podría contener coeficientes mayores que uno y aún así estaría representando un PLG. Así que los ARMA deben cumplir tanto las obligaciones de la parte MA (invertibilidad) como de la parte AR (convergencia).

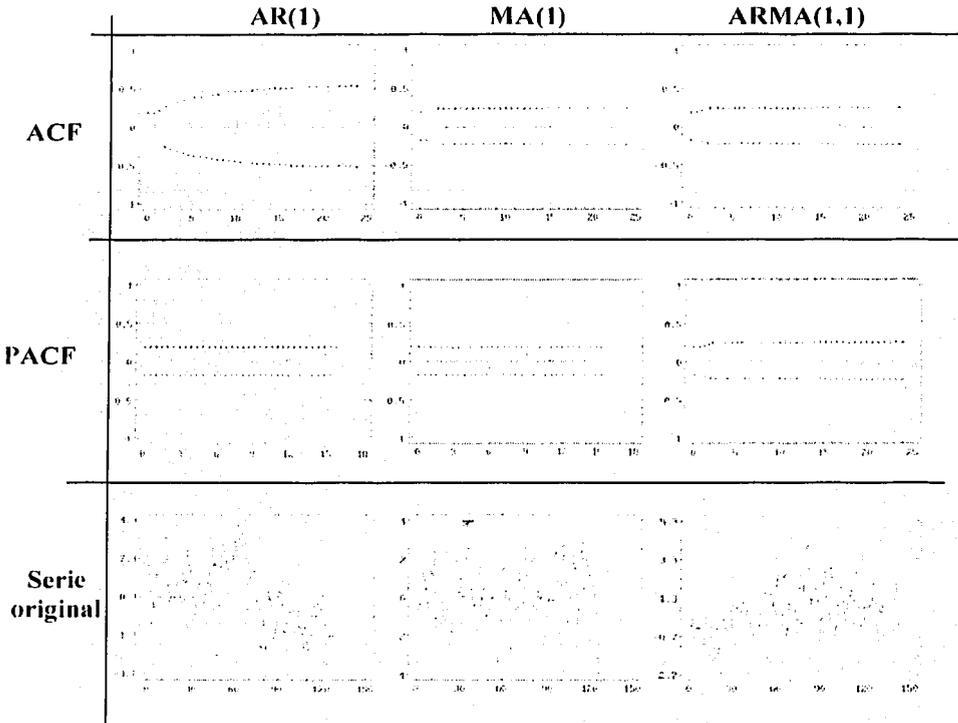
Los procesos AR y MA se pueden expresar en términos de ARMA. Por ejemplo un AR(1) se escribe ARMA(1,0). En la realidad casi nunca nos vamos a encontrar con modelos ARMA con p y q mayor a 2 (Brockwell, 1995).

ARMA(1,1)

$$y_t = \phi y_{t-1} + \theta e_t + e_t$$

Puede suceder que para una misma serie tanto el ρ_k como el ρ_{kk} disminuyan paulatinamente, pero lo que no puede pasar es que los dos se corten bruscamente. Debemos escoger cuál de los dos se está cortando bruscamente, entonces el otro tendrá que ser el que está disminuyendo. Para ello debemos ser cuidadosos al observar las demás herramientas que existen además de los ρ_k y ρ_{kk} .

CUADRO COMPARATIVO DE LOS CORRELOGRAMAS DE TRES SERIES DE TIEMPO SIMULADAS



TESIS CON FALLA DE ORIGEN

TESIS CON
FALLA DE ORIGEN

Capítulo VIII. Modelos no estacionarios

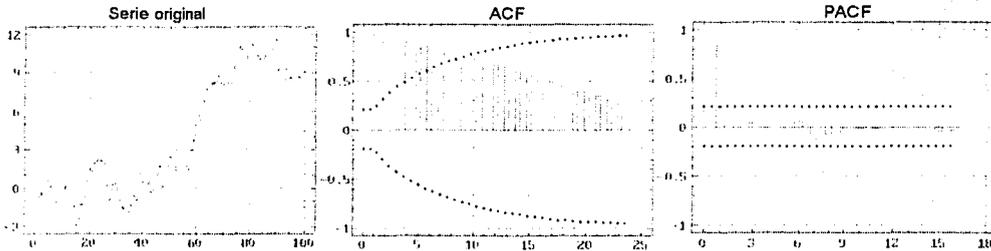
Los procesos no estacionarios son los que no satisfacen alguna de las condiciones de estacionalidad: no son ni invertibles o su representación en PLG converge. Por tanto tienen una varianza y autocovarianza cambiantes a través del tiempo, lo que los hace impredecibles. Asimismo, pueden presentar un cambio sistemático en su nivel, lo cual permite que exista cualquier clase de tendencia.

Los procesos no estacionarios son los que más nos vamos a encontrar en la realidad, por ser los que representan las series económicas, series de producción y muchas otras más. Es posible transformarlas a series estacionarias. Veremos en éste capítulo cómo eliminar media y varianza no estacionarias. En capítulos más adelante nos encargaremos especialmente de la estacionalidad.

¿CÓMO SE VEN LOS PROCESOS NO ESTACIONARIOS?

Para saber si una serie es estacionaria o no, primero debemos observarla. A simple vista es posible identificar alguna de las características ya mencionadas, como un crecimiento en la media y varianza, por ejemplo. Una tendencia implica una media no estacionaria. Un crecimiento en la varianza es una varianza no estacionaria. Si los valores son trimestrales, podemos esperar una estacionalidad de tamaño cuatro.

La ACF de una serie no estacionaria muestra coeficientes que tardan mucho en disminuir y permanecen significativos hasta valores de k muy grandes ($k > 5$). La PACF se muestra significativa sólo en $k=1$ pero con un valor muy grande, cercano a 1.



VIII.1 Procesos con tendencia

El primer ajuste que estaremos haciendo es para la tendencia. Las series que presentan cambio sistemático en su nivel pueden estar siendo generadas por uno de dos posibles modelos de tendencia: el determinístico y el estocástico.

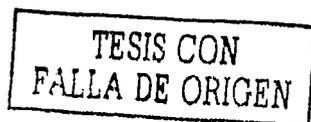
TENDENCIA DETERMINÍSTICA

La tendencia determinística es una que ya conocemos y que utilizamos en los primeros capítulos del libro*. Es simplemente la suposición de que la serie está creciendo alrededor de una línea, ya sea ésta curva o recta.

$$y_t = a + b(t) + e_t$$

$$y_t = a + b(t) + c(t^2) + e_t$$

* Ver capítulo 3



donde $e_t \sim WN(0, \sigma^2)$

Por lo general, las series no presentan una tendencia mayor a una de segundo grado.

Como vemos, estos modelos contienen la variable tiempo (t), lo cuál ya concluimos que era un impedimento para que un proceso fuera estacionario. Así que en un intento de eliminar esta tendencia, tendríamos que encontrar la manera de eliminar t.

Pero existen ciertas incongruencias en la metodología de tendencia determinista y de hecho es rara vez apropiada, porque todas las observaciones tienen el mismo peso (b) a la hora de hacer los pronósticos, lo cual no es razonable. Suponer una tendencia estocástica es mucho más realista, ya que a la hora de pronosticar no es forzoso que la serie siga el mismo patrón en el futuro del que siguió en el pasado.

Aunque en honor a la verdad, las series sí podrían estar siendo generadas por una tendencia determinística. Pero conocer si se trata de una serie determinística o una estocástica es sumamente difícil, por lo que se opta por suponer una estocástica *siempre*. Esta suposición da buenos resultados, sobre todo si la serie tiene una gran cantidad de datos.

TENDENCIA ESTOCÁSTICA

De acuerdo con el PLG estacionario, los parámetros θ de un MA de orden finito podrían valer cualquier cosa. Los AR son los que tenían una restricción (ambos sin tomar en cuenta la aplicabilidad). Por tanto un proceso no es un PLG estacionario si alguna de las raíces de su polinomio autorregresivo está dentro del círculo unitario.

Cuando en especial alguna de las raíces del polinomio autorregresivo ϕ vale uno, se dice que el modelo tiene una **raíz unitaria**.

El caso más sencillo y común es el de AR(1) con raíz unitaria.

$$y_t = y_{t-1} + e_t$$

Si obtenemos los retrasos consecutivos del modelo, de manera análoga a [7.10] la serie queda:

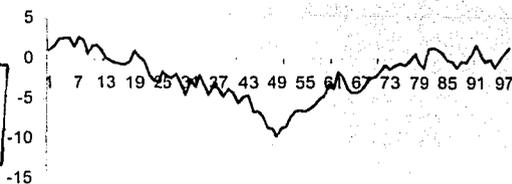
$$y_t = e_t + e_{t-1} + e_{t-2} + e_{t-3} + e_{t-4} + \dots$$

De tal forma que el primer golpe aleatorio es tan determinante como el último en el posicionamiento de la serie.

A este proceso se le conoce como caminata del borracho o **caminata aleatoria** (*Random Walk*), porque es como seguir los pasos de un borracho, donde su posición actual está determinada por su posición anterior más un golpe aleatorio.

Si simulamos una caminata aleatoria se ve:

$$y_t = y_{t-1} + e_t$$



Una de las características de esta serie es que presenta largos períodos de aparentes tendencias, crecientes o decrecientes, que pueden cambiar de dirección imprevisiblemente. También se puede advertir que los períodos adyacentes están extremadamente próximos el uno al otro.

¿Que pasa si le agregamos una constante a la caminata aleatoria?, ¿cree el lector que se comportará igual que antes?

$$y_t = \delta + y_{t-1} + e_t$$

Para obtener una respuesta, supongamos que el primer dato con el que contamos es y_0 .

$$y_1 = \delta + y_0 + e_1$$

adelantemos consecutivamente ahora el modelo en vez de retrasarlo, sólo para variar

sustituyendo y_1

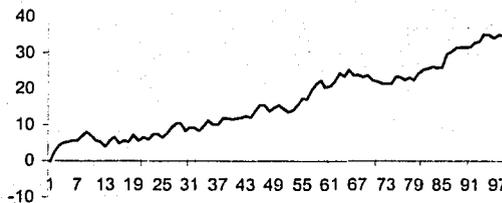
$$\begin{aligned} y_2 &= \delta + y_1 + e_2 \\ &= \delta + (\delta + y_0 + e_1) + e_2 \\ &= 2\delta + y_0 + e_1 + e_2 \\ y_3 &= \delta + y_2 + e_3 \\ &= \delta + (2\delta + y_0 + e_1 + e_2) + e_3 \\ &= 3\delta + y_0 + e_1 + e_2 + e_3 \end{aligned}$$

de manera que

$$y_t = (t)\delta + y_0 + \sum_{i=1}^t e_i$$

Así vemos que la caminata aleatoria con constante adquiere **tendencia**. Debido al carácter recursivo de los procesos ARMA la constante se multiplica por el tiempo, con lo que la serie pierde su media estacionaria. A este modelo también se le llama caminata aleatoria con deriva y es el generador de la tendencia estocástica.

Simulémosla, con un δ del 0.5:



TESIS CON FALLA DE ORIGEN

$$y_t = \delta + y_{t-1} + e_t$$

Este tipo de tendencia es muy común y afortunadamente los procesos que la contienen, llamados procesos *Autorregresivos Integrados y de Medias Móviles (ARIMA)*, son fáciles de modelar como si fueran ARMA.

ELIMINACIÓN DE LA TENDENCIA

Existen varias razones para preferir modelar con tendencias estocásticas en vez de determinísticas. Una de ellas ya la hemos mencionado: no son adecuados en sus ponderaciones de las observaciones a través del tiempo. Otra es que existe una metodología muy estudiada para éstas, que ha mostrado ser de gran efectividad. Por esta razón siempre

estaremos suponiendo que el crecimiento sistemático que se presente en una serie es una tendencia estocástica, la cual se elimina con **diferencias**.

Recordemos que la diferencia de primer orden de un proceso es:

$$\Delta y_t = y_t - y_{t-1}$$

que en forma de operador rezago queda:

$$\Delta y_t = (1-B) y_t$$

Así que al aplicarla a un modelo de tendencia estocástica

$$y_t = \delta + y_{t-1} + e_t$$

queda

$$\Delta y_t = \delta + y_{t-1} + e_t - \delta - y_{t-2} - e_{t-1}$$

$$= y_{t-1} + e_t - y_{t-2} - e_{t-1}$$

y si sustituimos y_{t-1}

$$= \delta + y_{t-2} + e_{t-1} + e_t - y_{t-2} - e_{t-1}$$

$$= \delta + e_t$$

lo que resulta no es más que la constante (que podemos pretender que sea cero) más ruido blanco, ¡lo que es muy útil! porque ya volvimos un proceso no estacionario a uno estacionario.

Claro que al diferenciar no nos va a quedar siempre un modelo de ruido blanco. En todo caso nos va a quedar un ARMA estacionario, lo que no le quita nada de fantástico.

Al trabajar con las diferencias, en vez de con la serie original, estamos diciendo que *los cambios que sufre la serie entre valores adyacentes están comportándose estacionariamente, se están comportando como un ARMA(p,q)*. Las diferencias de segundo grado se dan como

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$$

$$= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

$$= y_t - 2y_{t-1} + y_{t-2}$$

Con operador rezago

$$= (1-B)^2 y_t = y'_t$$

En donde y'_t es ya un proceso estacionario y y_t no lo es. Generalizando:

$$\Delta^d y_t = (1-B)^d y_t = y'_t$$

A este proceso y'_t se le conoce como un ARIMA(p,d,q).

VIII.2 Procesos Autorregresivos Integrados y de Medias Móviles ARIMA(p,d,q)

Si un proceso ARIMA se diferencia suficientes veces, va a dar como resultado un ARMA. O en otras palabras, un ARIMA es un ARMA *integrado* d veces, dado que la integración es la operación inversa a la diferenciación, lo que explica la letra I (Integración) en ARIMA.

Cuando un modelo ARIMA se haya tenido que diferenciar una vez para poder obtener un proceso estacionario, la I será igual a uno.

Si por ejemplo tomamos el modelo de caminata aleatoria

$$y_t = y_{t-1} + e_t$$

No sabemos qué modelo ARIMA es, pero le aplicamos una diferencia de orden uno se obtiene:

$$\begin{aligned} \Delta(y_t) &= y_{t-1} + e_t - (y_{t-2} + e_{t-1}) \\ &= y_{t-1} - y_{t-2} + e_t - e_{t-1} \end{aligned}$$

y dado que $y_t - y_{t-1} = e_t$, se puede inferir que $y_{t-1} - y_{t-2} = e_{t-1}$. Sustituyéndolo queda:

$$\begin{aligned} \Delta(y_t) &= e_{t-1} + e_t - e_{t-1} \\ &= e_t \end{aligned}$$

que es un ARMA(0,0). Así que y_t es un ARIMA(0,1,0)

Un modelo ARIMA(p,d,q) está definido como:

$$\Phi^*(B)y_t = \Theta(B)e_t$$

Donde alguna de las raíces de Φ^* está dentro del círculo unitario. Si podemos factorizar por lo menos una vez $(1-B)$ del lado izquierdo, entonces se dice que el modelo es un ARIMA de orden d .

$$\begin{aligned} \Phi(B)(1-B)^d y_t &= \Theta(B)e_t \\ \Phi(B)\Delta^d y_t &= \Theta(B)e_t \quad d \geq 1 \end{aligned}$$

donde las raíces de ambos polinomios $\Phi(B)$ y $\Theta(B)$ están todas fuera del círculo unitario.

Los casos que más nos interesan son en los que $d=0$ y $d=1$, por ser los más comunes. Algunos casos necesitaremos hacer segundas diferencias, pero esto se presenta las menos de las veces. Esto se debe, y es equivalente, a las mismas razones por las que la aproximación determinística sólo necesita hasta polinomios de grado dos (Diebold, 1998).

La primera diferencia de una tendencia determinista de segundo grado es

$$\begin{aligned} y_t &= a + b(t) + c(t^2) \\ \Delta y_t &= a + b(t) + c(t^2) - (a + b(t-1) + c(t-1)^2) \\ &= b + ct^2 - ct^2 + 2ct - c \\ &= 2ct - c \end{aligned}$$

Que no es más que la tendencia determinista. Al repetir la diferencia queda:

$$\begin{aligned} y_t &= 2ct - c \\ \Delta y_t &= 2ct - c - (2c(t-1) - c) \\ &= 2c \end{aligned}$$

Con lo que se elimina todo rastro de tendencia y resta una constante.

SOBREDIFERENCIACIÓN

En el caso que la raíz unitaria se encuentre en el polinomio de promedios móviles y no en el autorregresivo, se dice que se ha sobrediferenciado al proceso. Es decir, la única manera de encontrar una proceso de promedios móviles con raíz unitaria es por haber diferenciado y la gráfica mostrará el paso de estacionalidad a no estacionaridad. Es por eso que se debe tener mucho cuidado al hacer las diferencias, porque si una serie ya es estacionaria, al seguir aplicando diferencias, va a seguir estando horizontal, pero el modelo se va tornar más complejo. Por ejemplo, apliquemos segundas diferencias al modelo de tendencia estocástica.

Ya hicimos la primer diferencia, que resultó:

$$\Delta y_t = \delta + e_t$$

¿Qué sucede si le aplicamos Δ otra vez?

$$\begin{aligned} \Delta^2 y_t &= (\delta + e_t) - (\delta + e_{t-1}) \\ &= e_t - e_{t-1} \end{aligned}$$

Este es un modelo AR(1) con $\theta = 1$ que nos sacamos de la manga, porque ya teníamos al proceso estacionario ¡y lo volvimos no estacionario!. Nótese que con cada diferencia se pierde un dato, por lo que necesitamos muchos datos para poder utilizar los modelos ARIMA(p,d,q).

¿Pero por qué tanta importancia a las raíces autorregresivas unitarias?. Las raíces unitarias en promedios móviles ya dijimos que son el resultado de sobrediferenciación y son imposibles de encontrar en casos reales. Pero de todas maneras simulemos una para conocer su apariencia

MA(1) con raíz unitaria

TESIS CON FALLA DE ORIGEN



¿Y qué hay de un MA(1) con θ mayor que uno?. Simulemos una con $\theta = 1.4$

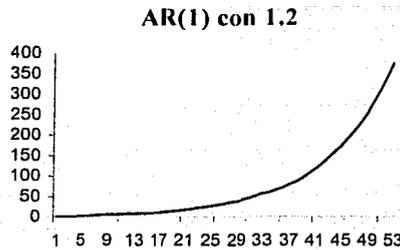


Se pueden seguir incluyendo rezagos de golpes aleatorios y los efectos son similares.

¿Qué concluimos con los MA con raíces dentro del círculo unitario?. Además de que no son de uso práctico, se confunden visualmente con procesos estacionarios. Lo más

importante es que al obtener los estimadores de los parámetros AR y MA, estos no detectan ninguna raíz unitaria. Sólo detectan que se trata de un proceso estacionario. Interesante ¿no?. A pesar de que es no estacionaria "teóricamente" genera una serie estacionaria "prácticamente" y además con estimadores estacionarios.

Ahora pasemos a los AR. Simulemos un AR(1) una con raíz igual a 0.83, es decir con el parámetro ϕ igual a 1.2



Se aprecia que este proceso no tiene nada de estocástico, y es simplemente una función cuadrática. Cuanto mayor sea el orden del AR, cuanto más acelerado será el crecimiento de la línea.

Estos efectos se contagian a los MA cuando se conjuntan en los ARMA(p,q).

Si estimamos los valores de los parámetros de las simulaciones realizadas, el software (Statgraphics) los va a obtener como lo haría normalmente, ya que no le importa si el proceso tiene raíces estacionarias o no, como se dijo anteriormente.

Podemos concluir entonces que la razón por la que se han escogido las raíces unitarias para su estudio en los procesos estocásticos es por su gran utilidad y su fácil representación y manejo matemático.

| Características visuales de procesos con raíces que no satisfacen condiciones del PLG | | |
|---|--|---|
| | AR | MA |
| Raíz unitaria | Tendencia estocástica | Se confunden con procesos estacionarios |
| Raíces inversas fuera del círculo unitario | Se puede aproximar determinísticamente | Se confunden con procesos estacionarios |

VIII.3 Estabilización de la varianza

Algunas veces el hacer ciertos ajustes a las series antes de proceder a modelarlas, conlleva a modelos más simples y más interpretables de pronóstico.

En lo tocante a la varianza, existen ciertas transformaciones matemáticas que pueden estabilizarla de manera muy efectiva. Las transformaciones más eficientes son la raíz cuadrada y el logaritmo (en realidad no importa la base, ya que lo que nos interesa es en un cambio de escala en la serie). Además de éstas, existe la raíz cúbica y el recíproco negativo. Si denotamos a w_t la serie transformada y y_t la serie original, la familia de transformaciones estabilizadoras de varianza está definida de la siguiente manera:

TESIS CON FALLA DE ORIGEN

$$w_i = \begin{cases} -y_i^p & p < 0 \\ \log(y_i) & p = 0 \\ y_i^p & p > 0 \end{cases}$$

En este contexto p es el exponente de la función. Para $p=1$ la transformación es simplemente $w_i = y_i$, lo cual deja los datos intactos. Al escoger $p = 1/2$ la transformación es la raíz cuadrada y $p=-1$ da el recíproco inverso. La razón por la cual $p=0$ es el logaritmo es porque, cuando p está muy cercano a cero, y_i^p se comporta mucho como un logaritmo. Para $p < 0$ se utiliza el signo negativo para que la transformación incremente cuando la función decrece. El parámetro p puede valer cualquier cosa si todos los datos de la serie son positivos, pero p debe ser mayor que cero si los datos contienen algún cero. Si los datos contienen valores negativos, no se les puede aplicar ninguna transformación, en cuyo caso será necesario añadir una constante a la serie.

Los pronósticos se formulan con los datos transformados, pero como los pronósticos que nos interesan son los de los datos originales, es necesario regresarlos como estaban, aplicando la función inversa a la que se utilizó.

Como regla, es preferible utilizar valores sencillos de p que den transformaciones como el logaritmo o raíz cuadrada. Los modelos y pronósticos son relativamente indiferentes al tipo de transformación que se elija, ya que valores colindantes de p producen resultados similares. Además, valores sencillos como 0, -1 ó $1/2$ hacen los resultados más fáciles de interpretar que un número como 0.2439804.

Un proceso que es de tendencia estacionaria, es decir de media constante, no es necesariamente de varianza estacionaria, pero una serie que tiene tendencia tendrá también varianza no estacionaria. Para ver qué implicaciones tiene esto retomemos la fórmula de la tendencia estocástica:

$$y_t = \delta + y_{t-1} + e_t$$

Ya comprobamos* que su representación era también:

$$y_t = t\delta + y_0 + \sum_{i=1}^t e_i$$

A primera vista ya se puede ver que la sumatoria de los golpes aleatorios está acrecentando cada vez más el ruido, pero procedamos a obtener la varianza, para cerciorarnos:

$$\begin{aligned} \text{var}[y_t] &= \text{var}[t\delta] + \text{var}[y_0] + \text{var}\left[\sum_{i=1}^t e_i\right] \\ &= \text{var}[e_1 + e_2 + e_3 + e_4 + \dots + e_t] \\ \text{var}[y_t] &= t\sigma^2 \end{aligned}$$

Así que, dependiendo la varianza del tiempo, lo siguiente es válido:

* Ver tema VII.1 "Procesos con tendencia"

$$\lim_{t \rightarrow \infty} \text{var}[y_t] = \infty$$

Con esto se comprueba que la tendencia estocástica tiene una varianza creciente, lo que la haría aspirante a alguna transformación, pero ya vimos que cuando se hacen las diferencias, el modelo resulta de varianza estacionaria.

En realidad no existe una regla clara de cuál es el camino a seguir. Se puede diferenciar y esto podría eliminar la tendencia, pero puede ocurrir que aún después de diferenciar siga habiendo varianza creciente, así que sería necesario aplicar alguna transformación.

En realidad se procede a manera de prueba y error, normalmente empezando por probar las transformaciones, y después encontrando alguna combinación de diferencias y transformaciones que presenten los datos más estables y predecibles, tomando siempre en cuenta la importancia de no sobrediferenciar.

**TESIS CON
FALLA DE ORIGEN**

Capítulo IX: Pronósticos óptimos con los modelos estacionarios y no estacionarios

IX.1 Pronósticos de error cuadrático mínimo

Ahora sí, finalmente, vamos a pronosticar. ¿Cómo se hace esto? Es muy sencillo, son dos simples pasos. Primero escribimos el modelo para el período que queremos hacer el pronóstico y después todo lo que desconozcamos de ese modelo en el presente, le asignamos su valor esperado.

Es más fácil verlo con un ejemplo. Tomemos un modelo MA(2)

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

Ahora el primer paso. Escribimos el modelo para el período que queremos el pronóstico. Empecemos por pronosticar para un sólo período, simplemente sumando una unidad de tiempo a toda la ecuación.

$$y_{T+1} = e_{T+1} + \theta_1 e_T + \theta_2 e_{T-1}$$

Hemos puesto una T mayúscula en vez de la minúscula para indicar que T es el *presente*, el momento en el cual se está formulando el pronóstico. El segundo paso es sustituir todo lo que esté en el futuro, y que no conozcamos, por su valor esperado.

$$P(y_{T+1} | \Omega_T) = y_{T+1,T} = \theta_1 e_T + \theta_2 e_{T-1}$$

A esta operación que acabamos de hacer en dos pasos se le llama proyección sobre el conjunto de información en el momento T o **proyección lineal**. Se puede entender como un pronóstico responsable, porque toma en cuenta toda la información con la que dispone en el momento de hacer el pronóstico.

Hagamos ahora el pronóstico pero para dos etapas hacia el futuro

$$y_{T+2} = e_{T+2} + \theta_1 e_{T+1} + \theta_2 e_T$$

Y al hacer la proyección lineal queda $y_{T+2,T} = \theta_2 e_T$

Para tres períodos adelante $y_{T+3} = e_{T+3} + \theta_1 e_{T+2} + \theta_2 e_{T+1}$

por ende $y_{T+3,T} = 0$

Directamente se ve que cualquier otro pronóstico más adelantado hacia el futuro con un modelo MA(2) dará como resultado cero.

$$y_{T+h,T} = 0 \quad \text{para todo } h \text{ mayor que dos.}$$

Nótese que se utiliza la notación $y_{T+h,T}$ y $P(y_{T+h} | \Omega_T)$ indistintamente.

La proyección lineal es la mejor manera de pronosticar con la que contamos y será en la que nos basaremos de aquí en adelante.

Pero de hecho ya nos adelantamos. deberíamos haber comenzado por el concepto de pronóstico óptimo para comprender cómo se llega a la proyección lineal.

PRONÓSTICO ÓPTIMO

Un pronóstico óptimo es aquel que minimiza el error de pronóstico, el cual está dado como la diferencia entre el valor real y el valor pronosticado.

$$e_{t+h} = \hat{y}_{t+h} - y_{t+h}$$

donde h es el tiempo hacia el futuro para el que se está haciendo el pronóstico, \hat{y}_{t+h} es el pronóstico; y_{t+h} es el valor observado.

Para contar con una cuantificación que no involucre cancelaciones de números positivos con negativos, se utiliza el error al cuadrado

$$(e_{t+h})^2 = (y_{t+h, T} - \hat{y}_{t+h})^2$$

Y finalmente su esperanza

$$\text{MSE}(y_{t+h, T}) = E[(y_{t+h, T} - \hat{y}_{t+h})^2]$$

La cual se denomina error cuadrático promedio o MSE por sus siglas en inglés (Mean Square Error). El pronóstico que tiene el mínimo MSE es el **pronóstico óptimo**. Este concepto es el pilar de la teoría que utilizamos para seleccionar el mejor modelo, así que estamos buscando, porque todavía no le tenemos, un *predictor* que cumpla con ser óptimo.

ESPERANZA CONDICIONAL

Una forma muy intuitiva de pronosticar es la esperanza condicional o promedio condicional

$$E(y_{t+h} | \Omega_T)$$

donde $\Omega_T = \{e_t, e_{t-1}, e_{t-2}, \dots\}$ es el conjunto de información que se tiene hasta el momento de formular el pronóstico. Recordemos que, para un proceso estocástico estacionario, dentro de Ω_T también pueden venir incluidos los valores observados y_t , pero dado que siempre es posible representarlos como choques e_t hemos considerado al conjunto de información con los golpes aleatorios únicamente.

Se sabe que para condiciones de estacionalidad "razonablemente" débiles (Hamilton, 1997), el pronóstico con esperanza condicional genera un MSE mínimo y por lo tanto un pronóstico óptimo.

¿Ya encontramos el predictor óptimo entonces? No. El promedio condicional nos puede ocasionar complicaciones a la hora de tratar de calcularlo. Este promedio puede trabajar de manera lineal o no lineal con los valores de Ω_T ; cuando se comporta de manera no lineal puede generar cálculos verdaderamente complejos, por lo que nos vemos forzados a utilizar el caso lineal únicamente. Así concluimos que necesitamos trabajar con un pronóstico que sea óptimo y a la vez que sea lineal.

PRONÓSTICO ÓPTIMO LINEAL

Un predictor, para ser lineal, necesita ser de la forma:

$$\text{Pronóstico} = \alpha \Omega_T$$

donde α es un vector de constantes.

Y dijimos que necesita también tener el mínimo MSE para poder ser óptimo. Es decir,

$$E[y_{T+h} - \alpha\Omega_T]^2 \quad [9.1]$$

debe ser mínimo. Ω_T puede ser por ejemplo, un valor observado y_t , o tres valores y_t, y_{t-1}, y_{t-2} o cuatro choques aleatorios, e_t, e_{t-1}, e_{t-2} y e_{t-3} , o cualquier mezcla de estos. El vector α contendría, desde el enfoque ARIMA, a los parámetros θ y ϕ .

Para encontrar el MSE mínimo de [9.1] necesitamos encontrar el valor de α que satisfaga la minimización del error. Esto se resuelve como lo haríamos minimizando una función comúnmente, con α como la incógnita, Ω_T y y_{T+h} como constantes.

El primer paso es encontrar los puntos críticos, obteniendo las raíces de la primera derivada de [9.1] igualada a cero.

$$\begin{aligned} \frac{d}{d\alpha} [y_{T+h} - \alpha\Omega_T]^2 &= 0 \\ = -2\Omega_T[y_{T+h} - \alpha\Omega_T] &= 0 \end{aligned} \quad [9.2]$$

Pero el MSE puede ser solamente un punto y éste tiene que ser forzosamente el mínimo, ¿por qué?, porque el máximo puede ser cualquier cosa. es decir, ∞ , y los puntos de inflexión no pueden existir.

Así que únicamente debemos resolver [9.2] para α .

Obsérvese lo que sucede al restaurar la esperanza que habíamos omitido en [9.2]

$$= -2E[(y_{T+h} - \alpha\Omega_T) \Omega_T] = 0$$

Ésta es la covarianza entre el error de pronóstico y el conjunto de información, lo que es más fácil de interpretar que [9.2]. De hecho se puede prescindir del -2 y la condición no se ve afectada.

Condición de optimalidad

$$E[(y_{T+h} - \alpha\Omega_T) \Omega_T] = 0$$

Así tenemos que el pronóstico lineal que tenga una correlación nula entre su error de pronóstico y el conjunto de información será el de MSE mínimo y, por tanto, el óptimo lineal.

Al comenzar el capítulo ya estábamos utilizando la proyección lineal para pronosticar como si fuera el óptimo. La cuestión es entonces ¿satisface la proyección lineal las condiciones de optimalidad y linealidad?

Pongamos a la proyección lineal a trabajar sobre el PLG y dependiendo de los resultados que nos proporcione podremos decidir.

El PLG está formulado como

$$y_t = \sum_{i=0}^s \psi_i e_{t-i}$$

donde $\psi_0=1$ y $e_t \sim WN(0, \sigma^2)$.

El primer paso de la proyección lineal es fijar un presente T y formular el modelo para las etapas que deseamos pronosticar. En este caso generalizaremos para h etapas.

$$y_{T+h} = e_{T+h} + \psi_1 e_{T+h-1} + \psi_2 e_{T+h-2} + \dots + \psi_{h-1} e_{T+1} + \psi_h e_T + \psi_{h+1} e_{T-1} + \psi_{h+2} e_{T-2} + \dots \quad [9.3]$$

Toda la primera parte del lado derecho de la ecuación (es decir, antes de los primeros tres puntos) desaparece al hacer la proyección al presente (segundo paso):

$$y_{T+h,T} = \psi_h e_T + \psi_{h+1} e_{T-1} + \psi_{h+2} e_{T-2} + \dots \quad [9.4]$$

Ahora calculemos el error de pronóstico:

$$\begin{aligned} e_{T+h,T} &= (y_{T+h,T} - y_{T+h}) = \\ &= \text{primer paso} - \text{segundo paso} = \\ &= e_{T+h} + \psi_1 e_{T+h-1} + \psi_2 e_{T+h-2} + \dots + \psi_{h-1} e_{T+1} \end{aligned} \quad [9.5]$$

El error de pronóstico, aunque está autocorrelacionado, no está correlacionado con el conjunto de información ($e_t, e_{t-1}, e_{t-2}, \dots$) con lo que se satisface la condición de optimalidad. Además $y_{T+h,T}$ sí es lineal con el conjunto de información, con lo que se satisface la linealidad. Hemos comprobado así que la proyección lineal es el mejor pronóstico con el que podemos contar porque es el predictor lineal óptimo que estábamos buscando.

PRONÓSTICOS DE MODELOS MA(q)

Ya que comprendemos por qué utilizamos la proyección lineal, retomemos los pronósticos que estábamos realizando con los procesos MA

$$\text{MA}(3): \quad y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3}$$

1er paso para un período en el futuro

$$y_{T+1} = e_{T+1} + \theta_1 e_T + \theta_2 e_{T-1} + \theta_3 e_{T-2}$$

2do paso

$$y_{T+1,T} = \theta_1 e_T + \theta_2 e_{T-1} + \theta_3 e_{T-2}$$

Repetimos para dos períodos

1er paso

$$y_{T+2} = e_{T+2} + \theta_1 e_{T+1} + \theta_2 e_T + \theta_3 e_{T-1}$$

2do paso

$$y_{T+2,T} = \theta_2 e_T + \theta_3 e_{T-1}$$

Para tres períodos

1er paso

$$y_{T+3} = e_{T+3} + \theta_1 e_{T+2} + \theta_2 e_{T+1} + \theta_3 e_T$$

2do paso

$$y_{T+3,T} = \theta_3 e_T$$

Y finalmente para 4 períodos

1er paso

$$y_{T+4} = e_{T+4} + \theta_1 e_{T+3} + \theta_2 e_{T+2} + \theta_3 e_{T+1}$$

2do paso

$$y_{T+4,T} = 0$$

No sería nada incongruente afirmar, después de ver estos resultados, que para cualquier proceso MA(q) su capacidad de pronóstico termina en q períodos en el futuro, como de hecho sucede.

¿Pero por qué decimos que su capacidad de pronosticar termina?, ¿acaso el pronóstico igual de cero no es éste suficiente? No.

Si calculamos la esperanza comúnmente usada, también llamada incondicional, de los procesos MA(q), veremos que pronosticar igual a cero resulta una obviedad

$$\begin{aligned} E[MA(q)] &= E[e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}] \\ &= E[e_t] + \theta_1 E[e_{t-1}] + \theta_2 E[e_{t-2}] + \dots + \theta_q E[e_{t-q}] \\ &= 0 + 0 + 0 + \dots + 0 \\ &= 0 \end{aligned}$$

(Recordemos que hemos retirado el recorrimiento vertical para simplificar las cosas al suponer una media de cero. Así que cuando decimos “igual a cero”, equivale a decir “a la media de la serie”).

Si el pronóstico que no cuenta con información y el que sí cuenta con ella están arrojando el mismo resultado, entonces, ¿de qué le sirve todas esa información (Ω_T) a la proyección lineal?

No podemos culpar a la proyección lineal de esto. Después de todo son los procesos reales los que se están comportando como MA(q) de orden finito. Además, nadie ha dicho que esto sea malo necesariamente. Si actualizamos los pronósticos cada q etapas de tiempo, o menos, tendremos siempre pronósticos muy vigentes que nunca pronosticarán cero (o la media).

PRONÓSTICOS CON PROCESOS AR(P)

Los pronósticos de los promedios móviles finitos que acabamos de ver son un caso especial de los infinitos. Los procesos AR(p) se pueden representar con promedios móviles infinitos. El pronóstico de estos ya lo realizamos en [9.3] y [9.4]. Vale la pena recapitular porque se nos escaparon algunos detalles.

Decíamos que la proyección lineal de un PLG para h etapas realizado en el momento T era

$$y_{T+h,T} = \varphi_h e_{T+1} + \varphi_{h+1} e_{T+1} + \varphi_{h+2} e_{T+2} + \dots \quad \text{de [9.4]}$$

y el error de pronóstico para la misma cantidad de etapas

$$e_{T+h,T} = e_{T+h} + \varphi_1 e_{T+h-1} + \varphi_2 e_{T+h-2} + \dots + \varphi_{h-1} e_{T+1} \quad \text{de [9.5]}$$

Ahora, ¿qué pasa si el horizonte de pronóstico se hace muy grande, digamos infinitamente grande?. Recordemos que estamos hablando de procesos estacionarios, así que los valores de θ más antiguos deben ser más pequeños que los recientes. Si esto se cumple, cuando h tienda a infinito, tendremos $\varphi_h e_{T+1} + \varphi_{h+1} e_{T+1} + \varphi_{h+2} e_{T+2} + \dots$ y el pronóstico [9.4] tenderá a cero, lo que es igual a la esperanza incondicional del proceso, que es la esperanza que conocemos

ESPERANZA INCONDICIONAL DEL PLG

$$E[PLG] = E[e_t] + \varphi_1 E[e_{t-1}] + \varphi_2 E[e_{t-2}] + \dots = 0 \quad [9.6]$$

por lo que la información que tenemos se tornará cada vez menos útil conforme nos alejemos hacia el futuro.

Recordemos que los procesos AR(p) tienen un infinito número de choques en su conjunto de información, por lo que su capacidad de pronóstico es mucho mayor a la de los MA(q). Vemos en [9.4] que para un h razonablemente pequeño (menor que p), el pronóstico estaría bastante alejado del pronóstico obvio, cero, y por ende nos estaría aportando información muy valiosa.

REGLA DE LA CADENA PARA LOS PRONÓSTICOS

Afortunadamente, a diferencia del enredado método de los MA(q), los procesos AR(p) tienen una forma mucho más fácil de ser computados, por medio de un método llamado **regla de la cadena para pronósticos**.

Veamos la explicación de éste método al mismo tiempo que lo vamos aplicando a un ejemplo. Tomemos el modelo AR más sencillo: AR(1)

$$y_t = \phi_1 y_{t-1} + e_t$$

Primero determinamos el pronóstico a una etapa, reescribiendo el proceso para la etapa T+1

$$y_{T+1} = \phi_1 y_T + e_{T+1}$$

Después realizamos la proyección sobre la información que tenemos en el tiempo T

$$y_{T+1,T} = \phi_1 y_T$$

Ahora repetimos para dos etapas

$$y_{T+2} = \phi_1 y_{T+1} + e_{T+2}$$

Y volvemos a aplicar la proyección

$$y_{T+2,T} = \phi_1 y_{T+1,T}$$

Ya sabemos lo que vale $y_{T+1,T}$. Repetimos el proceso a tres periodos, para que no quede duda.

Primer paso: $y_{T+3} = \phi_1 y_{T+2} + e_{T+3}$

Segundo paso: $y_{T+3,T} = \phi_1 y_{T+2,T}$

Y también ya conocemos lo que vale $y_{T+2,T}$ de tal manera que para pronosticar a cualquier etapa en el futuro un proceso AR(1) sólo se necesita el valor más reciente de y_t . Para un proceso AR(p) se necesitan los p valores más recientes.

PRONÓSTICOS PARA PROCESOS ARMA(P,Q)

Gracias a que los procesos mezclados ARMA(p,q) contienen una parte AR, es posible calcular los pronósticos con la regla de la cadena para los pronósticos.

Si tomamos un ARMA(1,1):

$$y_t = \phi y_{t-1} + \theta e_{t-1} + e_t$$

El proceso representado a una etapa en el futuro

$$y_{T+1} = \phi y_T + \theta e_T + e_{T+1}$$

y aplicando la proyección lineal

$$y_{T+1,T} = \phi y_T + \theta e_T$$

Para dos periodos de tiempo en el futuro

$$y_{T+2} = \phi y_{T+1} + \theta e_{T+1} + e_{T+2}$$

la proyección lineal de esta ecuación

$$y_{T+2,T} = \phi y_{T+1,T}$$

repetiendo los mismos pasos para tres y cuatro periodos

$$y_{T+3} = \phi y_{T+2} + \theta e_{T+2} + e_{T+3}$$

$$y_{T+3,T} = \phi y_{T+2,T}$$

$$y_{T+4} = \phi y_{T+3} + \theta e_{T+3} + e_{T+4}$$

$$y_{T+4,T} = \phi y_{T+3,T}$$

Como se aprecia fácilmente, el pronóstico solamente necesita los estimados de los parámetros y los datos más recientes ($\phi y_T + \theta e_T$), ya que toda proyección realizada a cualquier cantidad de periodos en el futuro es simplemente un múltiplo de éstos.

MEDICIÓN DE LA INCERTIDUMBRE EN LOS PRONÓSTICOS

Aún queda por resolver una cuestión, que es la de la precisión del pronóstico, tema tan importante como el pronóstico puntual mismo en cualquier metodología. Sin una medición de la incertidumbre de los pronósticos, éstos no nos son de mucho uso porque no contamos con una idea clara de su precisión. Sería como decir "mi pronóstico es éste, pero lo más seguro es que quién sabe".

En muchos casos la palabra precisión se refiere también a la bondad de ajuste, que a su vez se refiere a la capacidad del modelo de ajustarse a los datos históricos. Pero en la realidad, lo que interesa es que el modelo se ajuste a los datos del futuro, no del pasado. Esto último es muy importante y no se debe olvidar.

INTERVALOS DE CONFIANZA

Una manera acostumbrada de expresar la incertidumbre en los pronósticos es a través del intervalo de confianza, esto es, un intervalo que contenga el valor futuro que estamos calculando con una cierta probabilidad, ajustable por cierto, que preferimos que sea alta, como de .95 ó .99.

Los intervalos de confianza son muy útiles, porque proveen a la persona que está utilizando un modelo de pronóstico de estimados del mejor y peor caso, así como de una idea de cuán confiable es cierto modelo. Curiosamente de paso protege al pronosticador de las críticas que puedan surgir: no se puede esperar que ningún pronóstico sea perfecto y los intervalos de confianza enfatizan precisamente este hecho.

Adicionalmente se debe señalar que el pronosticar es una tarea dinámica: cada vez que surgen nuevos datos es necesario actualizar los pronósticos, tomando en cuenta la nueva información. Esto se verá en el tema IX.3 "Actualización de pronósticos".

INTERVALOS DE CONFIANZA PARA PRONÓSTICOS DE PROCESOS AR(p)

En [9.3] y [9.5] ya contamos con el pronóstico y su error de los AR(p). Obtenemos ahora su varianza

$$e_{T+h,T} = e_{T+h} + \phi_1 e_{T+h-1} + \phi_2 e_{T+h-2} + \dots + \phi_{h-1} e_{T+1}$$

$$\begin{aligned} \text{var}(e_{t+h,t}) &= \text{var}(e_{t+h}) + \text{var}(\varphi_1 e_{t+h-1}) + \text{var}(\varphi_2 e_{t+h-2}) + \dots + \text{var}(\varphi_{h-1} e_{t+1}) \\ &= \delta^2 + \varphi_1^2 \delta^2 + \varphi_2^2 \delta^2 + \varphi_3^2 \delta^2 + \dots + \varphi_{h-1}^2 \delta^2 \\ &= \delta^2 (1 + \varphi_1^2 + \varphi_2^2 + \varphi_3^2 + \dots + \varphi_{h-1}^2) \end{aligned} \quad [9.7]$$

$$\sum_{i=0}^{h-1} \varphi_i^2 \delta^2 \quad \text{con } \varphi_0 = 1$$

Esta fórmula es la que nos proporciona el intervalo de confianza para procesos AR(p), MA(∞) o PLG, que para el caso son lo mismo. Con [9.7] es más fácil ver que conforme nos alejamos hacia el futuro la varianza del error del pronóstico, y por tanto el intervalo de confianza, se va haciendo cada vez más grande, lo que nos avisa que podemos confiar más en los pronósticos de plazo pequeño que los de plazo grande.

Además, como ya habíamos concluido antes, cuanto mayor sea el horizonte de pronóstico, menos útil se torna la información del pasado.

$$\text{var(PLG)} = \sum_{i=0}^{\infty} \psi_i^2 \delta^2 \quad \text{con } \psi_0 = 1$$

Hasta ahora hemos supuesto que el intervalo de confianza debe estar dado por el error de pronóstico, pero ¿de dónde sale esa idea? ¿qué garantiza que esto sea cierto? La teoría del porqué se utiliza el error de pronóstico como intervalo de confianza está bien fundamentada y vale la pena profundizar en ella, ya que envuelve muchos conceptos necesarios para entender en el futuro cualquier extensión que se haga de la metodología de pronósticos que estamos estudiando.

En un principio supusimos que los errores de pronóstico se deberían comportar como una distribución Normal con media cero, varianza constante, y no estar correlacionados entre sí. El punto es que estamos *suponiendo* una distribución Normal porque tiene características bien conocidas para nosotros. Además, la similitud entre cómo el error *debería estar* bastante cerca a cero la mayoría de las veces y cómo la distribución Normal *ya cuenta con esta característica*, es comprobación suficiente para que utilicemos las propiedades de la Normal, aunque en realidad es una suposición no comprobada teóricamente.

Supongamos que y_{t+h} es nuestro *presente*, que ahí es donde nos encontramos y que nuestra serie está *trunca*, es decir, nos hacen falta h elementos en nuestra serie: los h valores inmediatamente previos al presente. Así que podemos rellenar la serie, con un modelo basado en los datos precedentes que sí tenemos. Pero dado que el proceso es estocástico, cada elemento de la serie puede tomar un sólo valor de un conjunto de valores (la distribución de probabilidad): debido a que los valores de la serie están relacionados, por el carácter recursivo de los modelos ARIMA, una pequeña variación en uno de ellos generaría una serie completamente diferente.

Aunque e_t sea pequeña, puede tomar un valor entre -3 y $+3$ porque la estamos simulando como una Normal. Si al rellenar la serie asignamos el valor de -3 aleatoriamente y el valor real que sucede es de $+3$, no importa que acertemos en los siguientes e_t , la serie será una completamente distinta y, a saber, errónea.

Lo que la varianza del error de pronóstico hace es precisamente tomar en cuenta ésta incertidumbre inherente en los procesos estocásticos y calcularla.

Si tenemos que rellenar un sólo espacio (pronosticar), la cosa no es tan grave, porque el intervalo de confianza es bastante pequeño, a saber de (-3,3) si nos interesa tener una exactitud del 99%. Pero si nos atrevemos a rellenar dos espacios, entonces el intervalo cambia. Porque si, como dijimos, escogemos un -3 y resulta un +3, el siguiente período podemos volver a escoger el -3 y nos podría volver a salir un +3 como valor real. El error de pronóstico advierte de esta situación.

Claro que este escenario que acabamos de suponer es el peor posible y en verdad tiene muy pocas posibilidades de suceder; además, como θ es menor a 1, la varianza no se dispara tan drásticamente como en el ejemplo que acabamos de poner.

Esto nos hace recordar que al pronosticar estamos suponiendo uno de tantos caminos que pudo haber tomado el proceso y, gracias a la estacionaridad, estos diferentes caminos no están tan alejados el uno del otro. Aunado a esto, en algún momento supusimos que el modelo se repite cada cierto ciclo de tiempo, así que si no atinamos en el primer período, tenemos la posibilidad de hacerlo en el siguiente.

De hecho las ventajas de la estacionaridad se hacen patentes con la condición de θ menor a 1, ya que hace a los intervalos de confianza más pequeños y a los pronósticos más confiables: trata de dar un respiro en cada nueva etapa, perdona un poco el hecho de que nos hallamos equivocado en el pasado (que es lo más probable) y con θ menor que uno, el error que cometimos pasa a valer menos de lo que valió en el pasado.

El error de pronóstico:

$$MSE = \sum_{i=1}^n \frac{(y_i - F_i)^2}{n}$$

donde F_i es el pronóstico

Si tomamos por hecho que los errores de pronóstico se comportan como una Normal, es decir, que se comporten alrededor de cero, podemos utilizar al MSE para formar los intervalos de pronóstico, ya que provee de un estimador de la varianza del error de pronóstico a una etapa.

Entonces el 95% de los datos debería estar en $y_i \pm 1.96(S)$. El intervalo de pronóstico a una etapa está dado por:

$$Y_{T+1,T} \pm Z [MSE]^{1/2}$$

Generalizando para cualquier horizonte de pronóstico:

$$y_{T+h,T} \pm Z [1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2]^{1/2}$$

ψ puede representar ya sea a ϕ (representación de promedios móviles de un AR) o a θ (los factores tomados directamente de como viene representado el modelo MA).

El valor de Z determina el ancho y la probabilidad del intervalo. Por ejemplo 1.96 es el intervalo para un 95% de confianza. Esto es, el intervalo tiene una probabilidad del 95% de contener el valor verdadero.

| Z | Probabilidad |
|-------|--------------|
| 0.674 | 0.50 |
| 1.000 | 0.68 |
| 1.150 | 0.75 |
| 1.282 | 0.80 |
| 2.576 | 0.99 |

Fuente: (Makridakis, 1979)

INTERVALOS DE CONFIANZA PARA PRONÓSTICOS DE PROCESOS MA(Q)

En el caso de promedios móviles, la única diferencia es que el conjunto de información está limitado a una cantidad q en vez de un infinito número de e_t .

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

$$y_{t+h} = e_{t+h} + \theta_1 e_{t+h-1} + \theta_2 e_{t+h-2} + \dots + \theta_q e_{t+h-q}$$

Contamos con $q-1$ datos (choques) para pronosticar y dependiendo del horizonte de pronóstico algunos o todos puede desaparecer y volverse ceros al realizar la proyección. Esto es, si $h > q$ todos los datos con los que contábamos desaparecerán, y si $h \leq q$ tendremos al menos un parámetro para poder hacer un pronóstico diferente a cero

$$y_{T+h,T} = \begin{cases} \theta_h e_T + \theta_{h+1} e_{T-1} + \theta_{h+2} e_{T-2} + \dots + \theta_q e_{T+1-h} & \text{para } h = 1, 2, 3, \dots, q \\ 0 & \text{para } h = q+1, q+2, \dots \end{cases}$$

Al encontrar el error de pronóstico para h etapas en el futuro encontraremos el intervalo de confianza

$$e_{T+h,T} = e_{T+h} + \theta_1 e_{T+h-1} + \theta_2 e_{T+h-2} + \dots + \theta_{h-1} e_{T+1}$$

$$\text{var}(e_{T+h,T}) = \begin{cases} \delta^2 & \text{para } h = 1 \\ \delta^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_{h-1}^2) & \text{para } h = 2, 3, 4, \dots, q \\ \delta^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) & \text{para } h = q+1, q+2, \dots \end{cases}$$

IX.2 Cálculo de los pronósticos: Ajuste de los modelos para su operatividad

Ya que contamos con esta teoría es justo poner manos a la obra, es decir, aterrizar la teoría.

Para pronosticar vamos a utilizar todas las fórmulas que hemos visto hasta ahora, sólo que los e_t se reemplazarán por sus estimados \hat{e}_t , así como los ϕ y θ por sus estimados.

Estos dos últimos se estiman por el método del error cuadrático mínimo, MSE. Estudiemos ahora la estimación de los errores de pronóstico.

ESTIMACIÓN DE LOS RESIDUALES

Tomemos un ejemplo muy sencillo, un AR(1).

$$y_t = \phi y_{t-1} + e_t$$

Volvamos a realizar su pronóstico a una etapa:

Primer paso: $y_{T+1} = \phi y_T + e_{T+1}$

En el siguiente paso ya sabemos que $e_{T+1,T}$ va a desaparecer.

Segundo paso: $y_{T+1,T} = \phi y_T$

Y el error de pronóstico para una unidad de tiempo es:

$$e_{T+1,T} = y_{T+1} - \phi y_T$$

¿Por qué era que desaparecía el $e_{T+1,T}$, recuerda el lector?, porque no contamos con la observación de y_{T+1} . Si la tuviéramos, podríamos *calcular* $e_{T+1,T}$, no hacerla cero. Esto se puede lograr si pronosticamos dentro de la muestra.

Por ejemplo, se puede pronosticar para y_2 con el primer dato de la serie.

$$y_{2,1} = \phi y_1$$

Entonces

$$y_{2,1} = \phi y_1$$

y ahora sí, podemos obtener el error:

$$\begin{aligned} \hat{e}_{2,1} &= y_2 - y_{2,1} \\ &= y_2 - \phi y_1 \end{aligned}$$

Al contar con el valor y_2 simplemente lo sustituimos y obtenemos así el primer residual \hat{e}_1 y podemos establecer las diferentes particularizaciones del modelo AR(1) $y_{T+1} = \phi y_T + e_{T+1}$

$$\begin{aligned} y_2 &= \phi y_1 + \hat{e}_2 \\ y_3 &= \phi y_2 + \hat{e}_3 \\ &\text{etc.} \end{aligned}$$

Así que mientras estemos dentro de la serie, no es necesario *generar* las e_t como v.a.i.i.d. $N(0, \sigma^2)$, por el contrario, las calculamos y deberían comportarse como tales. Si no lo hacen significa que el modelo está mal.

Para cotejar los residuales con la distribución normal podemos realizar una análisis de residuales. Para empezar graficamos los residuales contra el tiempo y estos deberían tener la apariencia de ruido blanco, si no lo tienen algún patrón se nos ha escapado. Los analizamos como si se tratar de una serie de tiempo más, buscando patrones en sus ACF, PACF y periodogramas y en sus serie original. En el capítulo de la metodología de Box-Jenkins se verá a detalle cada uno de los pasos en el análisis de residuales.

La razón por la que es necesario analizar los residuales deriva del hecho que el algoritmo de estimación minimiza los residuales en base a la cantidad de parámetros que le indiquemos, pero no se preocupa por eliminar las interrelaciones dentro de los residuales.

INICIALIZACIÓN DE LOS RESIDUALES

Existe un detalle. Si intentamos calcular el residual número uno, $\hat{\epsilon}_1$, surge la cuestión de qué valor debemos poner como y_0 , ya que no se cuenta con él. De hecho, el problema de datos faltantes va acompañando a la metodología desde que empezamos, ya que en un principio dijimos que el presente está determinado por la suma *infinita* de choques aleatorios del pasado (el proceso lineal general), lo que en la práctica es imposible.

$$y_1 = \phi y_0 + \epsilon_1$$

$$y_0 = ?$$

Una solución sencilla es sustituir el primer pronóstico por el valor observado, que quiere decir que ϵ_1 sea igual a 0. Esta solución es suficiente siempre y cuando se cuente con una serie lo suficientemente grande (Box y Jenkins, 1970). A las sustituciones que hacemos en estos casos para datos pre-históricos se llama *inicialización de la serie*.

IX.3 Actualización de pronósticos

La actualización de los pronósticos tiene como fin contar con pronósticos que utilicen la información más actual, pero sin sacrificar los pronósticos ya realizados. Tienen como principio el que un pronóstico que fue hecho para un cierto horizonte h se puede mejorar con cada dato más que se agregue a la serie. En otras palabras, el pronóstico realizado el día de hoy lunes para lo que sucederá dentro de una semana (lunes a lunes), no está mal, pero se puede hacer que se acerque al valor que en realidad sucederá si observamos e integramos al modelo lo que sucederá mañana (martes).

No se elimina ningún dato y se actualiza la información.

Recordemos que el modelo general de pronóstico es el que está dado por:

$$y_{t+h} = \psi_0 \epsilon_{t+h} + \psi_1 \epsilon_{t+h-1} + \psi_2 \epsilon_{t+h-2} + \dots$$

Se puede extender fácilmente sin perder validez, a un horizonte de $h+1$ en lugar de h :

$$y_{t+h+1} = \psi_0 \epsilon_{t+h+1} + \psi_1 \epsilon_{t+h} + \psi_2 \epsilon_{t+h-1} + \dots$$

Proyectamos sobre 2 diferentes conjuntos de información Ω_t y Ω_{t+1} y observamos qué sucede.

$$y_{T+h+1, T+1} = \psi_h \epsilon_{T+1} + \psi_{h+1} \epsilon_T + \psi_{h+2} \epsilon_{T-1} + \dots$$

$$y_{T+h+1, T} = \psi_{h+1} \epsilon_T + \psi_{h+2} \epsilon_{T-1} + \dots$$

Al restar estas dos ecuaciones obtenemos el *actualizador de pronósticos*, por llamarlo de alguna manera.

$$y_{T+h+1, T+1} = \psi_h \epsilon_{T+1} + y_{T+h+1, T}$$

Digamos que ya obtuvimos hoy el pronóstico para dentro de una semana (lunes y lunes). Asimismo ya obtuvimos también el pronóstico para una semana y un día (lunes de esta semana y martes de la siguiente semana). El día de mañana podremos hacer el pronóstico para el mismo martes, al igual que lo hicimos hoy para el siguiente lunes. Pero una forma alternativa, con el actualizados de pronósticos, es sumándole al pronóstico que

hicimos hoy para el segundo lunes, la información que recibamos el día de mañana. Así se ahorra una gran cantidad de operaciones.

**TESIS CON
FALLA DE ORIGEN**

Capítulo X. Metodología de Box-Jenkins*

En 1976, George E. P. Box y Gwilym M. Jenkins publicaron su libro "*Time Series Analysis, Forecasting and Control*"**, en el cual se mencionan aplicaciones prácticas del pronóstico de series de tiempo. Este libro generó una reacción en cadena que dio cabida a que muchos investigadores probaran, estudiaran y publicaran acerca de la eficiencia y alcances de este método, además de que proveyó a los matemáticos de una prueba empírica de los alcances eficientes que la metodología cuantitativa puede tener.

El método de Box-Jenkins consiste en extraer los movimientos predecibles de los datos observados. La serie de tiempo se descompone en varios componentes que se pueden denominar "filtros". Cada filtro intenta *filtrar* precisamente uno de los componentes de la serie de tiempo. En los métodos de suavizamiento y descomposición los componentes eran la tendencia, estacionalidad, ciclicidad y el componente aleatorio. Ahora, bajo el esquema del método de Box-Jenkins, aunque no se dejan de considerar los anteriores componentes ya que no pierden validez sus definiciones, los componentes del método se tornan en tres: el componente de media móviles, el autorregresivo y el de integración. Es decir, la serie de tiempo se ve bajo la misma óptica que con los métodos de descomposición, a saber: todo proceso estocástico genera una serie de tiempo que contiene una cierta combinación de los componentes señalados y se debe de encontrar cuánto influye cada componente en ella. Con la única diferencia que los componentes cambian de nombre.

El componente autorregresivo es un componente que indica cuánto depende el valor de la serie en el momento T del valor de esa misma serie en el momento T-n en el pasado.

El componente de medias móviles indica cuánto depende el momento actual T de los golpes aleatorios del pasado.

Y el componente de integración indica si los dos anteriores componentes están tratando con la serie original o con una o varias diferencias de la serie original. Por ejemplo, una serie diferenciada*** genera otra serie de tiempo, que bien puede no tener tendencia, cuando la original sí la tenía. O puede carecer de varianza creciente, cuando la original sí tenía una varianza no constante. En corto se analizará la importancia de una tendencia inexistente, es decir una media constante, y una varianza constante, es decir que la serie tenga una propiedad llamada homoscedasticidad.

A muy *grosso modo*, los 4 pasos del método son:

Identificación.- Ésta es la etapa donde se vuelven estacionarios a los datos y se asigna un modelo tentativo, es decir, provisional. La estacionaridad o la ausencia de

* George E.P. Box nació el 18 de octubre de 1819 en Inglaterra. Fungió como químico para la Armada Británica de su país durante la segunda guerra mundial. Después de la guerra recibió su título de matemáticas y estadística en Inglaterra. Durante el tiempo que trabajó con ICI Ltd, el gigante inglés de los químicos en los años 50, comenzó a trabajar como profesor invitado en el Instituto de Estadística de la Universidad de Carolina del Norte en los Estados Unidos. En 1960 se mudó a Wisconsin, donde fungió como el presidente del departamento de Estadística. Recibió la Medalla del Imperio Británico en 1946 y la Medalla Shewhart en 1968, entre otras.

** Box and Jenkins, "Time Series Analysis, Forecasting and Control," Holden Day, 1976, Revised Edition.

*** Ver el tema 1.3 para una referencia de lo que es una serie diferenciada.

ésta se verifica tanto de manera visual, con la gráfica de la serie, como con las gráficas de los coeficientes de autocorrelación y autocorrelación parcial.

Estimación.- En la etapa de estimación se inicializan los parámetros para después obtener los estimados de los mismos. Se inicializan normalmente con 0 ó 0.1 y después la computadora hace el resto.

Diagnóstico.- Se revisa la proximidad entre pronósticos y los datos reales. Se analizan los residuales $y_t - \hat{y}_t$ como si fueran datos de la serie, observando sus gráficas de ACF y PACF. Si estas gráficas son diferentes a las del ruido blanco, es necesario reconsiderar el modelo.

Pronóstico.- Una vez que contamos con el modelo adecuado, se integran los datos (en caso de que se haya diferenciado) y se aplica la operación inversa de la que se utilizó para hacer estacionaria a la serie. Por ejemplo, si usamos \ln para hacerla estacionaria, pronosticamos con e^x . Después de esto sólo hay que sustituir valores en el modelo. En esta etapa es también cuando se calculan los intervalos de confianza.

La metodología de Box-Jenkins es una sistematización de todo lo que hemos visto en la segunda parte de este trabajo. Su principio se basa en que las observaciones y_t deben ser independientes entre sí y los e_t deben ser pequeños, independientes e impredecibles, es decir, ruido blanco. Si no es así, debemos utilizar un modelo conformado por valores rezagados de y_t y e_t .

La metodología de Box y Jenkins es iterativa; primero trata de captar el patrón en los datos. Después coteja el modelo con los datos históricos para verificar que efectivamente se estén generando residuales como ruido blanco. Si el modelo no es el adecuado, se procede a encontrar otro modelo que funcione mejor.

El proceso se repite hasta que se encuentra un modelo satisfactorio, siempre guardando en mente el principio de parsimonia. Todo este procedimiento se aplica a datos estacionarios y el modelo resultante es del tipo

$$y_t = C + \sum_{i=1}^{\infty} \psi_i e_{t-i}$$

donde c es una constante para ajustar al recorrimiento vertical.

REQUISITOS

Si no existe una suficiente cantidad de datos, podría resultar tan efectivo utilizar los métodos de suavizamiento o descomposición. Para Box y Jenkins se presupone una estacionaridad débil, datos recolectados en intervalos equidistantes de tiempo y se necesitan por lo menos 50 datos.

El método también requiere que no existan datos faltantes en la serie. La selección del intervalo de tiempo entre cada dato es importante. Si los datos cambian mes con mes pero solo contamos con un dato recolectado por año, nos resultará imposible pronosticar los cambios mensuales. Y viceversa, si nos interesa pronosticar cambios que suceden cada 11 años, necesitaremos una gigantesca cantidad de datos. De no contar con ellos, no podremos captar el patrón que nos interesa.

VENTAJAS Y DESVENTAJAS

Una desventaja del método es que puede ser tardado y tedioso ya que trabaja de forma iterativa. Aún así, con las velocidades que alcanzan las actuales computadoras personales, el tiempo de trabajo se ha reducido al del razonamiento, por nuestra parte, del modelo. así como sus pruebas de ajuste. Las operaciones se hacen prácticamente de manera instantánea.

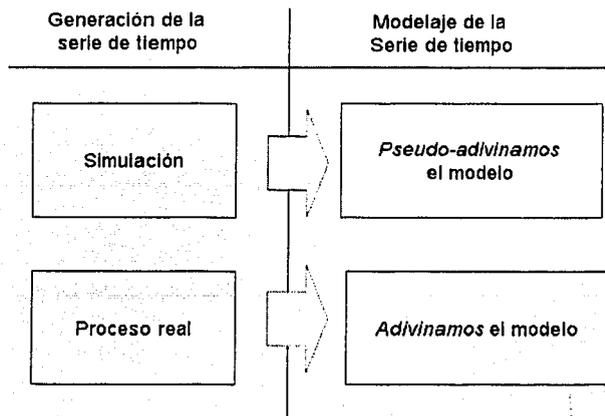
Otra desventaja es que se necesitan muchos datos, ya que no se puede trabajar con menos de 60. Dado que los estimadores se obtienen de manera no lineal, el esfuerzo computacional puede resultar considerable y los estimadores pueden arrojar más de un resultado. Además, los resultados del método deben ser interpretados subjetivamente, lo cual puede hacer parecer al método más un arte que una ciencia.

Una de las grandes ventajas que tiene el método de Box-Jenkins sobre los demás métodos cuantitativos. es que para el corto plazo es muy difícil de vencer en cuanto a exactitud. Tiene también la ventaja de ser mucho menos sensible a las suposiciones de la naturaleza subyacente de las fluctuaciones de los datos que muchos otros métodos.

SIMULACIÓN VS. SERIES DE TIEMPO REALES

Para dar un mejor seguimiento a cada paso de la metodología, iremos tratando con dos enfoques distintos. El primero será el de simular nosotros mismos un proceso estocástico. una serie de tiempo. Con ésta trabajaremos como si se tratara de un proceso real, aunque de antemano sabemos el resultado que deberíamos obtener. El otro, obviamente, es modelar series de tiempo con datos reales. donde desconocemos por completo el proceso generador.

El fin de hacer esto es el de conocer e identificar las discrepancias que surgen al momento de estimar. Por ejemplo, cuando generamos la serie de tiempo y la estimamos con el mismo modelo con el que fue generado, podremos ver que los valores de los estimadores no son idénticos a los que utilizamos. Esto se debe, entre otras cosas, a los sesgos y problemas inherentes de consistencia de los estimadores. Este ejercicio, práctico, ya no mental, permitirá que nos familiaricemos con las dificultades con las que nos encontraremos al tratar con procesos reales.



TESIS CON FALLA DE ORIGEN

X.1 Identificación

La identificación es la etapa más importante de toda la metodología y a la vez la más difícil. Es bueno contar con una actitud precavida a través de toda la identificación, ya que estamos trabajando con una muestra y ésta podría estar indicando algo muy diferente a lo que en realidad está sucediendo. Por eso se debe concentrar en las características generales al principio y después podremos afinar detalles. No debemos sentir apego inexplicable por un modelo, es decir, el modelo que seleccionemos es solamente tentativo y podremos en las siguientes etapas desacreditarlo, con lo cual tendremos que regresar a esta etapa a buscar uno mejor.

En esta etapa el fin es asignar un modelo, el mejor que podamos tomando muy en cuenta la parsimonia porque los coeficientes se obtendrán automáticamente por la computadora tomando en cuenta el modelo que nosotros escojamos.

Existen 2 pasos básicos para la identificación de un modelo apropiado de Box y Jenkins: transformar los datos, si es necesario, para que estos sean estacionarios y determinar el modelo tentativo analizando las funciones de autocorrelación y autocorrelación parcial, es decir, determinar p.d.q.

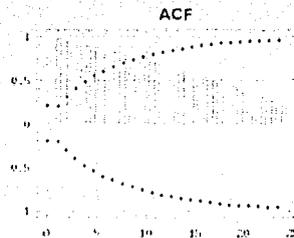
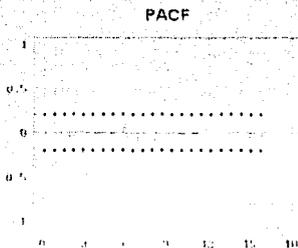
PASOS A SEGUIR EN LA IDENTIFICACIÓN

1.- Empezamos viendo la gráfica de la serie. Identificamos visualmente si tiene tendencia o varianza creciente. Una vez que concluimos si la serie es estacionaria o no, verificamos esto con las ACF y PACF de la serie.

Características de la ACF y PACF de los procesos estacionarios

| Proceso | ACF | PACF |
|--------------|--------------------|--------------------|
| $AR(p)$ | Decae poco a poco | Se trunca en $k=p$ |
| $MA(q)$ | Se trunca en $k=q$ | Decae poco a poco |
| $ARMA(p, q)$ | Decae poco a poco | Decae poco a poco |

Pero es muy probable que nos encontremos gráficas de ACF y PACF como las siguientes:



TESIS CON
FALLA DE ORIGEN

Una autocorrelación parcial muy alta, de casi 1, en $k=1$ y una autocorrelación que tarda mucho en disminuir nos indican que hace falta diferenciar*.

Los correlogramas no van a ser idénticos a los teóricos poblacionales, por eso debemos buscar cualquier semejanza que exista. El modelo subyacente se puede considerar que está escondido y lo estamos buscando.

Una vez que hemos decidido el tipo de modelo que creemos adecuado se debe intentar siempre por el menor orden e ir añadiendo paulatinamente más parámetros. Por ejemplo si consideramos que se trata de un AR(3), deberíamos comenzar por añadir el AR(1) y verificar la significación de este parámetro. Después avanzamos al AR(2) y así hasta llegar al 3. Esto ayuda a evitar sobrediferenciar.

Ahora es seguro mencionar que muchos autores alegan que los peligros de sobrediferenciar son menores a los de subdiferenciar. Así que si nos encontramos con casos en donde hay ambigüedad con respecto a si se debe diferenciar o no, por lo general se recomienda la diferenciación. Claro que esto se debe de hacer con inteligencia, sabiendo que así aumentará el número de parámetros y disminuirá el número de datos.

El utilizar por sí solas a la ACF y PACF tiene sus debilidades, ya que se necesita hacer uso del juicio para establecer el orden de diferenciación y cuando se trata de un proceso mezclado ARMA, los decaimientos simultáneos de ambas gráficas impiden hacer una identificación clara del modelo que se trata.

2.- Aplicamos a nuestro juicio las transformaciones o diferencias necesarias para eliminar la no-estacionaridad. Si aplicamos primero las diferencias y después las transformaciones es lo mismo que si primero aplicáramos las transformaciones. Esto es posible debido a la linealidad del operador diferencia Δ .

$$\log[\Delta(y_t)] = \log(y_t) - \log(y_{t-1})$$

$$\Delta[\log(y_t)] = \log(y_t) - \log(y_{t-1})$$

Por otro lado, pueden surgir valores negativos al realizar las diferencias, por lo que algunas veces es imperativo que se apliquen las transformaciones antes que las diferencias.

Recordamos también que una serie con varianza no constante por lo general necesita una transformación logarítmica.

3.- El último punto es el de decidir si vamos a considerar la constante o no. Hasta ahora hemos estado suponiendo una igual a cero, pero en la realidad se tiene que probar la hipótesis de si se considera en el modelo o no. Si la consideramos siempre dentro del modelo, después se puede desechar si resulta ser estadísticamente no significativa.

El punto de la constante en el modelo necesita quedar muy claro antes de proseguir. En un principio se desechó por completo, no porque siempre vaya a ser igual a cero, sino por simplicidad. La ecuación de diferencias se iguala a la constante para obtener la raíz, independientemente de si la constante es cero o no. Esto de obtener la raíz es solamente

* Ver capítulo VII en el tema "Cómo se ven los procesos estacionarios" para ver por qué es necesario diferenciar en este punto

para fines didácticos, ya que a la computadora no le importa que el modelo sea o no estacionario.

Lo que puede resultar confuso es que desde que vimos que la diferenciación eliminaba la constante cuando se trataba de un modelo con tendencia estacionaria (es decir de caminata aleatoria con deriva) no consideramos que podríamos volver a encontrarnos con dicha constante. Pero después de diferenciar, efectivamente sí puede quedar una constante.

¿Sólo porque un modelo contenga una constante implica que el modelo deba tener tendencia?. No.

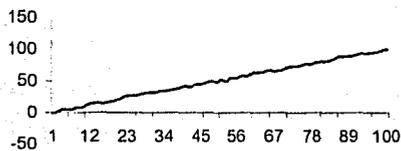
De los modelos AR, MA y ARMA, ¿cuál no es recursivo? Solamente los MA puros (es decir, MA que no contienen una parte AR ni una parte I), porque cualquiera que contenga aunque sea una parte AR, se contiene a sí mismo como variable, haciendo que cualquier constante que se añada al modelo le proporcione el potencial de tener tendencia.

RECURSIVIDAD Y MODELOS ARIMA

Para que un modelo MA(q) tenga tendencia, no hay otra posibilidad más que se esté sumado con el tiempo (t), ya que el tiempo siempre incrementa. Por ejemplo, un MA(1):

$$y_t = \theta e_{t-1} + e_t + t$$

Se vería como la siguiente figura:



Primeras diferencias:

$$\begin{aligned} \Delta y_t &= \theta e_{t-1} + e_t + t - (\theta e_{t-2} + e_{t-1} + (t-1)) \\ &= \theta e_{t-1} - \theta e_{t-2} + e_t - e_{t-1} + 1 \\ &= (\theta - 1)e_{t-1} - \theta e_{t-2} + e_t + 1 \end{aligned} \quad (10.1)$$

La primer diferencia de un modelo MA(1) con tendencia da como resultado un modelo estacionario MA(2) con constante, cuya gráfica es:



TRABAJOS CON
FECHA DE ORIGEN

Visualmente la serie es una estacionaria y recorrida en una unidad (circunda a la constante -1), como efectivamente lo marca el modelo.

Para las 2 anteriores gráficas se utilizó $\theta = 1$. Si resolvemos las raíces de la ecuación de diferencias para este valor de θ se verá que el modelo es invertible, además de estacionario*.

$$\begin{aligned} \text{MA}(2) : y_t &= (\theta-1)e_{t-1} - \theta e_{t-2} + e_t + 1 \\ y_t &= B(\theta e_{t-1} - \theta^2 e_{t-2}) + e_t + 1 \\ y_t &= e_t [B(\theta-1) - (B^2\theta) + 1] + 1 \end{aligned}$$

Ahora, igualamos todo el factor de e_t (que es la ecuación de diferencias Θ) a cero, para poder encontrar las raíces.

$$\Theta = [B(\theta-1) - (B^2\theta) + 1] = 0$$

En este caso, dado que se tiene un término que no es factor de e_t , uno, se iguala a uno, no a cero. Sustituyendo a θ por su valor real, uno, da como resultado

$$-1 = (1 - B^2)$$

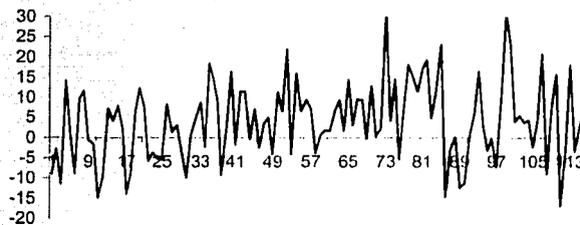
de donde

$$B = \pm \sqrt{2}$$

Las raíces inversas son $B = \pm 0.7071$ que están ambas dentro del círculo unitario, lo que prueba que el modelo es invertible y por ende estacionario. Todos los modelos MA son estacionarios, independientemente de sus valores de θ y de que contengan o no constante.

Ahora tomemos un MA con raíces inversas fuera del círculo unitario. Es decir, un modelo estacionario pero no invertible.

$$y_t = e_t + 8e_{t-1} + 3$$



Obtenemos la raíz de la ecuación de diferencias

$$-3 = (1 + 8B)$$

$$B = -0.5$$

$$\text{Raíz inversa } (1/B) = -2$$

$$|-2| > 1$$

TESIS CON FALLA DE ORIGEN

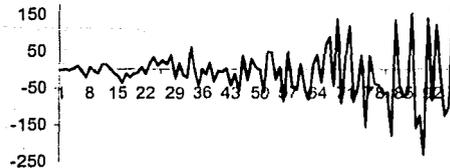
* Para ver por qué se hace uso del requisito de la raíz inversa, referirse al capítulo VII en el tema de Condiciones de Invertibilidad para procesos AR y MA.

Y observamos que la gráfica es estacionaria pero el modelo no es invertible.

La única otra manera de tener un MA no estacionario es que contenga varianza creciente. Esto se puede lograr multiplicando los efectos por el tiempo, en vez de sumarlos.

Por ejemplo

$$y_t = (e_t)(e_{t-1})(t)$$



En estos casos, se puede eliminar la heteroscedasticidad* con logaritmos, así se transformaría el modelo a uno aditivo.

$$\log(y_t) = \log(e_t) + \log(e_{t-1}) + \log(t)$$

CONCLUSIONES DE PROCESOS MA

En los procesos MA es indiferente si se contiene una constante sumada o no, ésta se puede retirar del modelo sin que se vea afectado. Para ser invertibles los procesos MA necesitan satisfacer la condición de que las raíces inversas estén dentro del círculo unitario, pero por formación, todos los MA son estacionarios. Las únicas dos maneras de tener MA no estacionarios es que se quebranten las reglas fundamentales de los modelos ARIMA: que se esté sumando una "constante" que no es constante (la variable tiempo) o que se estén multiplicando los efectos incluyendo al tiempo (un modelo que de entrada no es un modelo ARIMA aditivo como los que hemos visto hasta ahora).

SIMULACIÓN DE MODELOS MA: SERIES NO ESTACIONARIAS QUE RESULTAN SER ESTACIONARIAS

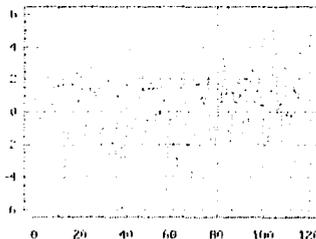
Generemos un MA(1) con constante 0 y coeficiente $\theta=2$

$$y_t = 2e_{t-1} + e_t$$

$$(2L+1)=0$$

$$B = -.5$$

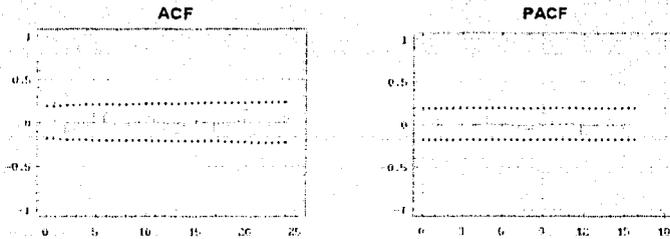
Vemos enseguida que no es invertible. Pero su gráfica es la de un proceso estacionario



TESIS CON
FALLA DE ORIGEN

* Varianza no constante

Más aún, si vamos más allá y obtenemos sus ACF y PACF, éstas nos están indicando que el proceso ¡sí es invertible!



Si desconociéramos el modelo original, tendríamos que empezar por estas dos gráficas y crear un modelo tentativo a partir de ellas.

PROCESOS AR Y TENDENCIA

Ahora, están los AR(p). En estos procesos no es tan fácil retirar la constante del modelo, porque ésta podría estar formando activamente parte del modelo, es decir, podría estar ayudando a realizar un mejor pronóstico, ya que los AR son recursivos y la constante no está involucrada de manera lineal, sino exponencial.

Es cierto que la recursividad puede hacer que una constante, por pequeña que sea, haga que el proceso se vuelva inestable. Pero no porque contenga constante un proceso AR implica que necesariamente se trate de un proceso no estacionario. Por eso es necesario probar su significación.

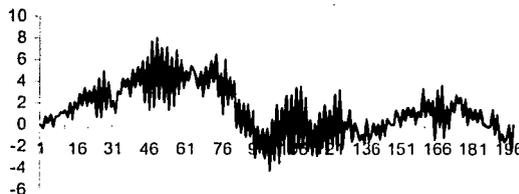
Por ejemplo, tomemos un AR(1) estacionario y convergente

$$y_t = 0.95y_{t-1} + e_t$$

Obtenemos la raíz de la ecuación de diferencias

$$(1 - 0.95L) = 0$$

$$B = \frac{1}{0.95}$$



TESIS CON FALLA DE ORIGEN

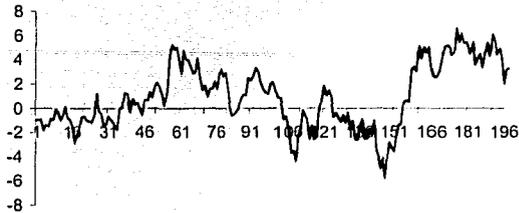
La raíz inversa (1/B) es 0.95, convergente. Pero si añadimos una constante, aunque sea muy pequeña, las cosas cambian.

$$y_t = 0.95y_{t-1} + e_t + 0.1$$

$$(1 - 0.95L) = 0.1$$

$$B = \frac{0.9}{0.95} = 0.947$$

La raíz inversa es 1.055 que ya no cumple con convergencia, y eso que es una constante muy pequeña ¡de 0.1!



Podemos observar que su serie se parece mucho a la de una caminata aleatoria. Ahora vemos otro modelo AR(1) que sí es convergente

$$y_t = 0.1y_{t-1} + e_t$$

$$0 = y_t [1 - (0.1)B] + e_t$$

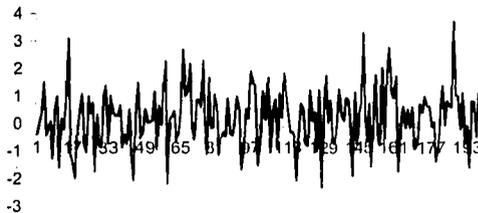
Iguamos las ecuación de diferencias a cero, da como resultado

$$1 - B(0.1) = 0$$

$$B = -1/0.1$$

$$B = 10$$

TESIS CON FALLA DE ORIGEN



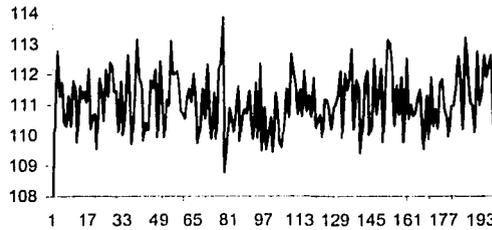
La raíz inversa (1/B) está 0.1 dentro del círculo unitario. Agreguémosle una constante muy grande, para forzarla a volverse no estacionaria

$$y_t = 0.1y_{t-1} + e_t + 100$$

$$(1 - 0.1B) = 100$$

$$B = \frac{-99}{0.1} = -990$$

Con raíz inversa -0.001 que satisface la estacionaridad a pesar de que escogimos a propósito una constante muy grande.



TESIS CON
FALLA DE ORIGEN

Visualmente la serie con constante es la de un proceso estacionario.

CONCLUSIONES DE MODELOS AR

Podemos concluir que existe un balance muy delicado entre los coeficientes y la constante y no se puede concluir nada si, al evaluar el modelo de la serie, obtenemos una constante aparentemente grande. Necesitamos probar siempre estadísticamente su significancia: si se incluye la constante o no en el modelo.

Las conclusiones de los AR se extienden a los ARMA.

CONCLUSIONES ARMA

Pero nótese que son dos puntos diferentes de los que se acaba de hablar: uno es el de procesos estocásticos cuya ecuación de diferencias tiene como raíz una raíz que no satisface las condiciones de estacionaridad. Ese ya se había mencionado antes cuando se explicó que *todos* los modelos MA son, por formación, estacionarios, pero que para poder ser útiles en la vida real debían satisfacer la condición de convergencia.

El otro punto es el del **balance**: la estabilidad que puede tener un proceso AR a pesar de contar con una constante, que supusimos en un principio que podría traer un crecimiento desmesurado debido a que los modelos AR son autorreferenciales, es decir, recursivos. Pero se vio, a través de las ecuaciones de diferencias y de las gráficas, que esto no es así necesariamente y existe precisamente un equilibrio que se debe mantener para que el proceso no crezca exponencialmente.

X.2 Estimación

El algoritmo de obtención de parámetros se basa en la minimización de la varianza.

Un modelo ARIMA se puede especificar como

$$\Phi(B)z_t = \Theta(B)e_t$$

Donde $z_t = \Delta^d y_t$. Al despejar e_t se obtienen los residuales a ser minimizados, una vez que se hayan sustituido los vectores Φ y Θ por sus estimados $\hat{\Phi}$ y $\hat{\Theta}$.

$$\hat{e}_t = \hat{\Theta}^{-1}(B) \hat{\Phi}(B)z_t \quad (10.2)$$

e_t sería ruido blanco si conociéramos los valores reales de los parámetros. Dado que no se pueden conocer, suponemos como *reales* aquellos vectores que nos proporcionen unos

residuales que se comporten como ruido blanco y cuya suma de residuales al cuadrado* sea mínima

$$\sum_{i=1}^T \hat{e}_i^2 = S(\hat{\Phi}, \hat{\Theta})$$

Se presentan 3 obstáculos naturales para resolver para los vectores requeridos así como se acaba de plantear:

- 1) Como se puede apreciar, 10.2 es no lineal si contiene al menos un término MA
- 2) Para iniciar el procedimiento de obtención de parámetros, hacen falta valores prehistóricos o no observados $z_0, z_{-1}, \dots, e_0, e_{-1}, e_{-2}, \dots$
- 3) Existe más de un método para obtener los estimadores.

Se empezará por tratar este último punto y se trabajará hacia el primero.

MÉTODOS PARA ESTIMAR LOS PARÁMETROS DE MODELOS ARIMA

Pueden existir dos razones para que los residuales no se comporten como ruido blanco:

- a) Un modelo adecuado pero una estimación de parámetros pobre
- b) El modelo es inadecuado

Existen tres métodos para estimar los parámetros: Método de Máxima Verosimilitud, Mínimos Cuadrados Condicionales y Mínimos Cuadrados Incondicionales. El más importante es el de Máxima Verosimilitud.

Una de las razones para que la función de verosimilitud sea tan importante en la teoría de la estimación es por el llamado *principio de verosimilitud*, que dice que, suponiendo que el modelo sea el correcto (es decir, descartando el punto 2), todo lo que los datos *están tratando de decirnos* acerca de los parámetros está contenido en la función de verosimilitud y todo lo demás es irrelevante.

ESTIMACIÓN DE LOS PARÁMETROS Θ Y Φ A PARTIR DEL CÁLCULO DE RESIDUALES

La serie original y_t tiene $T+d$ elementos, porque se eliminan d datos cuando se diferencia d veces, y la serie ya estacionaria tiene T datos. Suponiendo que esta serie *diferenciada* está siendo generada por un modelo ARIMA, se deben entonces estimar los valores de $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ en

$$e_t = z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2} - \dots - \phi_p z_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} \quad (10.3)$$

donde $z_t = \Delta^d y_t$.

Si se sustituyen directamente los valores de e_t y z_t del lado derecho, se pueden calcular los residuales e_t del lado izquierdo. Esto no se puede hacer *al principio*, es decir, en el tiempo $t=1$, ya que hacen falta $p+q$ valores no observados: $z_0, z_{-1}, z_{-2}, \dots, z_{-p}, e_0, e_{-1}, e_{-2}, \dots, e_{-q}$. Si se obtuvieran valores razonables para estos datos faltantes se podrían sustituir en la fórmula 10.3 directamente y recorrer con ella toda la serie, desde $t=1$ hasta $t=T$, para

* Por como funciona el MSE

determinados parámetros Φ y Θ , obteniendo así los residuales del modelo deseados y a la vez poder obtener iterativamente los parámetros que minimicen la suma de los residuales.

¿QUÉ ES LA FUNCIÓN DE VEROSIMILITUD?

Una función de verosimilitud es una fórmula de probabilidad. La estimación por máxima verosimilitud usualmente comienza por una función de verosimilitud que se debe maximizar o minimizar.

Cuando las observaciones son independientes unas de otras, como en el tiro de un dado, la probabilidad de las múltiples ocurrencias consecutivas está dada por el producto de sus probabilidades individuales. La probabilidad de obtener un 3 en un tiro es de 1/6. La probabilidad de obtener dos 3 en dos tiros es de $(1/6)(1/6) = 1/36$.

En una serie de tiempo, si existe ergodicidad*, cada e_t es independiente y distribuida normalmente y, por tanto, tiene una distribución de probabilidad $[N(0, \sigma^2)]$ y una función de densidad de probabilidad de

$$p(e_t) = \frac{e^{-\frac{e_t^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

donde la e en la base del numerador es el número irracional e . Se puede eliminar el término $1/\sqrt{2\pi}$ (ya que es una constante y no afecta la razón de cambio) y el producto de T golpes aleatorios se puede expresar entonces como

$$\left(\frac{e^{-\frac{e_t^2}{2\sigma^2}}}{\sigma} \right)^T = \sigma^{-T} e^{-\left(\sum \frac{e_t^2}{2\sigma^2} \right)} = \sigma^{-T} e^{-\frac{S(\Theta, \Phi)}{2\sigma^2}}$$

Para que se pueda apreciar como una suma, ésta se puede reformular logarítmicamente

$$B(\Phi, \Theta, \sigma) = -T \ln(\sigma) - \frac{S(\Theta, \Phi)}{2\sigma^2} \quad (10.4)$$

Esta función se denomina de verosimilitud logarítmica "condicional", porque la suma de errores al cuadrado está *condicionada* a los valores iniciales no observables. Obsérvese que el criterio de minimización al que está sujeto $S(\Phi, \Theta)$ coincide perfectamente con el de la maximización de la verosimilitud en 10.4, dado que contiene el signo de menos en la fórmula, con lo que minimizar la suma de cuadrados equivale a maximizar la verosimilitud, siempre y cuando se considere la normalidad en los residuales. Esto simplifica los cálculos futuros porque sabiendo esto se puede trabajar con mínimos cuadrados en lugar de verosimilitud, ya que la máxima verosimilitud es difícil de manejar y se acaba de comprobar que mínimos cuadrados le es equivalente.

* Un proceso ergódico es en el cual una realización del proceso es representativo del proceso entero.

MÍNIMOS CUADRADOS INCONDICIONALES

El método de mínimos cuadrados incondicionales asigna la esperanza incondicional a los datos no observados. Así los valores de e_t faltantes se sustituyen por ceros y, en caso de que la media sea cero, los valores de z_t también serían cero. Ésta es la única diferencia con los condicionales.

INICIALIZACIÓN DE LA SERIE Y MÍNIMOS CUADRADOS CONDICIONALES

Los valores no observables se pueden reemplazar ya sea por ceros o por valores pronosticados hacia atrás. Se llaman métodos de mínimos cuadrados condicionales e incondicionales, respectivamente. Para series suficientemente grandes, los resultados son muy similares. Pero si las raíces están cercanas a la no estacionaridad, el utilizar pronóstico hacia atrás para obtener datos faltantes resulta más eficaz que si no se hiciera, ya que son valores iniciales más adaptados al proceso con el que se está trabajando.

Utilícese el siguiente ejercicio como ejemplo del procedimiento de pronóstico hacia atrás para la inicialización de valores no observables y de cómo trabajan los mínimos cuadrados condicionales.

El algoritmo de mínimos cuadrados condicionales se basa en la minimización de la varianza residual. Tómese un modelo MA(1). Primero se despeja el término del error, se calcula este error hacia atrás, empezando por el último término de la serie. Es decir, se pronostica el error hacia atrás, inicializando con cero los valores futuros no observables, para obtener así un estimado del dato o datos faltantes del pasado. En la segunda etapa se pronostican los errores hacia delante, utilizando los valores inicializados diferentes de cero. Estos errores serán los residuales que deben ser sumados y minimizados.

Se utilizarán dos nombres diferentes para los errores, uno para pronosticar hacia atrás a_t , y otro para pronosticar hacia delante e_t , con el fin de evitar confusiones.

Se está utilizando una serie de datos ya estacionaria, por lo que yendo de atrás hacia delante o de adelante hacia atrás la serie se ve igual y se comporta indistintamente.

$$z_t = (1 - \theta L)e_t = (1 - \theta F)a_t$$

donde F es el operador *salto hacia delante*. Entonces el problema a ser resuelto es, en el caso de un MA(1), obtener un valor inicial para e_0 . En el caso general ARMA se deben obtener $p+q$ valores iniciales.

$$z_t = e_t - \theta e_{t-1}$$

$$e_t = z_t + \theta e_{t-1}, \quad a_t = z_t + \theta a_{t+1}$$

$$e_1 = z_1 + \theta e_0$$

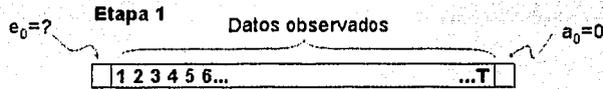
Se utiliza una serie de tamaño once para demostrar todo el procedimiento. Primero a a_{12} se le asigna el valor de cero, así como a e_0 . Después, con a_t despejado, pronosticamos hacia atrás

$$\text{it. 1: } a_{11} = z_{11} + \theta a_{12}$$

$$a_{11} = z_{11}$$

$$\text{it. 2: } a_{10} = z_{10} + \theta a_{11}$$

Se obtienen recursivamente los demás errores hasta llegar a a_0 . El objetivo es obtener un valor inicial para z_0 y por ende para e_0 . Para realizar los cálculos, tómesese a $0=0.53$ como estimado inicial.



Dirección del pronóstico (hacia atrás)

$$a_{11} = -3$$

$$a_{10} = 6 + 0.53(-3) = 4.41$$

$$a_9 = -19 + 0.53(4.41) = -16.66$$

...

$$a_1 = 3 + 0.53(1.83) = 3.97$$

Ahora, para a_0 las cosas cambian, porque ya no contamos con el valor de z_0 , por lo que se tiene que hacer el pronóstico para éste como se haría normalmente.

$$z_0 = a_0 - (.53)a_1 =$$

$$z_0 = 0 - (.53)3.97 = -2.10$$

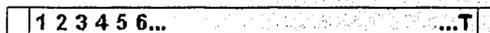
Primera etapa

| T | z_t | e_t | a_t |
|----|-------|-------|--------|
| 0 | -2.1 | 0 | 0 |
| 1 | 3 | - | 3.97 |
| 2 | 2 | - | 1.83 |
| 3 | -3 | - | -0.33 |
| 4 | 4 | - | 5.04 |
| 5 | 2 | - | 1.96 |
| 6 | 1 | - | -0.08 |
| 7 | 0 | - | -2.03 |
| 8 | 5 | - | -3.83 |
| 9 | -19 | - | -16.66 |
| 10 | 6 | - | 4.41 |
| 11 | -3 | - | -3 |
| 12 | - | - | 0 |

TESIS CON FALLA DE ORIGEN

Una vez que se cuenta con z_0 se puede obtener e_0 , asignando a e_{-1} el valor de 0 (que pertenece a los datos no observables y no estimados). Así se comienza la segunda etapa, pronosticando los residuales hacia adelante.

Etapa 2 $e_0 = -2.10$



Dirección del pronóstico (hacia adelante)

$$e_0 = z_0 + (.53)e_{-1} = -2.10$$

$$e_1 = z_1 + (.53)e_0 = 3 + (.53)(-2.10) = 1.89$$

$$e_2 = 2 + (.53)1.89 = 3.00$$

...

$$e_{11} = -3.00 + (.53)(-2.43) = -4.29$$

Segunda etapa

| T | z_t | e_t |
|----|--------|--------|
| 0 | -2.10 | -2.10 |
| 1 | 3.00 | 1.89 |
| 2 | 2.00 | 3.00 |
| 3 | -3.00 | -1.41 |
| 4 | 4.00 | 3.25 |
| 5 | 2.00 | 3.72 |
| 6 | 1.00 | 2.97 |
| 7 | 0 | 1.58 |
| 8 | 5.00 | 5.84 |
| 9 | -19.00 | -15.91 |
| 10 | 6.00 | -2.43 |
| 11 | -3.00 | -4.29 |

TESIS CON FALLA DE ORIGEN

Los residuales son elevados al cuadrado, sumados y almacenados por la computadora. Después comienza nuevamente todo el procedimiento, pero para un valor de 0 diferente. Este proceso itera hasta encontrar un valor para el parámetro que genere una suma de residuales al cuadrado mínima.

La diferencia entre utilizar mínimos cuadrados condicionales y mínimos cuadrados incondicionales se desvanece para muestras que son grandes en relación con p y q . Si la serie es pequeña en relación con p y q , se puede ganar alguna exactitud proveyendo al método de valores iniciales prehistóricos, aunque este aumento de exactitud no se puede cuantificar. Pero dado que desde un principio se estableció que la serie necesita tener más de 50 datos, y además es difícil encontrar modelos reales con p y q mayores a 3, utilizar mínimos cuadrados condicionales resulta una opción viable y evita los cálculos adicionales de la inicialización. La mayoría de los paquetes estadísticos tienen la opción de seleccionar *backcasting* o no, lo que nos permite comparar resultados.

Se pasa finalmente al punto que queda sin resolver, el punto número uno: la estimación.

ESTIMACIÓN DE MODELOS AR Y MA

La estimación de parámetros AR se puede hacer por simple regresión lineal múltiple, si tomamos a cada valor atrasado de z_t ($z_{t-1}, z_{t-2}, \dots, z_{t-p}$) como una variable independiente y a z_t como la variable dependiente. Este enfoque hace que el modelo AR sea idéntico al de la regresión.

$$\text{Modelo AR}(p): z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + e_t$$

Así que se puede formar un sistema de ecuaciones lineales, ya que para cada t va a haber una z_t diferente

$$\begin{aligned} z_1 &= \varphi_1 z_0 + \varphi_2 z_{-1} + \dots + \varphi_p z_{-p+1} + e_1 \\ z_2 &= \varphi_1 z_1 + \varphi_2 z_0 + \dots + \varphi_p z_{-p+2} + e_2 \\ &\dots \\ z_T &= \varphi_1 z_{T-1} + \varphi_2 z_{T-2} + \dots + \varphi_p z_{T-p} + e_T \end{aligned}$$

Esta matriz tiene T renglones y p columnas si se consideran los valores prehistóricos es decir, tiene la misma cantidad de renglones que datos tiene la serie. Pero si no se consideran los valores prehistóricos, tendrá T-p renglones, lo que es casi lo mismo, porque p rara vez es mayor a tres para procesos reales. Resolviendo este sistema de ecuaciones para los φ que minimicen los e_i al cuadrado se obtienen los estimados $\hat{\varphi}$. El proceso es bastante directo. Se puede entonces estimar el vector Φ con regresión lineal múltiple, ya que el modelo

$$e_i = \Phi(B)z_i$$

es de la forma general

$$z_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \dots + X_{ip}\beta_p + e_i$$

se considera a β_0 debido a la constante. Entonces podemos resolver

$$\Phi = (X'X)^{-1}X'Y$$

donde

$$X = \begin{bmatrix} z_p & z_{p-1} & \dots & z_1 \\ z_{p+1} & z_p & \dots & z_2 \\ & & \dots & \\ z_{T-1} & z_{T-2} & \dots & z_{T-p} \end{bmatrix}$$

y

$$Y = \begin{bmatrix} z_{p+1} \\ z_{p+2} \\ \dots \\ z_T \end{bmatrix}$$

ESTIMACIÓN DE MODELOS MA

Ahora, para los MA está la cuestión de la no linealidad

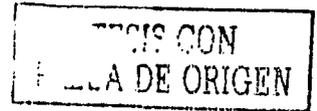
$$\hat{e}_i = \hat{\Theta}^{-1}(B)\hat{\Phi}(B)z_i$$

Este problema se resuelve haciendo uso del polinomio de Taylor, el cual está definido por

$$f(x) = P_n(x) + R_n(x)$$

donde

$$P_n(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$



Si tomamos los dos primeros términos por ser los únicos que son lineales, se puede aproximar e_i como si fuera la función $f(x)$

$$e_i \approx e_{i,0} + \sum_{l=1}^{p+q} (\beta_l - \beta_{l,0}) \left. \frac{\partial e_i}{\partial \beta_l} \right|_{\beta=\beta_0} \quad (10.5)$$

donde β_i es el vector de parámetros $\theta_1, \theta_2, \dots, \theta_q, \varphi_1, \varphi_2, \dots, \varphi_p$ y β_0 es un vector de valores iniciales. La sumatoria se introduce para que se tomen en cuenta todos los parámetros del vector β . Ahora para simplificar la notación hagamos

$$x_{i,t} = \left. \frac{-\partial e_i}{\partial \beta_l} \right|_{\beta=\beta_0}$$

dado que es una derivada evaluada en β_0 , finalmente se trata de un número. Reacomodando (10.5) queda

$$e_{i,0} + \sum_{l=1}^{p+q} \beta_{l,0} x_{i,t} = \sum_{l=1}^{p+q} \beta_l x_{i,t} + e_i$$

Dado que el lado derecho de esta ecuación es idéntica a la de la regresión lineal múltiple, gracias a que los parámetros β_l ya son lineales, es correcto suponer que todo el término del lado izquierdo se puede referenciar como una sola variable, digamos w_i ,

$$\begin{aligned} w_1 &= \beta_1 x_{1,1} + \beta_2 x_{2,1} + \dots + \beta_{p+q} x_{p+q,1} + e_1 \\ w_2 &= \beta_1 x_{1,2} + \beta_2 x_{2,2} + \dots + \beta_{p+q} x_{p+q,2} + e_2 \\ \dots & \dots \dots \dots \dots \dots \dots \\ w_T &= \beta_1 x_{1,T} + \beta_2 x_{2,T} + \dots + \beta_{p+q} x_{p+q,T} + e_T \end{aligned}$$

Se puede apreciar que la forma es idéntica a la de la regresión, por lo que se puede resolver como se hizo con los AR puros, obteniendo así un nuevo vector $\beta_{i,1}$. Éste de vuelve a introducir al polinomio de Taylor para ser linealizado y se repite el proceso, obteniendo así un nuevo vector de parámetros $\beta_{i,2}$. El proceso itera de esta manera hasta que se cumpla una de las siguientes 3 condiciones

- 1) $|\beta_{ki} - \beta_{(k+1)i}| < \epsilon$. La diferencia entre el vector de parámetros estimado para una iteración y la siguiente deja de variar significativamente.
- 2) $|\text{SSE}_{ki} - \text{SSE}_{(k+1)i}| < \epsilon$. La diferencia entre la suma de errores al cuadrado para el vector de parámetros estimado para una iteración y la siguiente deja de variar significativamente.
- 3) $k = \text{número límite}$. Se llega a un número límite de iteraciones permitidas, que puede ser establecido por el usuario.

INICIALIZACIÓN DE PARÁMETROS

La inicialización de los parámetros es importante hasta cierto punto. Lo peor que puede pasar es que no converjan los estimadores, es decir que y se alejen cada vez más con cada iteración, lo cual es bastante grave, porque nos quedamos sin estimados. Pero según varios autores asignar el valor de cero a todos los parámetros trae resultados satisfactorios.

y específicamente Gaynor y Kirkpatrick ("Introduction to Time Series Modeling and Forecasting in Business and Economics", 1996) alegan que 0.1 es el valor con el que se debería inicializar. Pero no todos los paquetes estadísticos cuentan con esta opción.

X.3 Diagnóstico

Antes de pronosticar con el modelo que hemos formado, es necesario hacer una pruebas de diagnóstico para validar la bondad de ajuste. El modelo estimado puede no ajustarse a los datos por dos razones: porque no fue correctamente seleccionado y no puede proporcionar un buen ajuste a los datos o porque fue pobremente estimado, aunque sea *potencialmente capaz* de un buen ajuste.

Un buen modelo debe satisfacer las siguientes condiciones:

- 1) El proceso iterativo debe converger. Esto quiere decir que el proceso de estimación se detiene cuando no se puede encontrar ningún otro estimador del parámetro que genere una suma de errores al cuadrado menor (con un cambio relativo menor de 0.001).
- 2) Se deben satisfacer las condiciones de invertibilidad y/o estacionaridad
- 3) Los residuales (errores de pronóstico) deberían ser aleatorios y aproximadamente distribuidos de manera normal.
- 4) Todos los estimadores de parámetros deberían ser significativamente diferentes de cero (valores de t significativa).
- 5) El modelo debe ser parsimonioso (esto es, debe estar en su forma más simple)
- 6) El modelo debe tener una RMSE (raíz del error cuadrático promedio) mínima.

Veamos a continuación cada punto*.

1) Convergencia del modelo

El modelo debe converger a una mínimo de suma de errores al cuadrado. La suma de los residuales debería ser bastante pequeña. El proceso de estimación también deberá de converger exitosamente a los estimados. Si los estimados de los parámetros están muy cerca de los límites de la estacionaridad, podría ser que el proceso no ha convergido. Si el proceso no convergió, pueden haber diferentes causas para ello: los parámetros pueden estar tan intercorrelacionados que la colinearidad entre ellos alisa la superficie de respuesta de la suma de los errores al cuadrado. Se puede ya sea incrementar el número permitido de iteraciones o dar más holgura al criterio de convergencia. Cualquiera de las dos acciones facilitaría la iteración hacia una solución.

2) Estacionalidad e invertibilidad

Una revisión de los parámetros puede revelar la adecuación o inadecuación del modelo. Su cantidad, significación, magnitud, intercorrelación, proximidad a los límites de estacionaridad o invertibilidad, así como el algoritmo que se utilizó para la estimación, tienen implicaciones en su inclusión o eliminación en el modelo.

* En adelante hablaremos de estacionaridad en lugar de convergencia, ya que este concepto se estará utilizando en el contexto del método de obtención de estimadores y no ya en el de PLG.

Las magnitudes de los estimados de los parámetros deben ser razonables. Los estimados de los parámetros deben estar considerablemente dentro de los límites de estacionaridad e invertibilidad. Si resulta necesario, se deberían formular los polinomios de diferencias y se deberían probar las raíces.

Si los parámetros están cerca de los límites de estacionaridad e invertibilidad es recomendable hacer las pruebas de raíz unitaria. A continuación se explica por qué.

RAÍCES UNITARIAS: CARACTERÍSTICAS DE SU ESTIMACIÓN POR MÍNIMOS CUADRADOS

Utilizaremos como ejemplo una caminata aleatoria, pero los resultados se aplican a todos los ARIMA(p, l, q) en general.

$$y_t = y_{t-1} + e_t$$

Pero no sabemos que el coeficiente autorregresivo es 1, así que estimamos el modelo

$$y_t = \phi y_{t-1} + e_t$$

Existen dos propiedades contrarias del estimador de mínimos cuadrados: la superconsistencia y el sesgo.

Empecemos por la superconsistencia. En el caso de la raíz unitaria, $\phi=1$, la diferencia entre $\hat{\phi}$ y 1 disminuye con cierta rapidez a medida que aumenta el tamaño muestral; de hecho, se contrae con una velocidad de $1/N$, donde N es el tamaño de muestra. En el caso estacionario, $\phi < 1$, la diferencia entre $\hat{\phi}$ y ϕ disminuye con más lentitud, como $1/\sqrt{N}$. A la convergencia extra rápida del caso unitario se le llama superconsistencia.

En el caso del sesgo se puede demostrar (Diebold, 1998) que el estimador de mínimos cuadrados, tiene sesgo hacia abajo, de modo que si el valor real de ϕ es ϕ^* , el valor esperado de $\hat{\phi}$ es menor que ϕ^* . En igualdad de circunstancias, cuanto mayor sea el valor real de ϕ , mayor será el sesgo, así que el peor caso es el de raíz unitaria. El sesgo disminuye a medida que aumenta el tamaño de muestra, porque el estimado converge al valor real de la población, pero en muestras del tamaño que nos concierne, el sesgo puede ser apreciable.

Conclusión: El estimador se acerca rápidamente a un valor por debajo de la unidad, cuando el valor real del parámetro es 1. Pero si el valor verdadero del parámetro no es la unidad, el estimador se acerca sin sesgo alguno, pero muy lentamente.

Así que si vemos como resultado un estimador de .95, por ejemplo, que al aumentar el tamaño de la serie también aumenta y se acerca a 1, es muy probable que se trate de una raíz unitaria. Pero en otro caso, un valor real muy por debajo de 1, no hay nada que hacer, porque el estimador se está acercando al valor verdadero, pero muy lentamente.

Esto no es para que desconfiemos de los resultados que nos da la computadora, por el contrario, son un arma teórica más que nos ayuda a interpretarlos mejor.

3) Análisis de residuales

Si las ACF y PACF de los residuales son diferentes a las del ruido blanco, se debe reespecificar el modelo.

Con el fin de tener una aproximación más cercana a la de χ^2 para muestras pequeñas se utiliza el estadístico de Ljung-Box:

$$Q = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{(n-k)}$$

Se puede apreciar que para valores de T moderados y grandes (>25), estos dos estadísticos difieren poco.

La selección de m permite equilibrar los pesos de competencia. Por un lado no se desea que m sea demasiado pequeña, porque se está tratando de hacer una prueba conjunta sobre una gran parte de la función de autocorrelación. Por otro lado, a medida que m crece en relación con T, la calidad de las aproximaciones se deteriora. En la práctica, lo razonable es asignar un valor de \sqrt{T} a m.

Para leer correctamente los residuales se utiliza asimismo el nivel de normalidad de los residuales, con las medidas de curtosis y asimetría**.

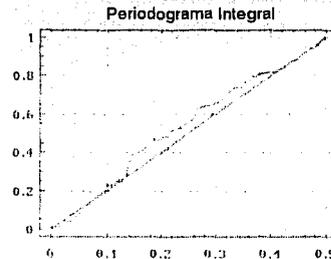
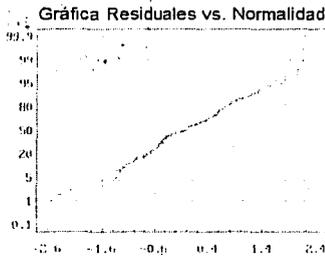
Otras dos herramientas que son de mucha ayuda son la gráfica de los residuales contra la normalidad y el periodograma integral de residuales. El primero nos ayuda a identificar la falta o presencia de normalidad en los residuales y el segundo es de especial utilidad cuando se están analizando series estacionales. El periodograma integral se define como:

TESIS CON FALLA DE ORIGEN

$$C(f_i) = \frac{\sum_{i=1}^i R^2(f_i)}{ns^2}$$

donde $R^2(f_i) = \frac{2}{n} \left[\left(\sum_{l=1}^n e_l \cos(2\pi e_l) \right)^2 + \left(\sum_{l=1}^n e_l \sin(2\pi e_l) \right)^2 \right]$; $f_i = \frac{i}{n}$ = frecuencia; $\frac{1}{f_i}$ = periodo; $s^2 = \hat{\sigma}^2$

y n = tamaño de la muestra. Para residuales de ruido blanco se deberían ver de la siguiente manera:

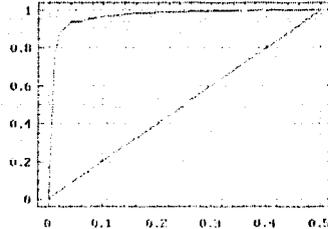
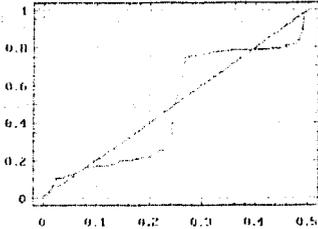


Las bandas de tolerancia se llaman límites de confianza Kolmogorov-Smirnov que están dados, para un 95% de confianza, a $\frac{\pm 1.36}{\sqrt{m}}$ con $m = \frac{n-2}{2}$ para n par y $m = \frac{n-1}{2}$ para n impar.

* Para todo lo referente a estacionalidad, ver capítulo VI: Modelos con variación estacional.

** Referirse al Anexo 4 para una explicación de cómo trabajan y cómo se usan la curtosis y la asimetría.

Unos residuales con patrón estacional de tamaño 4 presentarán valores significativos cada $R^2(1/4)$, $R^2(1/8)$, $R^2(1/12)$, etc. Que está representado por el primero de izquierda a derecha.



En este caso se debería eliminar con diferencias estacionales o modelar la estacionalidad con modelos estacionales. Cómo hacer esto se verá en el tema VII. También es posible fallar en captar la tendencia, en cuyo caso veríamos un periodograma integral de residuales como la segunda gráfica de periodogramas del esquema anterior.

4) Estimadores significativos

Una prueba rápida es verificar que, si se trata de un MA(2), la suma de θ_1 y θ_2 debe dar menos de 1. θ_2 menos θ_1 debe ser menor que 1. También el valor absoluto de θ_2 debe ser menor que uno.

No sólo deben ser de magnitud razonable los estimados de los parámetros (es decir, >0.06 y < 0.96 aproximadamente), deben ser también estadísticamente significativos.

Para saber si un parámetro permanece o es eliminado de un modelo, utilizamos la prueba t , que es la prueba de la significación de un parámetro en el modelo (de manera similar a como se hace en la regresión). Así podemos encontrar si estamos sobreespecificando el modelo. El estadístico t proporciona una prueba de la hipótesis de irrelevancia de la variable: que el parámetro poblacional verdadero, pero desconocido, es cero y en consecuencia que la variable correspondiente no contribuye con al pronóstico de la regresión y, por tanto, se puede omitir:

$$t = \frac{\text{estimado puntual del parámetro}}{\text{error estándar estimado}}$$

Un modelo correcto requiere que se conserven solamente aquellos términos cuyos valores t sean significativamente mayores a un valor crítico predeterminado. Utilizamos $|t| > 1.96$ para una confiabilidad del 95%. Los términos cuyos valores t no sean significativos, deben eliminarse y los parámetros restantes en el modelo deben re-estimarse.

Para evitar consultar las tablas, Statgraphics nos proporciona la probabilidad P de obtener un valor absoluto del estadístico t cuando menos tan grande, en valor absoluto, al que se obtuvo suponiendo que es verdadera la hipótesis de irrelevancia. Así si t fuera igual a 2, el valor correspondiente de probabilidad sería aproximadamente igual a 0.05. Cuanto menor es el valor de probabilidad, más fuerte es la evidencia contraria a la irrelevancia.

| Summary of Fitted Model for: MACUATRO.var1 | | | | |
|--|----------|-----------|----------|---------|
| Parameter | Estimate | Std.error | T-value | P-value |
| SAR(4) | .79624 | .05012 | 15.88640 | .00000 |

TESIS CON
 FALLA DE ORIGEN

Algunas veces los parámetros pueden estar cerca de ser significativos y deberán permanecer en el modelo, por razones de *inexactitud muestral*.

Si el modelo está correctamente especificado y los parámetros del modelo están abarcando toda la varianza sistemática, entonces los residuales deberían parecerse al ruido blanco.

5) Parsimonia

Los modelos estables y parsimoniosos tienen siempre preferencia.

El modelo debería ser razonable, parsimonioso y lo más elegante posible. Si se hicieron diferencias estacionales, debe existir una explicación lógica para que la serie esté comportándose estacionalmente. Asimismo, si se identificó una tendencia, deberíamos poder explicarla.

El mejor modelo ARIMA es aquel con menor cantidad de parámetros. Una de las razones para esto es la que de esta manera se reduce la incertidumbre debida al desconocimiento de los valores "verdaderos" del proceso.

Otra razón fuerte para preferir aquellos modelos que tienen un orden pequeño, es para evitar lo que se llama redundancia en los parámetros. Aunque los parámetros de un modelo de Box-Jenkins siempre estarán correlacionados, correlaciones muy altas, mayores a .8 o .9, por ejemplo, indican que el modelo tiene parámetros de más.

Aunque a veces se obtienen modelos con diferente cantidad de parámetros que presentan un ajuste igualmente bueno, los parámetros de los modelos de mayor orden siempre tienden a ser inestables y a cambiar drásticamente para cada muestra diferente del mismo proceso (fragilidad).

Así también el modelo debería plasmar la mayor varianza sistemática posible, dejando así residuales de ruido blanco.

6) Error Cuadrático Mínimo

En teoría la estimación de parámetros se debe intentar con diferentes algoritmos para ver si proporcionan resultados idénticos. Si el modelo proporciona similares resultados con esta técnica multimétodo, se dice que el modelo presenta confiabilidad, estabilidad y una relativa robustez a variaciones en las estimaciones. Si los resultados de los diferentes métodos difieren substancialmente, es un modelo al que algunos autores denominan como un *modelo frágil*.

El modelo necesita verificarse para sobre y sub-diferenciación. Si el modelo es óptimo, entonces ni añadiendo más parámetros, ni eliminando los parámetros dudosos, se debería formar una suma de errores al cuadrado menor.

Cuando se esté realizando el procedimiento de añadir más parámetros se debe tener cuidado de no *crear* una redundancia de parámetros. Ésta sucede cuando se añade un término AR y uno MA al mismo tiempo. Por ejemplo, tómese un modelo MA(1)

$$y_t = \theta_1 e_{t-1} + e_t \quad (10.6)$$

Añádasele un término AR y un MA, con lo que queda un modelo ARMA(1,2)

$$y_t = \theta_1 e_{t-1} + \theta_2 e_{t-1} + \phi y_{t-1} + e_t$$

Si le aplicamos una diferencia al modelo MA(1)

$$y_t - y_{t-1} = (\theta_1 e_{t-1} + e_t) - (\theta_1 e_{t-2} + e_{t-1})$$

despejemos y_t , queda:

$$y_t = [\theta_1 - 1] e_{t-1} - \theta_1 e_{t-2} + y_{t-1} + e_t$$

si sustituimos $\theta_2 = [\theta_1 - 1]$ y $\varphi = 1$, el modelo queda idéntico al ARMA(1,2)

$$y_t = \theta_2 e_{t-1} - \theta_1 e_{t-2} + \varphi y_{t-1} + e_t \quad (10.7)$$

Este último modelo (10.7) equivale al (10.6), pero además de ser más complejo, tiene los coeficientes correlacionados.

Es común encontrarse con uno o más modelos buenos. En estos casos se selecciona el que tenga la menor cantidad de parámetros y un RMSE mínimo

$$RMSE = \sqrt{\frac{\sum (Y_t - \text{pronóstico}_t)^2}{n - n_p}}$$

n = número de observaciones en la serie estacionaria

n_p = número de estimados de parámetros en el modelo

7) Comparación de pronósticos fuera de la muestra contra el modelo

Se puede agregar un punto más, importante por cierto, que va muy ligado al siguiente paso de la metodología, el del pronóstico: puede existir el sobreajuste del modelo, un ajuste a las discrepancias o al ruido, que hace que el modelo sea inservible. Para evitar esto se separa una pequeña porción de los datos, para compararlos con el pronóstico generado por la otra porción. Tanto los pronósticos como la porción más pequeña separada deben caer dentro del intervalo de confianza del pronóstico.

CRITERIOS ADICIONALES

Recordemos que cuantos más estimadores agreguemos al modelo, será posible un mayor ajuste, así que no es justo comparar 2 modelos que tienen diferente cantidad de parámetros, simplemente tomando en cuenta su suma de errores al cuadrado. En estos casos se pueden utilizar otros criterios, como su velocidad de convergencia: cuántas iteraciones hicieron falta para obtener los estimadores.

```

Estimation begins....
Initial:   RSS = 360.6   b = 0.1
Iteration 1: RSS = 179.66  b = 0.564364
Iteration 2: RSS = 157.947  b = 0.750109
Iteration 3: RSS = 157.053  b = 0.791386
Final:    RSS = 157.042  ...stopped on criterion 2
    
```

Existen también otras medidas de bondad de ajuste como la R cuadrada ajustada, el criterio de Schwarz y el criterio de Akaike*.

No se puede tomar a ningún indicador estadístico por sí sólo como evidencia suficiente para realizar un cambio en el modelo. Se tiene que obtener una visión global

* Referirse al Anexo 4 para una explicación de cómo trabajan y cómo se usan estas medidas de bondad de ajuste.

TESIS CON
 FALLA DE ORIGEN

conjuntando la información que nos proporcionan los diversos estimadores. Una vez que se decide hacer un cambio, el nuevo modelo se tiene que volver a someter al mismo diagnóstico.

X.4 Pronóstico

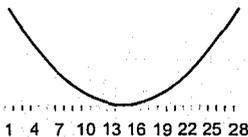
Ya se vio de manera amplia en el capítulo anterior el tema de cómo poner a funcionar los modelos de pronósticos. Al hacer el pronóstico se deben guardar en mente ciertos criterios que le agregan riesgo al pronóstico, como son el de la función de pérdida.

Función de pérdida

Y antes de proceder, tocaremos un punto que se ha dejado intacto, el de función de pérdida. Esta función puede ser cuadrática o lineal, simétrica o asimétrica. La función de pérdida nos dice cuánto estamos pagando por nuestros errores. Por ejemplo, si se nos pide que hagamos un pronóstico del clima para el día de mañana en grados centígrados, con el fin de preparar desde hoy el menú para la comida de mañana de una familia, podría resultar ser un pronóstico con función de pérdida bastante tolerante. Lo peor que puede pasar es que cocinemos algo muy caliente para un día caliente, o algo muy fresco que no se antoja tanto en un día frío. Si se considera que se trata de una familia que normalmente come con apetito, podríamos fallar por 10 o 15 grados y la función de pérdida sería bastante plana.

Pero si se trata del mismo pronóstico pero para un plantío de naranjas, por ejemplo, el equivocarnos por 10 grados podría resultar en la pérdida de toda la cosecha. En este caso la función de pérdida podría ser una cuadrática, que se comporta tolerante con pequeños errores pero para pérdidas mayores las cuantifica como mucho mayores.

TESIS CON
FALLA DE ORIGEN



Si se define a la función de pérdida por L , y a e como la diferencia entre el valor pronosticado y el valor real, una función de pérdida cuadrática podría ser

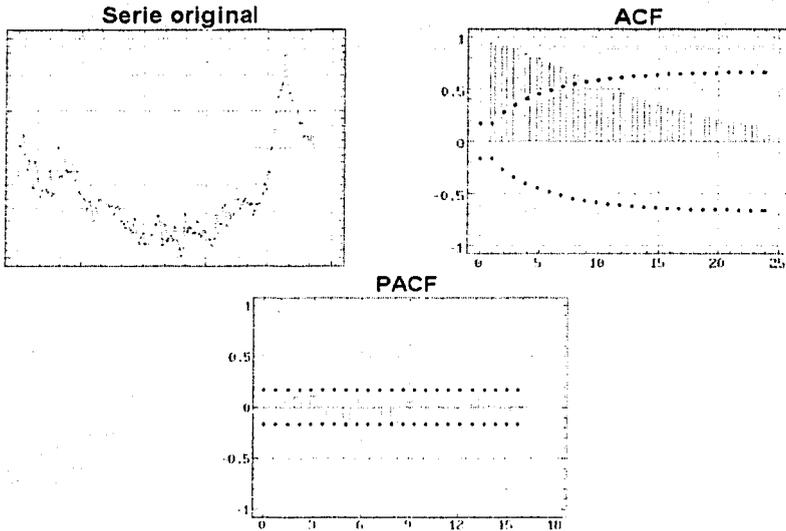
$$L(e) = e^2$$

Ésta es la que hemos estado utilizando, que es a su vez simétrica, lo que indica que el precio que se debe pagar por sobreestimar es el mismo que se paga por subestimar el pronóstico.

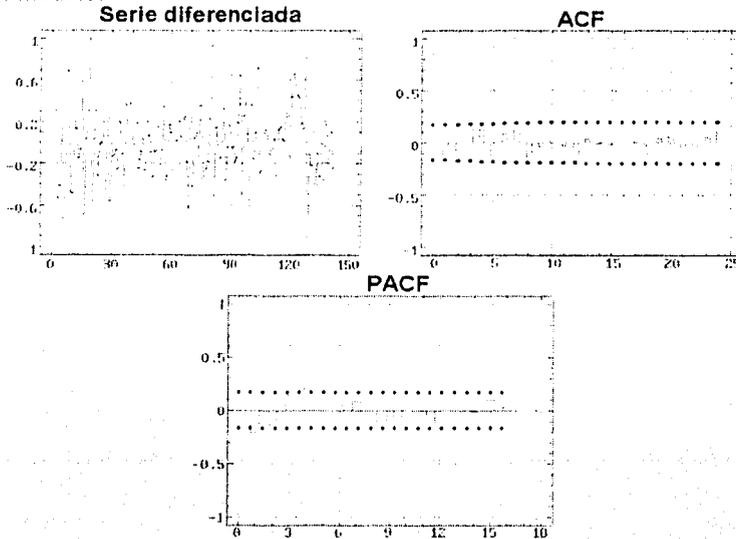
EJEMPLO PRÁCTICO

La siguiente serie de tiempo representa la tasa de desempleo abierto mensual en México de enero de 1985 a octubre de 1996*. Primero se observa la serie original junto con sus ACF y PACF.

* Fuente: INEGI



Se puede apreciar que no es estacionaria, ya que cuadra con las características de las serie no estacionarias. De hecho podemos atrevernos a decir que se trata de una caminata aleatoria, ya que la PACF es casi 1 y la ACF disminuye con demasiada lentitud. Se procede entonces a aplicar una diferencia ordinaria. La serie pasa a tener la apariencia de una serie estacionaria y la PACF y ACF se comportan de una manera que ya dan indicios de cuál podría ser el modelo.



**TESIS CON
FALLA DE ORIGEN**

El ACF y PACF ambos sobresalen en $k=1$. Para $k=2$ el PACF casi sobresale pero en el ACF claramente se trunca después de uno, así que se puede estimar el modelo para

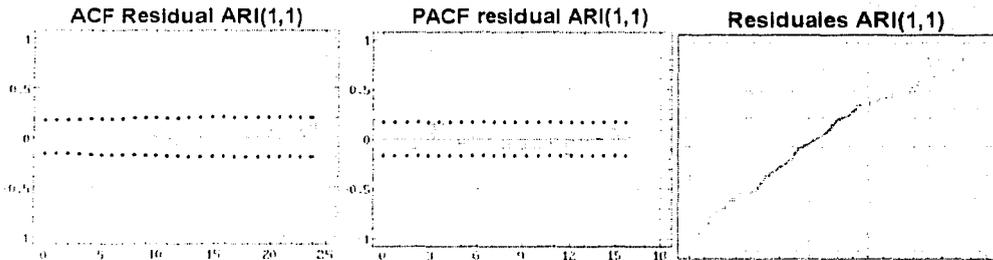
ARI(1,1), IMA(1,1) o incluso ARIMA(1,1,1). Recuérdese que ya se diferenci6 una vez por lo que en adelante los modelos que se forman llevan por fuerza $I=1$.

| Summary of Fitted Model for: DESEMP1.var1 | | | | |
|---|----------|-----------|----------|---------|
| Parameter | Estimate | Std.error | T-value | P-value |
| AR (1) | -.20531 | .08287 | -2.47764 | .01442 |
| Model fitted to differences of order 1 | | | | |
| Estimated white noise variance = 0.108903 with 140 degrees of freedom. | | | | |
| Estimated white noise standard deviation (std err) = 0.330004 | | | | |
| Chi-square test statistic on first 20 residual autocorrelations = 27.2618 | | | | |
| with probability of a larger value given white noise = 0.0987005 | | | | |
| Backforecasting: no | | | | |
| Number of iterations performed: 1 | | | | |

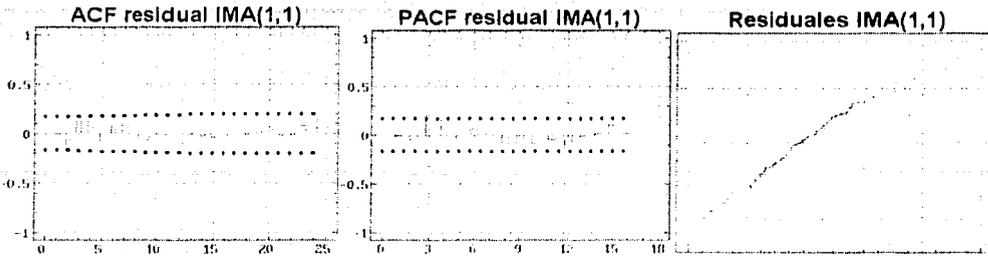
En los valores encerrados en elipses se aprecia que el valor T es significativo (>2), y el de P y el de la prueba de chi-cuadrada respaldan esto. Ahora intentemos un modelo IMA(1,1).

| Summary of Fitted Model for: DFSFMP1.var1 | | | | |
|---|----------|-----------|---------|---------|
| Parameter | Estimate | Std.error | T-value | P-value |
| MA (1) | .21614 | .08297 | 2.60493 | .01018 |
| Model fitted to differences of order 1 | | | | |
| Estimated white noise variance = 0.108281 with 140 degrees of freedom. | | | | |
| Estimated white noise standard deviation (std err) = 0.329061 | | | | |
| Chi-square test statistic on first 20 residual autocorrelations = 26.1913 | | | | |
| with probability of a larger value given white noise = 0.126573 | | | | |
| Backforecasting: no | | | | |
| Number of iterations performed: 1 | | | | |

Se puede ver que el modelo IMA(1,1), aunque tiene una ligera mayor representaci6n del proceso, no est6 pronosticando tan acertadamente como el ARI(1,1) ya que la prueba de bondad de ajuste as6 lo indica. El valor P est6 indicando que existe una probabilidad de 0.01 de que el valor verdadero del par6metro IMA(1,1) sea igual a cero, en lugar del estimado que est6 indicando: 0.216. Y el valor de chi-cuadrada est6 indicando que los residuales no son tan parecidos al ruido blanco como con el modelo ARI(1,1).

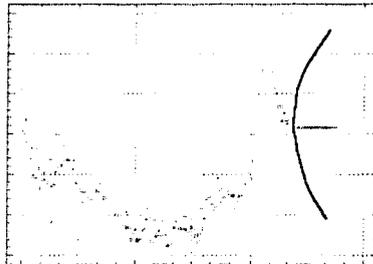


En los residuales del modelo ARI(1,1) se encuentra un comportamiento muy parecido al del ruido blanco. Todas las ACF y PACF est6n dentro de las barras de Bartlett. Adem6s siguen de cerca a la normalidad. En cambio el modelo IMA(1,1) tiene residuales que indican que el modelo se le escapa ciertos comportamientos que podr6an mejorarlo, como se muestra en las siguientes ilustraciones.

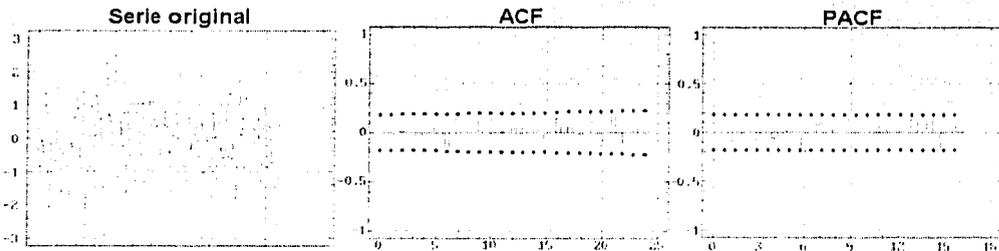


Una vez que nos hemos decidido por un modelo. ARI(1,1) con -0.2, realizamos el pronóstico.

Pronósticos



El mismo ejercicio que acabamos de hacer lo podemos realizar con datos simulados, de manera que pretendemos no saber cuál es el modelo. aunque lo generaremos nosotros mismos. Se modelará un modelo ARI(1,1) con -0.2 para ver las diferencias entre procesos reales y procesos simulados.



Se puede apreciar que el PACF tiene un valor aproximado de -0.2 en $k=1$ por lo que nos haría pensar que se trata de una ARI(1,1). Se asigna $l=1$ ya que se presupone que esta serie ya está diferenciada una vez. Se obtiene el estimador para ARI(1,1) y éste nos indica que hemos seleccionado el modelo correcto.

| Parameter | Estimate | Std.error | T-value | P-value |
|-----------|----------|-----------|----------|---------|
| AR (1) | -0.2003 | .06735 | -2.99379 | 0.0234% |

Estimated white noise variance = 1.25902 with 126 degrees of freedom.
 Estimated white noise standard deviation (std err) = 1.12206
 Chi-square test statistic on first 20 residual autocorrelations = 20.748
 with probability of a larger value given white noise = 0.350293
 Backforecasting: no
 Number of iterations performed: 1

TESIS CON
 FALLA DE ORIGEN

El valor del estimador es de -0.2 y el valor de que el valor del parámetro verdadero de $ARI(1,1)$ sea cero es muy bajo, apenas del 2 por ciento, por lo que hemos encontrado el modelo verdadero, que es precisamente el que se modeló desde un principio.

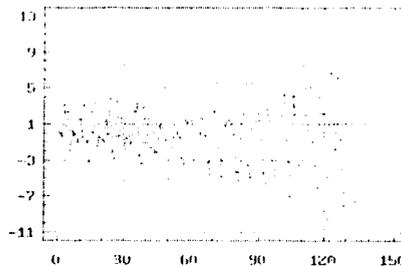
Capítulo XI. Modelos con variación estacional.

La diferencia entre los modelos ARMA estacionales y los no estacionales es que éstos son una representación del PLG como ya se vio, mientras que los primeros no lo son, ya que *se saltan* valores. Un AR(3) tiene y_{t-1} , y_{t-2} , y_{t-3} y un SAR(3) tiene solo y_{t-3} . Los modelos estacionales son necesarios en la vida real ya que muchos procesos se pueden representar exclusivamente en función de su pasado estacional o en función de una mezcla de su pasado estacional con su pasado reciente continuo.

Así un modelo estacional economiza parámetros, porque existen procesos que se pueden modelar en base al intervalo estacional y no lo que pase entre intervalos. Pensemos por ejemplo en el consumo interno que hay en el país en la época de navidad y lo poco que se gasta en la "cuesta de enero". Cada verano se gasta más en gasolina, trajes de baño, cremas protectoras y balnearios. Todos los meses del año se comportan de manera muy parecida a como se comportaron los años anteriores. Si el modelo se estableciera en función exclusivamente del consumo interno que hubo doce meses atrás se estaría hablando de un modelo **puramente estacional**. Por otro lado a veces hace falta una interrelación entre los efectos estacionales con efectos recientes, como podría ser el caso de la cantidad de vuelos para una aerolíneas, donde los vuelos del presente mes se puede explicar a través de las ventas de doce meses atrás junto con las ventas del mes anterior. Este tipo de modelo se llama **multiplicativo**, porque precisamente se multiplican los efectos.

Con los modelos estacionales se sigue exactamente la misma estrategia que con los no estacionales. Así que primero buscamos que éstos sean estacionarios.

Cómo transformar un modelo estacional no estacionario en uno estacionario



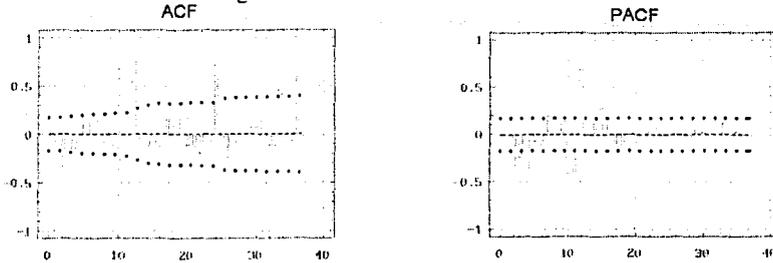
TESIS CON
 FALLA DE ORIGEN

Se procede primero a observar la serie. El análisis visual de la serie cobra mucha más importancia con series estacionales que con las demás, ya que el poder informativo de los indicadores como el ACF y PACF, disminuye debido a la complejidad en el funcionamiento de los modelos y la ampliación en la gama de posibles modelos en competencia.

Por el momento lo único que sabemos de esta serie es que se trata de datos mensuales. Podemos ver en una primera instancia que tiene varianza creciente, pero si se analiza un poco más se puede ver que esta aparente varianza tiene un patrón que se va repitiendo cada cierto tiempo. Los picos aparecen exactamente cada 12 datos, lo que es un

indicio de que se trata de una serie estacional, pero no es estacionaria ya que crecen los efectos con el tiempo.

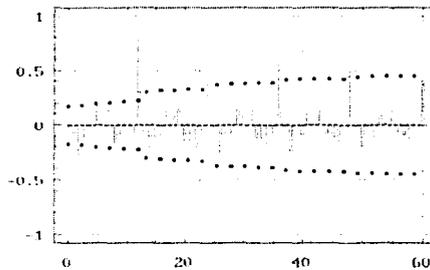
Se prosigue con los correlogramas. EL ACF decae poco a poco y el PACF se trunca. Los períodos estacionales según el ACF son de 12 lo que cuadra con el truncamiento del PACF en 12. Estos por sí solos nos indican que deberíamos intentar modelar con un $SAR(1)_{12}$. Si los títulos de estas dos gráficas se intercambiaran, se trataría de un $SMA(1)_{12}$.



Si se tiene duda con respecto al ACF en lo tocante a de si se trata de un corte brusco o es un decaimiento paulatino, podemos incrementar el tamaño de rezagos en la gráfica. Pongámoslos de 60 para que así se puedan apreciar cinco intervalos estacionales.

Haciendo esta expansión en el ACF es un recurso útil. Recordamos que son rarísimos los procesos ARMA con más de dos términos de rezago. Lo mismo va para los estacionales SARMA. Si se descartara la posibilidad de que fuera decaimiento en el ACF y que se supusiera truncamiento, se estaría afirmando que se trata de un $SMA(4)$ o incluso un $SMA(5)$, lo cual no cuadra con el principio de poca frecuencia de estos modelos ni con el principio de parsimonia. Se recomienda utilizar una cantidad grande de k rezagos cuando no haya completa claridad en la interpretación de correlogramas.

ACF



**ENCIC CON
FUELA DE ORIGEN**

Pero se puede notar, a pesar de la separación que hay entre ellos, que el decaimiento de los picos no es rápido: es más bien lento e indicante de que no hay estacionaridad, sobre todo dentro del contexto visual de la gráfica que nos indica exactamente lo mismo: la serie no es estacionaria. El criterio que se toma con los ACF y PACF no estacionales se toma también con los estacionales, pero se debe ser más cuidadoso con éstos y tener al criterio del análisis visual de la serie original como primordial.

Si estimamos aún así el modelo como un $SAR(1)_{12}$, el valor del parámetro efectivamente resulta ser de no convergencia. Es de casi 1.1 y al sustituirlo en la ecuación de diferencias, el inverso de la raíz resulta estar fuera del círculo unitario.

Estimation begins.....
 Initial: RSS = 1847.9 b = 0.1
 Iteration 1: RSS = 337.769 b = 0.76655
 Iteration 2: RSS = 156.554 b = 1.03317
 Iteration 3: RSS = 149.096 b = 1.09242
 Final: RSS = 149.004 ...stopped on criterion ?

Summary of Fitted Model for: SARNOE1.var1

| Parameter | Estimate | Std.error | t-value | P-value |
|-----------|----------|-----------|----------|---------|
| SAR(12) | 1.09939 | .02548 | 43.13994 | .00000 |

Estimated white noise variance = 1.10373 with 135 degrees of freedom.
 Estimated white noise standard deviation (std err) = 1.05059
 Chi-square test statistic on first 20 residual autocorrelations = 16.035
 with probability of a larger value given white noise = 0.654944
 Backforecasting: no Number of iterations performed: 4

El siguiente paso es entonces diferenciar estacionalmente. Tomamos diferencias de intervalo 12, ya que sabemos que se tratan de datos mensuales y es posible que haya un comportamiento similar cada 12 datos. Un intervalo de 10 sería contraintuitivo, por ejemplo, si se tratara de datos mensuales.

OPERADOR DIFERENCIA ESTACIONAL

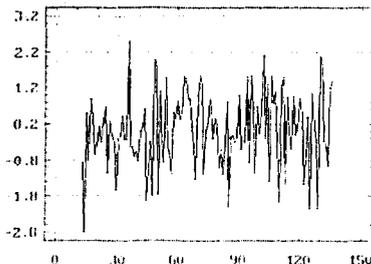
Hasta ahora sólo se ha utilizado el operador diferencia Δ^d . Para eliminar la estacionalidad de una serie con intervalo de estacionalidad de doce, no se puede hacer Δ^{12} , porque esto daría $\Delta^{11}(y_t - y_{t-1})$. El apropiado en este caso es el Δ_s^D , con $s=12$ y $D=1$, que hace exactamente lo que queremos:

$$\Delta_{12}(y_t) = y_t - y_{t-12}$$

Ésta es la diferencia estacional de primer orden con longitud de período $s=12$. Una de segundo orden con $s=4$ sería

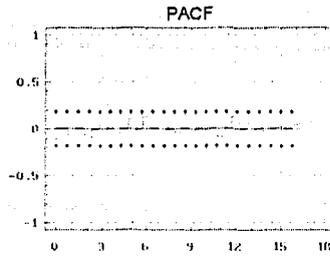
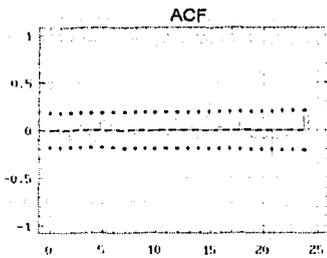
$$\begin{aligned} \Delta_4^2(y_t) &= \Delta_4(y_t - y_{t-4}) \\ &= (y_t - y_{t-4}) - (y_{t-4} - y_{t-8}) \\ &= y_t - 2y_{t-4} + y_{t-8} \end{aligned}$$

Una vez que se han aplicado las diferencias estacionales, se verificará la estacionalidad nuevamente, haciendo uso tanto de la gráfica como de los ACF y PACF.

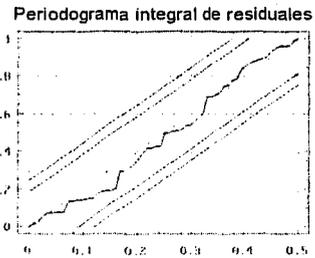
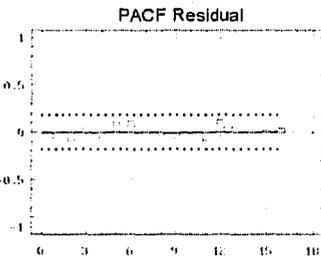
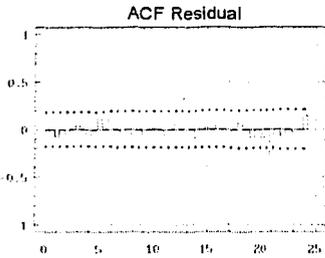


La serie ya parece estacionaria y ya no tiene picos estacionales. Las gráficas de autocorrelaciones lo confirman.

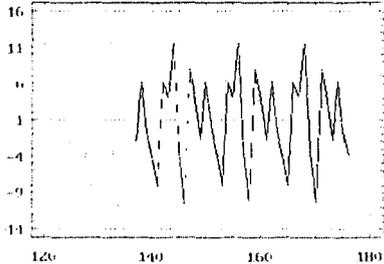
TESIS CON
 FALLA DE ORIGEN



De esta forma se llega a un modelo de predicción SARIMA(0,1,0)₁₂, ya que lo que resta después de diferenciar es ruido blanco únicamente. Hagamos el análisis de residuales:

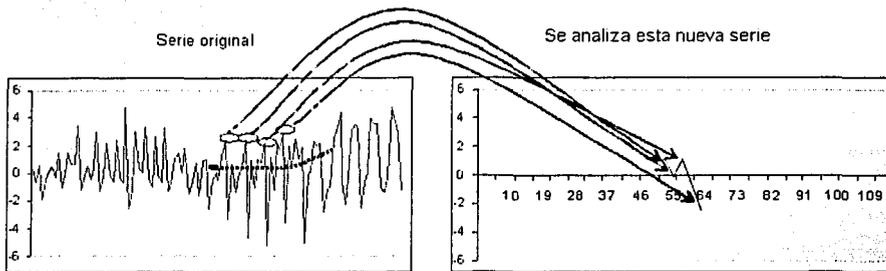


El ACF y el PACF nos indican que los residuales no tienen patrón alguno. El periodograma integral nos muestra que hicimos un buen trabajo en capturar la estacionalidad, ya que se mueve de manera errática y dentro de los límites. Por último echemos un vistazo al pronóstico:



Después del dato 138 empieza el pronóstico, que imita muy bien la estacionalidad de tamaño 12. El modelo que se simuló en este ejercicio fue un SAR(1)₁₂ con $\phi = 1.2$

XI.1 Modelos puramente estacionales

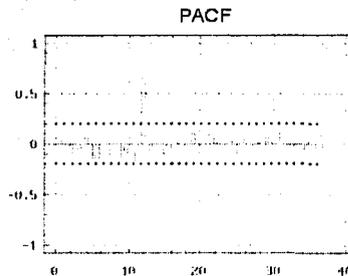
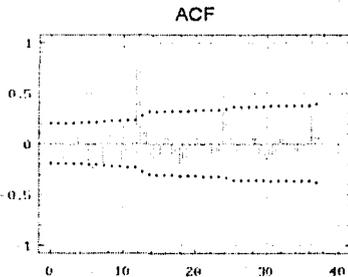


En esta ilustración es posible apreciar el enfoque que se toma para procesos estacionales. Se tratan de manera análoga que los modelos no estacionales: sólo que enfocándonos esta vez en la estacionalidad exclusivamente.

Al igual que con modelos no estacionales, los modelos ARMA y ARIMA, donde los segundos eliminaban la estacionaridad, para los modelos estacionales se tienen modelos que *modelan* la estacionalidad SARMA y otros que *eliminan* la estacionalidad para después modelar, los modelos SARIMA.

Así un modelo SARIMA(P,D,Q)_S estaría tomando D diferencias *estacionales*, y un modelo SARMA(P,Q) estaría modelando con P parámetros *estacionales* ϕ y Q parámetros *estacionales* θ , que para que no se confundan con los no estacionales pondremos en mayúsculas. Φ y Θ .

Tomemos por ejemplo un SAR(1)₁₂: $y_t = 0.9y_{t-12} + e_t$



Si suponemos que no existen los períodos no estacionales, la gráfica es similar a la de un proceso AR(1) no estacional. Es decir, sólo los múltiplos de S se muestran significativos y estos decrecen exponencialmente. La misma analogía se puede aplicar al PACF. El único k que se muestra significativo es el de 12, es decir en $k=S$ después de lo cual se trunca.

No obstante nótese que en 10 y en 8 aparecen unos picos que sobresalen en el PACF. En el ACF sucede en 6. Hay que tomar en consideración lo repetitivo de las series de tiempo, que sale a relucir más en estos casos donde hay estacionalidad que en los no estacionales.

TESIS CON FALLA DE ORIGEN

La serie de tiempo anterior se simuló con $S=12$, forzándola así a ser estacional. Lo que quiere decir que, hablando no muy rigurosamente, la serie se está "repitiendo" cada 12 datos. Y si se repite cada 12, es lógico que si la significación de las autocorrelaciones es alta, como es el caso (se utilizó $\Phi=.9$), las autocorrelaciones de k menores a 12 tiendan a ser significativas porque el estimador de la autocorrelación encuentra un comportamiento similar entre cada 2 datos, cada 3 datos, ... hasta 12, siendo este último el de mayor significancia. Este principio se aplica tanto a la ACF como la PACF. Es útil tener esto en mente cuando se analizan series reales para poder interpretar mejor los resultados.

MODELO SAR(1)

Un modelo SAR(1) es uno en donde el presente y_t puede expresarse a través de una ecuación lineal en función de y_{t-s} más un choque aleatorio e_t .

$$y_t = \Phi y_{t-s} + e_t$$

$$e_t = (1 - \Phi B^s) y_t$$

si sustituimos las y_t rezagadas, queda:

$$y_t = e_t + \Phi e_{t-s} + \Phi^2 e_{t-2s} + \Phi^3 e_{t-3s} + \dots$$

lo cual explica el comportamiento de la ACF:

$$\rho_{sk} = \Phi^k$$

Si Φ es mayor que uno en valor absoluto, el modelo no sería convergente y los golpes aleatorios pasados cobrarían mayor importancia que los más recientes. Por tanto el modelo SAR(1) tiene como condición de estacionaridad $|\Phi| < 1$.

MODELO SAR(P)

Un modelo SAR(P) aplica en series de tiempo donde la memoria estacional es mayor, ya que el presente y_t depende de lo que haya sucedido varios períodos estacionales "s" en el pasado.

$$y_t = \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + \Phi_3 y_{t-3s} + \dots + \Phi_P y_{t-Ps} + e_t$$

$$y_t = \sum_{i=1}^P \Phi_i y_{t-is} + e_t$$

El ACF para el SAR(P) se comporta de manera similar que el del SAR(1). El PACF de un SAR(P) sólo aparece significativo en los múltiplos de S hasta PS . Por ejemplo, un SAR(2)₄ tendrá valores significativos solamente en 4 y 8, después de lo cual se truncará.

Para ejemplo de cómo se ven las fotos de los procesos SAR puros tómese el ejercicio que ya se vio en esta sección al principio del actual capítulo.

MODELO SMA(1)

En un modelo SMA son los choques aleatorios, no los valores de la serie, los que *estacionalmente* dependen del pasado.

$$\Theta c_{t-s} + e_t = y_t$$

$$(\Theta B^s + 1) c_t = y_t$$

$$y_t = e_t + \Theta y_{t-s} - \Theta^2 y_{t-2s} + \dots$$

$|\Theta| < 1$ es la condición para la invertibilidad de los procesos SMA(1)

MODELO SMA(Q)

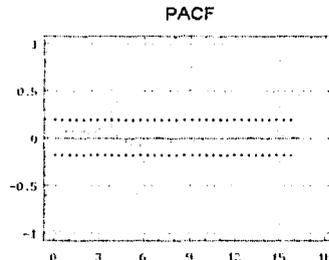
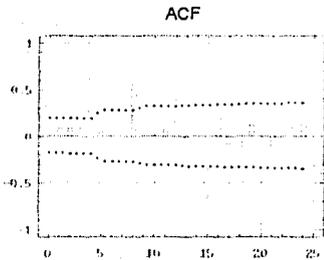
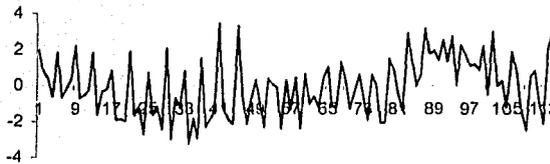
$$y_t = \sum_{i=1}^Q \Theta_i e_{t-is} + e_t = \sum_{i=1}^Q \Theta_i B^{is} e_t + e_t$$

Al igual que con los SAR(P), los SMA(Q) tienen ACF y PACF similares a los no estacionales. Las autocorrelaciones son sólo significativas en $k=S$, truncándose después del Q 'ésimo pico. Las autocorrelaciones parciales son decrecientes infinitas en los múltiplos de S .

Autocorrelación

$$\rho_k = \begin{cases} \frac{-\Theta}{1+\Theta^2} & \text{para } k=S \\ 0 & \text{para cualquier otro caso} \end{cases}$$

Tomemos por ejemplo el siguiente modelo SMA(1)₄ con $\Theta=0,9$



Se comprueba lo que se había dicho: la ACF es decreciente infinita y la PACF se trunca. Ya conocemos ahora las "fotos" de lo que estamos buscando.

Si el PACF de un SMA o el ACF de un SAR **no decaen** rápidamente significa que **hace falta diferenciar**.

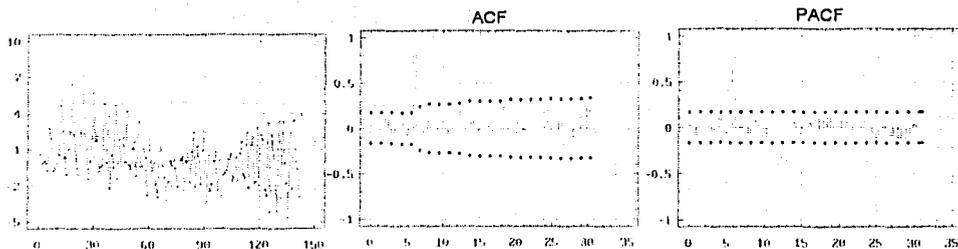
MODELOS MEZCLADOS SARMA(P,Q)

La única diferencia entre un modelo ARMA ordinario y un SARMA estacional es que en éstos las ACF y PACF aparecen únicamente en los intervalos estacionales S y al igual que aquellos tienen una doble restricción de invertibilidad y convergencia.

TESIS CON FALLA DE ORIGEN

TESIS CON FALLA DE ORIGEN

El siguiente es un SARMA(1,1)₆ con $\Phi=0.8$ y $\Theta=0.7$. En algunos valores de K del PACF donde se esperan coeficientes significativos, como en $k=18$, éstos apenas llegan a tocar la barrera de significación e incluso en 24 y 30 ni siquiera se notan.



Todo el conjunto de modelos estacionales que se han visto hasta ahora son **aditivos**, ahora procederemos a ver los **multiplicativos**.

XI.2 Modelos Multiplicativos

Al mezclar modelos estacionales con no estacionales se obtienen los modelos multiplicativos. De entrada se puede establecer que las condiciones de invertibilidad-estacionalidad-convergencia son una extensión lógica de las condiciones hasta ahora conocidas. Es decir, se tiene que resolver la ecuación de diferencias y la inversa de la raíz debe caer dentro del círculo unitario. Resolver estas ecuaciones a mano puede resultar extremadamente difícil porque ya nos encontraremos con ecuaciones de cuarto, sexto o doceavo grado.

La forma general de los modelos multiplicativos es:

$$\Phi(B^s)\varphi(B)y_t = \Theta(B^s)\theta(B)e_t + \delta$$

MODELOS AR(P) x SAR(P)_s

Tomemos un ejemplo sencillo para que se entienda su estructura

$$y_t = \phi_4 y_{t-4} + e_t \tag{11.1}$$

Se utiliza ϕ_4 pero el subíndice no quiere decir nada, es sólo un identificador. Supóngase ahora que e_t no es ruido blanco, sino que a su vez se trata de un proceso AR(1) en función de sí misma, de e_t

$$e_t = \phi_1 e_{t-1} + a_t \tag{11.2}$$

Donde a_t ya es ruido blanco normal. Ahora despejemos e_t de (11.1) y sustituyamos en ella esta última ecuación representada en función del operador B con a_t despejada.

De (11.2) $a_t = [1 - \phi_1 B]e_t$

De (11.1) $e_t = [1 - \phi_4 B^4]y_t$

Sustituyendo $[1 - \phi_1 B][1 - \phi_4 B^4]y_t = a_t$

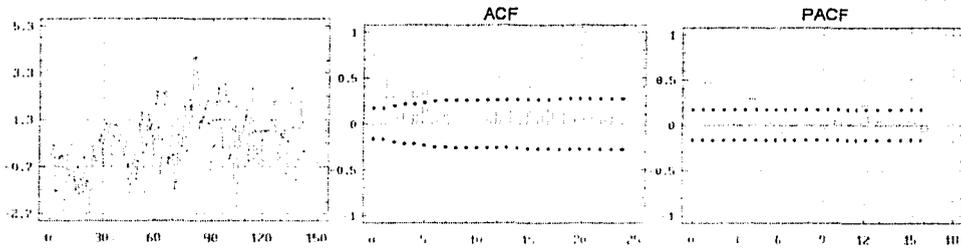
Que ya está en la forma general. Se trata de un AR(1) x SAR(1)₄. Si resolvemos la multiplicación que nos plantea queda:

$$y_t = \phi_1 y_{t-1} + \phi_4 y_{t-4} + \phi_1 \phi_4 y_{t-5} + a_t \tag{11.3}$$

Este modelo es una especie de AR(5) muy diferente a los que estábamos acostumbrados a ver hasta ahora, ya que se salta a y_{t-2} y y_{t-3} . Se hace hincapié en que se están usando dos parámetros para generar cuatro términos, que también es una novedad ya que antes un modelo con dos parámetros generaba su ecuación con tres términos (por el omnipresente e_t). Esta propiedad de los modelos multiplicativos lleva la capacidad parsimoniosa intrínseca, lo que le da gran potencial de representación con pocos parámetros.

La ventaja que nos da la versión desarrollada (11.3) del modelo, es que nos permite distinguir la estructura que tendrán las "fotos" ACF y PACF. El PACF tendrá que tener valores nulos en 2 y 3, independientemente de los valores de los parámetros. De cumplir estos con las condiciones de estacionalidad, la ACF será decreciente infinita.

El siguiente es el modelo (11.3) con valores de $\phi=.4$ y $\Phi=.3$



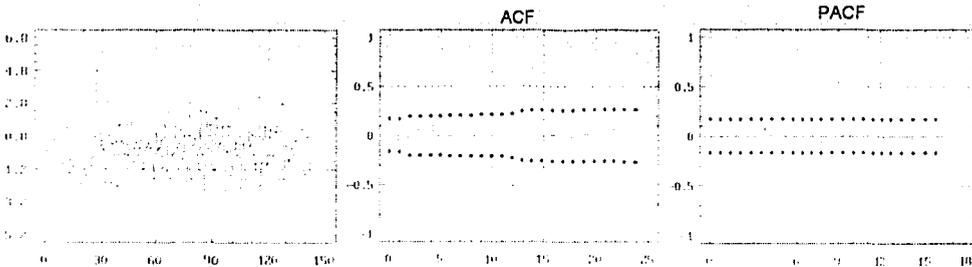
MODELOS MA(Q) x SMA(Q)_s

Análogamente un ARIMA(0,0,1)(0,0,1)₁₂:

$$y_t = (1 - \theta B)(1 - \Theta B^{12})e_t$$

$$y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}$$

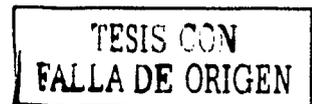
De aquí se pueden derivar los comportamientos de las autocovarianzas. Las siguientes son las gráficas de este modelo con $\theta = \Theta = 0.7$



MODELOS ARMA(P,D,Q) x SARMA(P,D,Q)

Por último desarrollemos el modelo de la aerolínea propuesto por Box y Jenkins (Box-Jenkins, 1976), un ARIMA(0,1,1)(0,1,1)₁₂, que modela la cantidad de vuelos. El modelo indica que para pronosticar la cantidad de vuelos es importante tanto el mes anterior como el mes del mismo nombre pero del año anterior.

$$(1 - \theta B)(1 - \Theta B^{12})e_t = (1 - B)(1 - B^{12})y_t$$



TESIS CON FALLA DE ORIGEN

El que el modelo cuente con una diferencia ordinaria sólo quiere decir que se debe incluir un término AR con parámetro igual a uno. Asimismo una diferencia estacionaria equivale a un modelo $(1,0,0)_s$ con parámetro $\Phi=1$. Desarrollando el modelo de la aerolínea:

$$y_t = e_t - \theta e_{t-12} - \theta e_{t-1} + \theta \Theta e_{t-13} + y_{t-12} + y_{t-1} - y_{t-13}$$

que da una ARMA de orden 13. Con 2 parámetros se llegó a un modelo con 6 términos.

PRONÓSTICO

Supongamos unas estimaciones de $\hat{\theta}=0.7$ y $\hat{\theta}=0.3$ y un pronóstico para $h=2$ etapas en el futuro. Para formular el pronóstico sólo basta con hacer funcional la última ecuación que vimos: todos los e_t desconocidos se sustituyen por ceros, los y_t desconocidos por sus pronósticos: los e_t conocidos se reemplazan por los errores de pronóstico a una etapa, es decir los residuales, que son la resta del pronóstico menos el valor observado $e_t = y_t - y_{t,T}$. Los parámetros se reemplazan por sus estimados.

$$\begin{aligned} Y_{t+2,T} &= e_{t+2} - 0.7e_{t-10} - 0.3e_{t+1} + 0.21e_{t-11} + y_{t-10} + y_{t+1,T} - y_{t-11} \\ &= -0.7[y_{t-10} - y_{t-10,T-11}] + 0.21[y_{t-11} - y_{t-11,T-12}] + y_{t-10} + y_{t+1,T} - y_{t-11} \\ &= 0.3y_{t-10} + 0.7y_{t-10,T-11} - 0.79y_{t-11} - 0.21y_{t-11,T-12} + y_{t+1,T} \end{aligned}$$

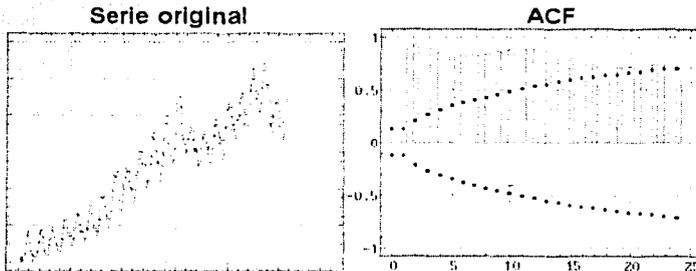
ACF'S Y PACF'S DE MODELOS ESTACIONALES

| Modelo | ACF | PACF |
|----------------|---------------------------|---------------------------|
| AR(p) x SAR(P) | Decae | Se trunca después de p+SP |
| MA(q) x SMA(Q) | Se trunca después de q+sQ | Decae |
| Mezclados | Decae | Decae |

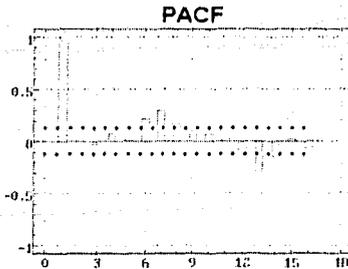
(Fuente: González Videgaray, 1990)

EJEMPLO PRÁCTICO

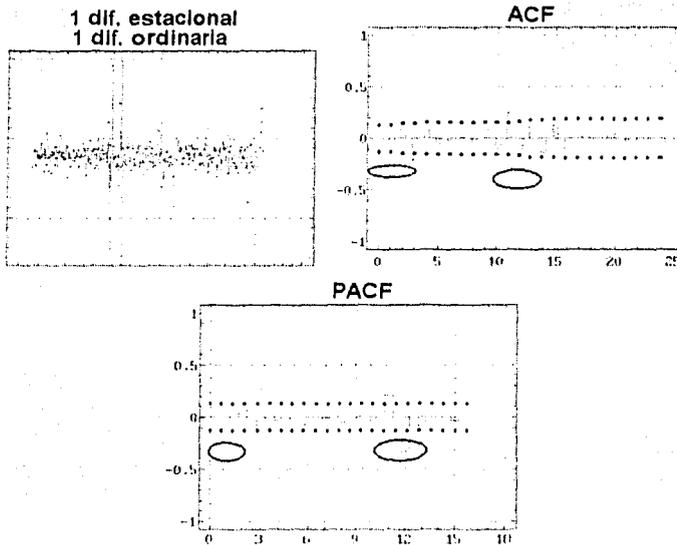
La siguiente serie de tiempo representa el consumo eléctrico doméstico mensual de enero de 1982 a junio del 2003*. Se trata de una serie que contiene tanto tendencia como estacionalidad de tamaño doce, como lo corroboran sus ACF y PACF.



* Fuente: INEGI



Al aplicar una diferencia estacional de tamaño doce, la serie se convierte en una estacionaria. La ACF y PACF nos indican que se puede tratar de un SARIMA₁₂(1,2,0)×ARIMA(1,0,0) ya que el ACF decae a partir de 1 y 12 y el PACF se trunca en 1 y 12 también.

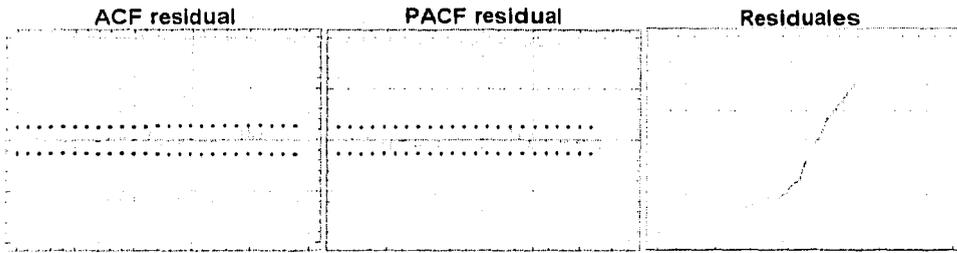


Al obtener los estimadores para el modelo señalado, los residuos se comportan como ruido blanco.

| Summary of Fitted Model for: ELECTRIC.var1 | | | | |
|--|----------|------------|----------|---------|
| Parameter | Estimate | Std. error | T-value | P-value |
| AR (1) | -.29573 | .06164 | -4.79785 | .00000 |
| SAR (12) | -.42811 | .06140 | -6.97222 | .00000 |

Model fitted to differences of order 1
 Model fitted to seasonal differences of order 1 with seasonal length = 12
 Estimated white noise variance = 5466.8 with 243 degrees of freedom.
 Estimated white noise standard deviation (std err) = 73.9513
 Chi-square test statistic on first 20 residual autocorrelations = 26.5646
 with probability of a larger value given white noise = 0.087531
 Backforecasting: no. Number of iterations performed: 4





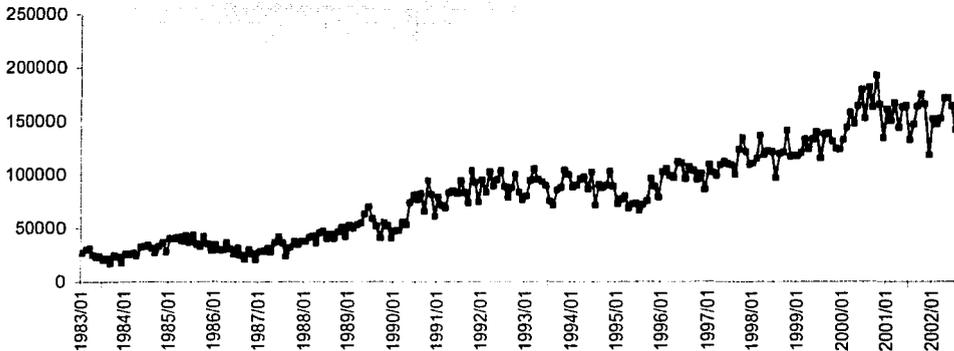
Con lo que se puede concluir que el mejor modelo es uno multiplicativo SARIMA₁₂(1,2,0) x ARIMA(1,0,0) con parámetros AR: -0.3 y SAR: -0.42. Que el modelo sea uno multiplicativo de esta forma no suena poco ilógico, ya que la utilización de energía eléctrica crece y disminuye cada 12 meses debido a dos factores principales: el alargamiento y acortamiento de la duración de luz solar a través del año. El otro factor se debe a los fríos de invierno, que son lo único que contrarresta al resto del año, en los cuales la temperatura es más o menos homogénea.

TESIS CON
FALLA DE ORIGEN

Capítulo XII. Aplicaciones

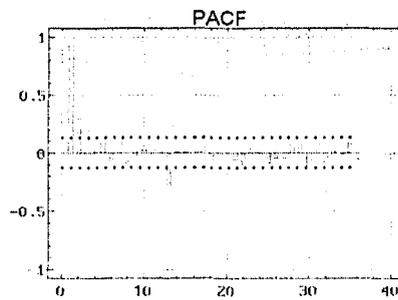
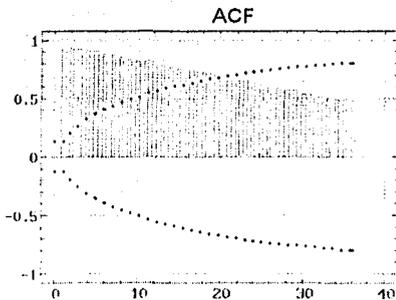
Caso I. Producción de vehículos automotores en México.

Se utilizará como primer aplicación la serie de datos de producción de automóviles. Los datos representan el total de la producción de unidades de automóviles por mes en México para el período que va de enero de 1983 a septiembre del 2002.



Fuente: INEGI. Asociación Mexicana de la Industria Automotriz, A.C.

Observamos una clara tendencia y no se encuentra una estacionalidad marcada, pero veamos también los ACF y PACF

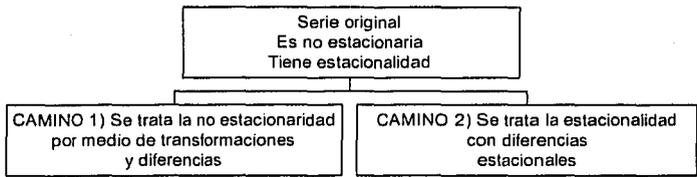


Podemos apreciar que aunque ya sabemos que no era estacionaria, no nos habíamos percatado de la estacionalidad de tamaño 12 indicada por el PACF. Esto se podría deber a que tenemos muchos datos en la serie y que las repeticiones de comportamiento cada cierto período no las podemos percibir tan claramente.

Tenemos ahora dos caminos a seguir: tratar de eliminar la estacionalidad de una vez y ver si de esta manera se vuelve también estacionaria la serie, o proseguir primero a volver estacionaria la serie y después tratar con la estacionalidad*.

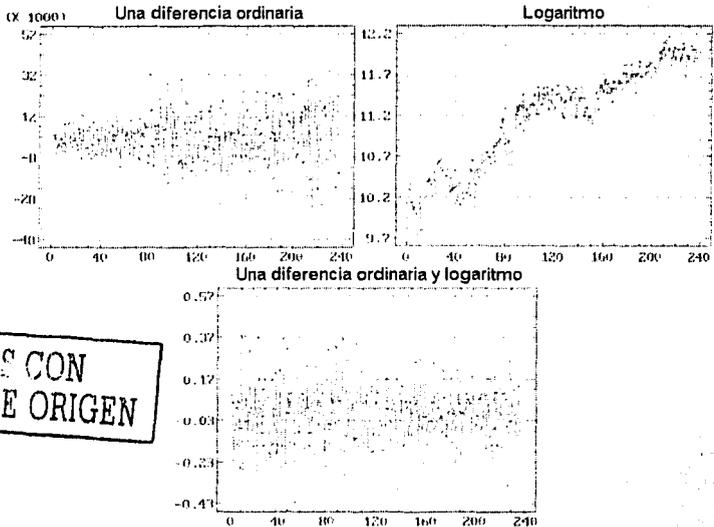
TESIS CON
FALLA DE ORIGEN

* A través de todo este ejemplo no se consideran todos los modelos posibles ya que para fines didácticos se consideran sólo los que son de valor para obtener el mejor modelo.



CAMINO 1

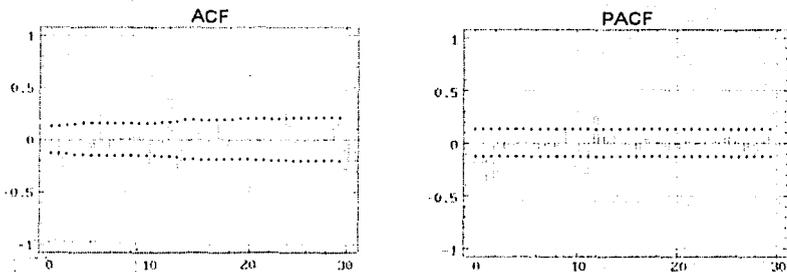
Empecemos entonces por tratar de eliminar la tendencia con diferencias ordinarias, con una transformación y con una combinación de estas dos.



TESIS CON FALLA DE ORIGEN

La diferencia ordinaria tiene problemas de varianza creciente. El logaritmo prácticamente la deja intacta, lo único que hace es reducir la escala. Pero conjugadas estas dos dejan a la serie visiblemente estacionaria, así que nos inclinamos por esta última.

Pero al analizar el ACF y PACF sale a relucir que la estacionalidad sigue ahí.



Éstas indican que es posible que se trate de un modelo multiplicativo. El ACF contiene un patrón que corresponde a un AR estacional de orden 12, con picos cada 12 y el PACF le corresponde con un solo valor en 12. Pero también estos picos están rodeados por

valores significativos. Si recordamos, al desarrollar las ecuaciones de los modelos multiplicativos, estos contaban con valores adyacentes al pico mayor estacional.

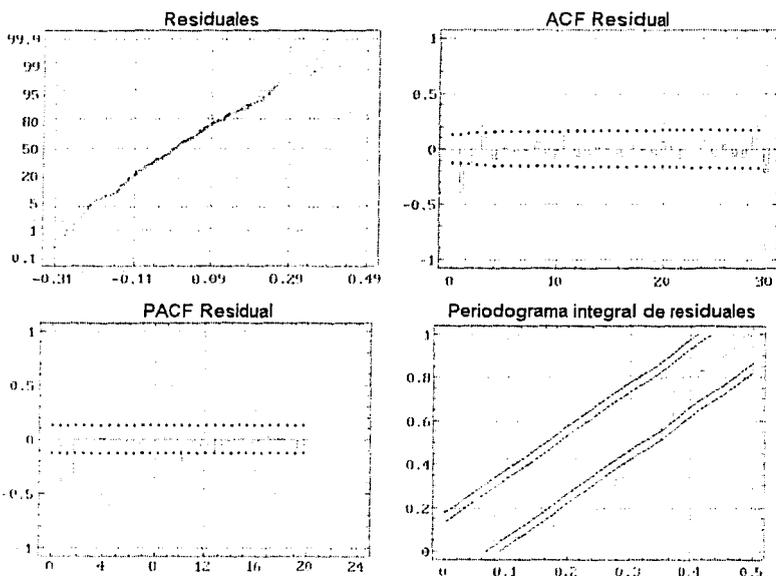
El primer modelo que probamos es un SAR(1)₁₂.

$$y_t = \Phi y_{t-12} + e_t$$

| Summary of Fitted Model for: log AUTOS2.var1 | | | | |
|--|----------|-----------|---------|---------|
| Parameter | Estimate | Std.error | T-value | P-value |
| SAR(12) | .44534 | .06016 | 7.40219 | .00000 |
| MEAN | .00607 | .01515 | .40093 | .68883 |
| CONSTANT | .00337 | | | |

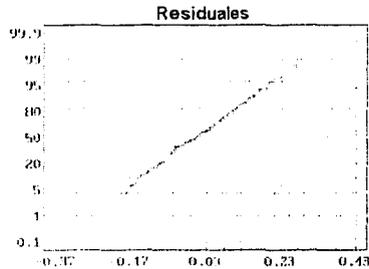
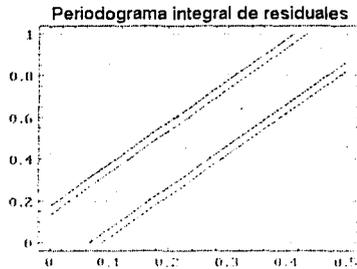
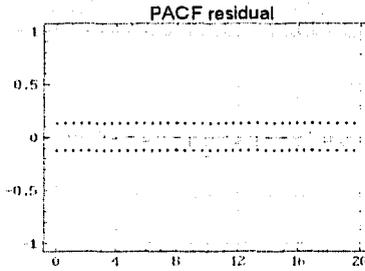
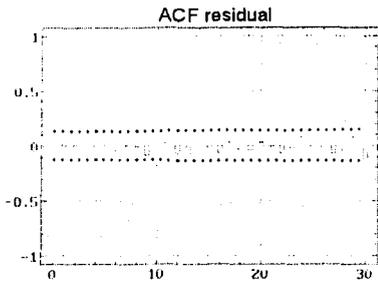
Model fitted to differences of order 1
 Estimated white noise variance = 0.0197619 with 234 degrees of freedom.
 Estimated white noise standard deviation (std err) = 0.140577
 Chi-square test statistic on first 20 residual autocorrelations = 78.4685
 with probability of a larger value given white noise = 1.58607E-9
 Backforecasting: no Number of iterations performed: 3

De esto concluimos que se puede eliminar la constante del modelo y conservar el parámetro del SAR(1)₁₂. Ahora se procede a analizar los residuales de este modelo.



Se pueden concluir dos cosas a partir del análisis de residuales. Una; el modelo está todavía lejos de ser el óptimo: dos; el siguiente parámetro que se debe agregar al modelo es un AR ordinario. El ACF y PACF residuales nos indican que hace falta una parte AR(2) en el modelo. El PACF se está truncando claramente en 2 y el ACF está decayendo. Existe un valor en la ACF y PACF que llega a ser significativo en 10, pero dado el contexto —que los datos son mensuales y resulta incoherente que suceda algo cada 10 meses— se puede considerar como un dato discrepante.

TESIS CON FALLA DE ORIGEN



| Residual Summary | | |
|---------------------------|------------|---------------------------------|
| Number of observations = | 236 | |
| Residual average = | 7.50917E-3 | |
| Residual variance = | 0.0152827 | |
| Residual standard error = | 0.123623 | |
| Coeff. of skewness = | 0.0696319 | standardized value = 0.436705 |
| Coeff. of kurtosis = | -0.0269663 | standardized value = -0.0845613 |

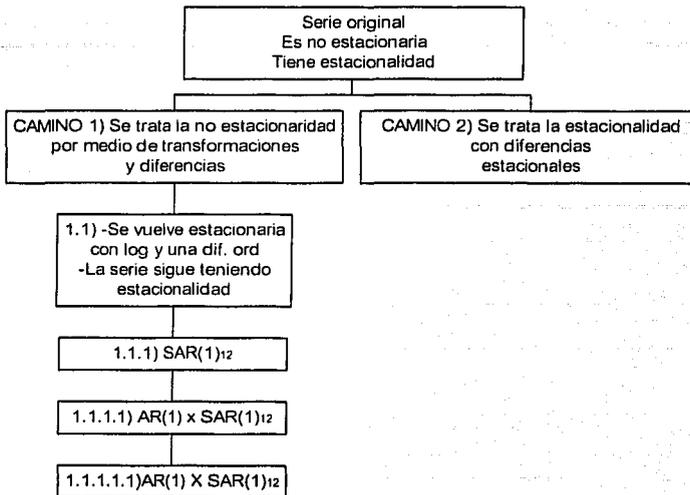
Vemos que el periodograma integral tiene un ajuste prácticamente perfecto. Los ACF, PACF y el resumen de residuales indican que éstos se están comportando como ruido blanco. Pero recordemos que esto podría deberse a una sobreparametrización, es decir, que no estamos siendo parsimoniosos.

Se procede entonces a hacer una prueba para sub y sobre especificación del modelo: quitamos y ponemos un término a la vez con el fin de buscar algún posible modelo que pudiera resultar mejor que el que tenemos.

A grandes rasgos lo que sucede cuando se retira el término SAR es que la ACF y PACF residuales indican claramente que hace falta algún término estacional en el modelo. Si se añade otro más (un SAR₂₄) los residuales no mejoran de manera alguna.

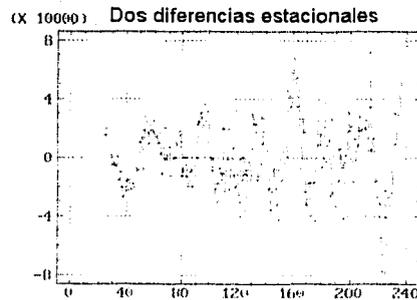
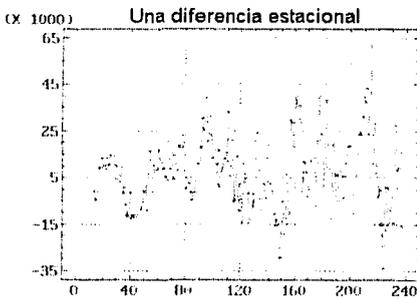
Ya se vio lo que sucede cuando no se cuenta con el término AR(1) o AR(2). Así que podemos concluir que nuestro trabajo con esta serie de tiempo ha terminado.

El modelo resultante es entonces: $y_t = 0.42y_{t-12} - 0.5y_{t-1} - 0.31y_{t-2} + e_t$



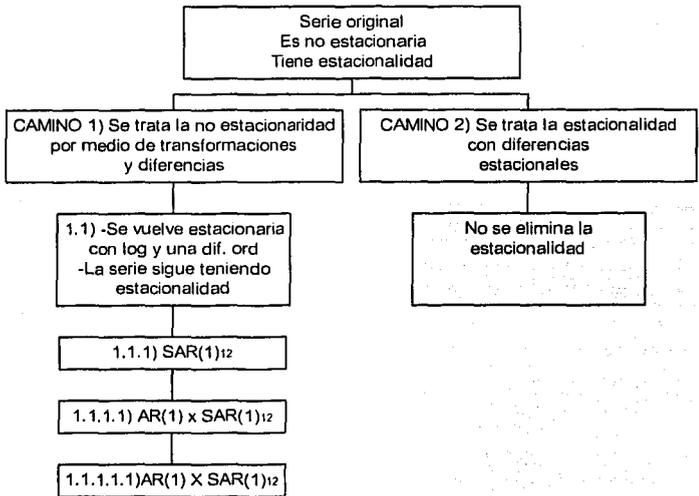
CAMINO 2

Retomando el otro enfoque, de tratar de eliminar la estacionalidad con diferencias, resulta infructuosa ya que, como se ve a continuación, la serie empeora ya sea con una o dos diferencias estacionales.

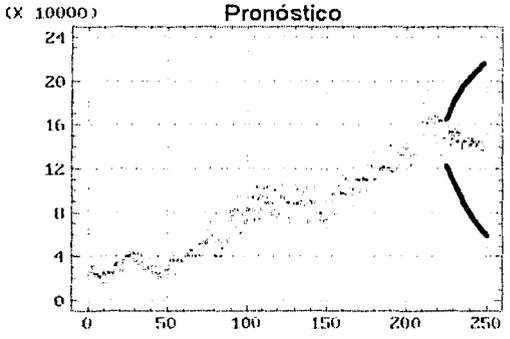


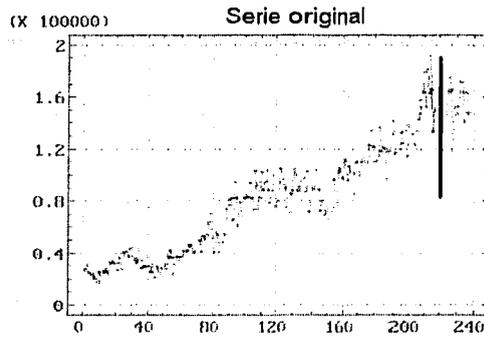
TESIS CON
FALLA DE ORIGEN

**TESIS CON
TÍTULO DE ORIGEN**



Así es como se llega a la conclusión de que el mejor modelo es un $AR(1) \times SAR(1)_{12}$. Por último lo que nos queda por hacer es comparar los pronósticos contra datos reales observados. Para hacer esto retiramos de la serie que acabamos de utilizar, los últimos 12 datos, de manera que nos quede una serie con datos de enero de 1983 a septiembre del 2001. Pronosticamos con esta serie recortada y comparamos los resultados de aplicarle el modelo contra los 12 datos que le extrajimos.





Los pronósticos no son completamente precisos, pero todas las observaciones caen dentro del intervalo de confianza del 95%.

TESIS CON
FALLA DE ORIGEN

TESIS CON
FALLA DE ORIGEN

Conclusiones

A través de toda la tesis se procuraron dos cosas: profundizar en el ciertos temas específicos de manera puntual y se trató de mantener una ubicación en el contexto mundial actual científico, tecnológico e incluso filosófico, con el fin de inculcar en el lector preocupaciones con respecto a las problemáticas actuales. Si bien el objetivo de la tesis es el de explicar el funcionamiento de los procesos estocásticos y series de tiempo, así como los métodos cualitativos y de juicio que existen, le atañen estos problemas al lector si se trata de un estudiante de la Universidad Nacional Autónoma de México.

TEMARIO

El temario sobre el que se trabajó en esta tesis tiene algunos puntos discutibles como cualquier temario. Por ejemplo en un principio fue difícil tratar con los métodos de descomposición y de suavizamiento sin haber explicado primero los conceptos fundamentales de series de tiempo, que no se introducen hasta en el sexto capítulo, cuando los pronósticos mencionados se comienzan a ver el principio de la tesis, en el capítulo 1. Pero dado que el objetivo era trabajar con un temario inamovible, fue posible trabajar con un mínimo suficiente de conceptos que hacían las explicaciones de la primera parte de la tesis menos cargadas de teoría e incluso más amenas. El trabajar con un mínimo de conceptos teóricos "nuevos" fue un artificio que resultó para mí un descubrimiento para dar una buena explicación y al concluir la tesis me percaté que esto es precisamente la intención del temario.

METODOLOGÍA BOX-JENKINS

Puede resultar sorprendente que tantos cálculos respaldados por teoría estadística puedan dar resultados eficientes solo para un periodo corto en el futuro. Después de estudiar diversas investigaciones recientes acerca de la metodología de Box-Jenkins* y después de realizar esta tesis, llego a la conclusión –al igual que algunos investigadores– de que las metodologías de pronósticos de juicio resultan ser por lo menos tan eficientes como una metodología de Box-Jenkins para horizontes de pronóstico de mediano y largo plazo y que la metodología de Box-Jenkins en el corto plazo se debe utilizar solamente cuando se desee una estricta exactitud, dado que es más laborioso que los pronósticos cuantitativos de suavizamiento y de descomposición.

GENERALES

Dada la necesaria personalidad y pertenencia de esta tesis –ya que no es ni pretende ser un trabajo de nivel muy avanzado de investigación– el objetivo al que está suscripto como tal se ha satisfecho. El objetivo de una tesis de nivel profesional es el de ser un trabajo de investigación en el que se trata de comprobar un supuesto que se basa en los conocimientos aprendidos durante los estudios profesionales.

* Geurts e Ibrahim, "Comparing the Box-Jenkins Approach with the Exponentially Smoothed Forecasting Model: Application to Hawaii Tourists" Periódico de Investigación del Mercado, mayo, 1975, 192-188; Hanke y Reitsch(1990)

La hipótesis en mi caso fue la de si era posible escribir y armar material de apoyo didáctico para dos materias de la carrera de Matemáticas Aplicadas y Computación, que son las materias de Series de Tiempo I y Series de Tiempo II. Todo mi empeño se enfocó a tratar de cumplir el objetivo, que a mi juicio se consiguió plenamente.

Todo conocimiento científico está siempre circundado por temas diversos, que no corresponden necesariamente a lo estudiado en la carrera. Por lo que el estudio y la investigación necesarias para hacer una tesis se torna *horizontal*; estudiando a fondo temas tan diversos como son los sociológicos, filosóficos o psicológicos inclusive.

Esto a su vez retroalimenta al lector de la tesis y da un impulso a una visión mucho más amplia del ámbito investigativo y de la sociedad en la que se desenvuelve, además de que encierra de una manera holística todos los conocimientos obtenidos durante la carrera y los relaciona con otras disciplinas.

Las Series de Tiempo y los métodos de pronóstico estudiados en esta tesis son un tema más de las matemáticas aplicadas que forman parte de un conglomerado de instrumentos matemáticos que son altamente aplicables y cuya vigencia y uso se trata de hacer evidente en esta tesis.

Bibliografía

- 1) Brockwell, Davis "Introduction to time series and forecasting". Springer. E.U.1995
- 2) Brown, Robert Goodell , "Smoothing forecasting and prediction of Discrete Time Series"
- 3) Bunge, Mario, "La ciencia, su método y su filosofía", Siglo Veinte, 1996, Argentina.
- 4) Cerejido Mario, "Ciencia sin seso. Locura doble". Siglo Veintiuno, 2000, México
- 5) Cerejido Mario, "Porqué no tenemos ciencia"
- 6) Diebold, Francis "Elementos de pronósticos" International Thompson Editores, 1998, México.
- 7) Durbin and Koopman "Time Series Analysis of State Space Methods"
- 8) Ekeland, Ivar "Al azar. La suerte, la ciencia y el mundo"Gedisa, Colección Limites de la ciencia. España. 1992.
- 9) Erick Jantsch, Hermann Kahn y otros. "Pronósticos del Futuro" Alianza Editorial.1965. México
- 10) González, Videgaray MariCarmen "Modelos de decisión con procesos estocásticos II" Universidad Nacional Autónoma de México, 1990, México.
- 11) González, Videgaray Maricarmen. "Modelos y Simulación". Universidad Nacional Autónoma de México
- 12) Hamilton, James "Time series analysis" Princeton University Press. E.U.1997
- 13) Hanke, John y Reitsch, Arthur."Pronósticos en los negocios". Prentice Hall. 1990
- 14) Harvey, Andrew C. "Time series models" MIT Press. Cambridge, Massachusetts. E.U.
- 15) Koopmans, Lambert H. "The spectral Analysis of time Series", 1998
- 16) Magec, Bryan"Popper" Fontana Modern Masters, 1975,Gran Bretaña.
- 17) Makridakis, Spyros and Wheelwright, Steven."Forecasting Methods for Managment" 5th edition John Wiley and Sons. E.U.
- 18) Makridakis, Spyros; Wheelwright, Steven C. and Hyndman, Rob J. "Forecasting. Methods and Applications" Third edition . John Wiley and Sons, Inc. E.U.
- 19) Makridakis, Spyros and Wheelwright, Steven."Manual de técnicas de pronóstico"
- 20) Meadows, Donella y otros,"Más allá de los límites del crecimiento" El País/Aguilar. México. 1993
- 21) Patricia E. Gaynor, Rickey C. Kirkpatrick. "Introduction to Time Series Modeling and Forecasting in Business and Economics" McGraw-Hill. 1996 E.U.
- 22) Peck, James (editor)"The Chomsky Reader"Pantheon Books, E.U. 1978
- 23) Peña, Daniel; Tiao, George C. and Tsay, Rucy S. "A course in Time Series Analysis" John Wiley and Sons. E.U.
- 24) Pérez Ransanz, Ana Rosa"Kuhn y el cambio científico" Fondo de cultura económica. 1999. México
- 25) Reyes Heroles, Federico "Memorial del mañana", Taurus, 1999, México
- 26) Toffler, Alvin "The Third Wave" William Morrow and Company, Inc.1980, E.U.
- 27) Von Ditfurth, Herbert "No somos sólo de este mundo" Planeta, 1983, México.
- 28) Wei, W.S. William "Time Series Analysis. Univariate and Multivariate Methods." Addison-Wesley Publishing Company, Inc. 1994.
- 29) Wright and Goodwin "Forecasting with Judgment"
- 30) Yaffee, Robert. "Time Series Analysis and Forecasting". Academic Press. E.U.

**TESIS CON
FALLA DE ORIGEN**

Anexos

Anexo 1

TEORÍA DE SISTEMAS

En el contexto de sistema, la Teoría de Sistemas sugiere prestar menos atención a lo individual y más a las interacciones entre estas individualidades. Esta teoría aduce que de esta manera se obtiene un mejor conocimiento del **objetivo superior** del que cada una de las individualidades —componentes— toman parte. Así, resulta inútil estudiar los comportamientos de cada uno de los componentes y luego “sumarlos” para obtener una comprensión total del sistema. El todo constituye más que la simple suma de sus partes.

Por ejemplo, al estar interpretando una pieza de música clásica, los integrantes de una orquesta sinfónica se concentran en lo que está escrito en su partitura y enfocan sus esfuerzos en interpretarla de la manera en que han aprendido a través del tiempo. El director no se preocupa por que todos toquen perfectamente bien ni se preocupa porque todos sigan sus instrucciones al pie de la letra, se preocupa por cómo interactúan todos los músicos. Lo más importante que hace el director es que enfoca toda su atención en guiar a los músicos para poder obtener un sonido “total” que se escuche de la manera que él considera que se debería de interpretar la pieza. Los músicos pueden no estar conscientes del granito de arena que está poniendo cada uno, pero el director sí. La forma en que el director quiere que suene la pieza de música se puede ver como el objetivo del sistema en conjunto.

Otra característica que destacan de los sistemas en este enfoque es el de que condiciones iniciales similares pueden llevar a resultados completamente distintos. Por ejemplo.

$$\begin{aligned} \text{Sistema A: } & (2 * 3) + 6 = 12 \\ \text{Sistema B: } & (2 + 3) * 6 = 30 \end{aligned}$$

$$\begin{aligned} \text{Sistema X: } & (5 * 3) + 30 = 45 \\ \text{Sistema Y: } & (6 + 3) * 5 = 45 \end{aligned}$$

Vemos que los sistemas A y B están compuestos por los mismos elementos y en el mismo orden, pero dan diferentes resultados. Asimismo los sistemas X y Y están compuestos por diferentes “condiciones iniciales” pero llegan a las mismas “conclusiones”.

Un sistema puede ser:

Abierto: Tiene relación permanente con su medio ambiente. Intercambia energía, materia, información. Tiene interacción constante entre el sistema y el medio ambiente.

Cerrado: Hay muy poco intercambio de energía, de materia, de información, etc. con el medio ambiente. Utiliza su reserva de energía interna.

TEORÍA DE COLAS

Definición

Teoría de Colas es el estudio matemático del comportamiento de líneas de espera. Estas se presentan cuando “clientes” llegan a un “lugar” demandando un servicio a un “servidor” el cual tiene cierta capacidad de atención. Si el servidor no está disponible inmediatamente y el cliente decide esperar, entonces se forma en la línea de espera.

El origen de la Teoría de Colas está en el esfuerzo que Agner Krarup Erlang (Dinamarca, 1878 - 1929) realizó en 1909 para analizar la congestión de tráfico telefónico con el objetivo de cumplir la demanda incierta de servicios en el sistema telefónico de Copenhague. Sus investigaciones acabaron en una nueva teoría llamada teoría de colas o de líneas de espera. Esta teoría es ahora una herramienta de valor en negocios debido a que muchos de sus problemas pueden caracterizarse como problemas de congestión llegada -partida.

Una Cola es una línea de espera y la teoría de colas es una colección de modelos matemáticos que describen sistemas de líneas de espera particulares o de sistemas de colas. Los modelos sirven para encontrar un buen compromiso entre costos del sistema y los tiempos promedio de la línea de espera para un sistema dado.

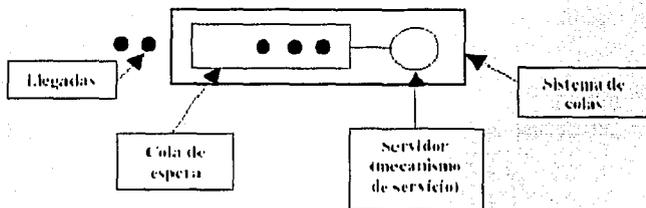
El problema es determinar que capacidad o tasa de servicio proporciona el balance correcto. Esto no es sencillo, ya que un cliente no llega a un horario fijo, es decir, no se sabe con exactitud en que momento llegarán los clientes. También el tiempo de servicio no tiene un horario fijo. Los problemas de Colas se presentan permanentemente la vida diaria: un estudio de EE.UU. concluyó que un ciudadano medio pasa 5 años de su vida esperando en distintas Colas, y de ellos casi 6 meses parado en los semáforos.

Problemas típicos de Teoría de Colas son:

PROBLEMAS TÍPICOS CON LINEAS DE ORIGEN

| Situación | Llegadas | Cola | Mecanismo de Servicio |
|---------------------|---------------------|----------------------|-----------------------|
| Aeropuerto | Pasajeros | Sala de espera | Avión |
| Dpto. de bomberos | Alarmas de incendio | Incendios | Dpto. De Bomberos. |
| Compañía telefónica | Números marcados | Llamadas | Conmutador |
| Panadería | Clientes | Clientes con números | Vendedor |
| Carga de camiones | Camiones | Camiones en espera | Muelle de carga |

Conceptos básicos



Clientes: Término usado en un sistema de colas para referirse a:

- 1) Gente esperando líneas telefónicas desocupadas.
- 2) Máquinas que esperan ser reparadas.
- 3) Aviones esperando aterrizar.

Instalaciones de Servicio: Este término se usa para referirse a:

- 1) Líneas telefónicas
- 2) Líneas telefónicas.
- 3) Talleres de reparación.
- 4) Pistas de aeropuerto.

Llegadas: Es el número de clientes que llegan a las instalaciones de servicio.

Tasa de Servicio: Este término se usa para designar la capacidad de servicio, por ejemplo un sistema telefónico entre dos ciudades puede manejar 90 llamadas por minuto o una instalación de reparación puede de media, reparar máquinas a razón una cada 8 horas. También una pista de aeropuerto en la que aterrizan dos aviones por minuto.

Número de servidores de servicio: Es la cantidad de servidores de que disponemos; número de conmutadores telefónicos; número de puestos de reparación; número de pistas de aterrizaje de un aeropuerto.

Anexo 2. Proporción áurea

Los fractales se encuentran en nuestro cuerpo. Hagamos el experimento de medir nuestra segunda falange del dedo índice (la de en medio) entre la primera (la que tiene uña). Es muy probable que el resultado sea 1.6 aproximadamente. Podemos medir también la distancia de los pies al ombligo y la de la cabeza al ombligo y el resultado será de 1.6 aproximadamente. Esto se debe a que la proporción áurea es una proporción resultante de una ecuación recursiva.

$$x = \sqrt{1 + \sqrt{1 + \sqrt{1 + \dots}}}$$

Hagamos, de manera análoga que con el atractor de Lorenz, una sustitución para x:

$x =$ "raíz de uno más raíz de uno, así hasta el infinito"

$$x = \sqrt{1 + x}$$

Con lo que ya tenemos una ecuación resoluble por álgebra:

$$x^2 - x - 1 = 0$$

Cuyas raíces son $x = \frac{1 \pm \sqrt{5}}{2}$, tomando la raíz positiva, porque es la representable físicamente, que vale 1.618... Esta ecuación lo que *dice* es: "'A' y yo formamos parte de un todo. 'A' que es proporcionalmente menor a mí, como yo soy proporcionalmente menor al todo".

Es imposible encontrar cualquier indicio de que tengamos algo en nuestro cuerpo en paquetes de cuatro, seis o siete. Más bien tenemos todo de uno, dos, tres, cinco. Esto se debe a la serie de fibonacci, que es una serie resultado de aplicar la ecuación que acabamos de ver, pero a los números enteros. Así es posible encontrar que el número de pétalos de una flor sea un número de fibonacci: que la cantidad de ramas entre ramas a la misma latitud con respecto del tallo, sean un número de fibonacci; y los ejemplos son interminables.

La serie de fibonacci es: 1,1,2,3,5,8,13,21,34,55, etc. Cualquier número es la suma de los dos anteriores. Y si se divide cualquier número entre su anterior menor, el resultado convergerá en el infinito al número áureo.

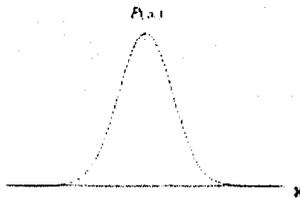
Esta es una prueba más de los resultados extraordinarios de los que es capaz la recursividad y de cómo está presente en toda la naturaleza.

Anexo 3. Ruido blanco

Desde el punto de vista de la acústica, el ruido blanco es el sonido que se produce cuando se combinan y se juntan los sonidos de todas las frecuencias. Si se tomaran todos los tonos imaginables para el oído humano y se combinaran se obtendría el ruido blanco. El adjetivo "blanco" se utiliza para describir a este tipo de ruido debido a la forma en que funciona la luz. La luz blanca es luz que está formada por todos los diferentes colores (frecuencias) de luz que existen. Debido a que el ruido blanco contiene todas las frecuencias, habitualmente se utiliza para enmascarar a otros sonidos. Si se está en un hotel y los sonidos del cuarto contiguo se alcanzan a escuchar, uno puede prender el ventilador y se dejarán de escuchar. El ventilador produce una buena aproximación al ruido blanco. ¿Por qué sucede esto? ¿Por qué el ruido blanco "ahoga" a otros sonidos?. Una forma de entenderlo es de la siguiente manera: cuando dos personas están platicando al mismo tiempo, el cerebro puede distinguir y escuchar una de ellas. Cuando están hablando tres personas a la vez, el cerebro probablemente alcance todavía a divisar una de ellas. Pero si mil personas hablan a la vez, es imposible para el cerebro distinguir una sola voz. Eso es lo que pasa con el ventilador: la voz del cuarto contiguo se vuelve el sonido mil y uno, por lo que el cerebro ya no puede distinguirlo.

Desde el punto de vista estadístico, el ruido blanco gaussiano es un proceso de ruido blanco con distribución Normal, la cual está dada por la función

$$P(x) dx = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2 / (2\sigma^2)} dx, \quad \int_{-\infty}^{\infty} P(x) dx = 1.$$



La curva Normal fue desarrollada matemáticamente en 1733 por DeMoivre como una aproximación de la distribución Normal. Sus estudios no fueron descubiertos hasta 1924 por Kart Pearson. Laplace utilizó la curva Normal en 1783 para describir la "distribución de errores". Después Gauss utilizó la curva Normal para analizar datos astronómicos en 1809. Se acostumbra llamar a la curva Normal la Distribución Gaussiana. La distribución Normal es la distribución estadística más usada. Las razones principales para esto son:

- 1) La Normalidad aparece naturalmente en muchas mediciones en la Biología, Física y Sociología.
- 2) La Normalidad es importante en la inferencia estadística.

TRIPS CON
FALLA DE ORIGEN

Anexo 4. Criterios de bondad de ajuste

SIMETRÍA

Si una distribución es simétrica, el sesgo vale cero. Cuanto mayor es el valor absoluto del estadístico de sesgo, más asimétrica es la distribución. Un valor positivo grande indica que el extremo o cola de la derecha es largo, y un valor negativo grande indica un extremo izquierdo largo. La simetría se estima como sigue:

$$\hat{S} = \frac{\frac{1}{T} \sum_{i=1}^T (y_i - \bar{y})^3}{\hat{\sigma}^3}$$

en donde T es el número de observaciones en la serie y

$$\hat{\sigma} = \sqrt{\frac{1}{T} \sum_{i=1}^T (y_i - \bar{y})^2} \quad \text{y} \quad \bar{y} = \frac{1}{T} \sum_{i=1}^T y_i$$

CURTOSIS

La curtosis de una variable aleatoria es una medida del espesor de los extremos de su curva de distribución en relación con el de la distribución normal. La curtosis de una variable aleatoria normal es igual a 3. Si la curtosis es mayor que 3 indica que los extremos son "gordos" o que la distribución es leptocúrtica; esto es, la distribución tiene mayor masa de probabilidad en los extremos que la distribución normal. Cuando la distribución está más concentrada en el centro se dice que es platocúrtica y la variable tiene platocurtosis.

La curtosis se estima como sigue

$$\hat{K} = \frac{\frac{1}{T} \sum_{i=1}^T (y_i - \bar{y})^4}{\hat{\sigma}^4}$$

Residual Summary

Number of observations = 150
Residual average = -3.34914E-3
Residual variance = 1.00371
Residual standard error = 1.00185

Coeff. of skewness = 0.0318814 standardized value = 0.159407
Coeff. of kurtosis = -0.197432 standardized value = -0.493579

R CUADRADA (R²)

R cuadrada debe estar entre cero y uno. Es el porcentaje de la varianza de y_i que explican las variables incluidas en el modelo. R^2 mide el éxito de la ecuación de regresión para predecir y_i dentro de la muestra. Así que lo ideal sería que estuviera lo más cercana posible a 1.

$$R^2 = 1 - \frac{\sum_{i=1}^T e_i^2}{\sum_{i=1}^T (y_i - \bar{y})^2}$$

204

TESIS CON
DE ORIGEN

R CUADRADA AJUSTADA (\bar{R}^2)

El problema con R^2 es que puede ser muy optimista con los modelos que incluyen más parámetros. Para compensar esto se usa \bar{R}^2 que resta la cantidad k de parámetros incluidos, es decir, grados de libertad.

$$\bar{R}^2 = 1 - \frac{\frac{1}{T+k} \sum_{i=1}^T e_i^2}{\frac{1}{T+1} \sum_{i=1}^T (y_i - \bar{y}_i)^2}$$

Como se puede ver en ambos casos se trata de la división de lo que varía el pronóstico con respecto de los valores dentro de la muestra entre lo que la variable cambia con respecto a su propia media.

CRITERIO DE INFORMACIÓN DE AKAIKE (AIC)

Este es un estimado de la varianza del error de pronóstico fuera de la muestra, que penaliza los grados de libertad.

$$AIC = e^{\left(\frac{2k}{T}\right)} \frac{\sum_{i=1}^T e_i^2}{T}$$

CRITERIO DE INFORMACIÓN DE SCHWARZ (SIC)

Este tiene la misma interpretación que el de Akaike, pero penaliza todavía más los grados de libertad

$$SIC = T \left(\frac{k}{T}\right) \frac{\sum_{i=1}^T e_i^2}{T}$$

De estos dos, el criterio de Schwarz selecciona modelos más parsimoniosos. Para verificar la validez de un modelo también se puede probar hacer la estimación tanto con *backcasting** como sin ella, y observar si se generan los mismos valores para los estimadores. Probar a su vez si los estimadores son consistentemente significativos.

* Pronosticar hacia atrás en inglés.

205

Anexo 5. Pronósticos mal realizados

"Las computadoras en el futuro puede que no pesen más de 1.5 toneladas"
–Mecánica Popular, pronosticando la marcha implacable de la ciencia, 1949–

"Yo creo que hay un mercado mundial para quizá cinco computadoras"
–Presidente de IBM, 1943 –

"He atravesado todo el país a lo largo y ancho y he hablado con la mejor gente, y les puedo garantizar que el procesamiento de datos es una moda que no pasará de este año "

–El editor encargado de los libros de negocios de Prentice Hall, 1957–

"Pero...¿para qué sirve?"

–Ingeniero de la división de Sistemas de Cómputo Avanzado de IBM, 1968,
comentando acerca del microchip–

"No hay razón alguna para que alguien quiera una computadora en su hogar"
–Presidente y fundador de Digital Equipment Corp., 1977–

"El 'teléfono' tiene demasiadas limitaciones como para ser seriamente considerado como un medio de comunicación. El dispositivo es inherentemente de ningún valor para nosotros "

–Memo interno de Western Union, 1876–

"La caja musical inalámbrica no tiene ningún valor comercial. ¿Quién pagaría por un mensaje que no va a dirigido a nadie en específico?"

–Los socios de David Sarnoff respondiéndole acerca de sus exigencias de invertir en el radio en la década de 1920 –

"El concepto es interesante y está bien formado, pero para obtener un calificación más alta que una 'C' la idea debe ser factible"

–Un profesor de administración de Yale en respuesta al trabajo de Fred Smith en el que propone un servicio confiable de mensajería que entregue de un día para otro. (Smith fundaría después Federal Express Corp.)–

"¿Quién diablos quiere oír a los actor hablar"

–Warner Brothers, 1927–

"Estoy contento de que sea Clark Gable el que va a hacer el ridículo y no Gary Cooper"

–Gary Cooper opinando con respecto a su decisión de no tomar el papel principal en "Lo que el viento se llevó"–

"Máquinas voladoras que pesen más que el aire son imposibles"

–Presidente de las Royal Society, 1895–

"Si lo hubiera razonado, no hubiera realizado el experimento. Los libros estaban llenos de ejemplos que decían que no se puede hacer esto"

–Spencer Silver acerca del trabajo que lo llevó a la invención de los adhesivos "Post-it" de 3-M –

"Entonces fuimos a Atari y les dijimos, 'Oigan, tenemos esta cosa increíble, incluso fabricada con algunas de sus partes de ustedes, ¿qué piensan acerca de patrocinarnos? O mejor se la damos. Sólo queremos fabricarlo. Páguenos un salario, vendremos a trabajar para ustedes' Y ellos dijeron 'No.' Entonces después fuimos a Hewlett-Packard

y ellos nos dijeron, 'Oigan, no los necesitamos a ustedes. Ni siquiera han terminado la Universidad.'"

–Steve Jobs, fundador de Apple Computer Inc. comentando acerca de sus intentos de lograr que Atari y HP se interesaran en el más reciente invento de él y de Steve Wozniak: la computadora personal–

"¿Perforar para extraer petróleo? ¿Quiere decir que quiere perforar el suelo para tratar de encontrar petróleo? Usted está loco."

–Perforadores a los que Edwin L Drake trató de contratar para su proyecto para perforar y extraer petróleo en 1859–

"Las acciones han alcanzado lo que parece un nivel alto permanente"

–Profesor de Economía de la Universidad de Yale, 1929–

"Los aviones son interesantes, pero de ningún valor militar"

–Profesor de Estrategia, Ecole Supérieure de Guerre–

"Todo lo que puede ser inventado, ha sido inventado"

–Comisionado de la Oficina de Patentes de Estados Unidos, 1899–

"La teoría de las bacterias de Louis Pasteur es una ficción ridícula"

–Profesor de fisiología de Toulouse, 1872–

"640 k deberían ser suficientes para cualquiera"

–Bill Gates, 1981–

**TESIS CON
FALLA DE OMBEN**

Glosario

Aleatoriedad. En el **contexto** de series de tiempo suele referirse al ruido blanco (ver).

AR. Nomenclatura con la que se abrevia a los procesos "Autorregresivos". Son procesos estocásticos cuyo estado actual depende de los estados posteriores en el tiempo.

Brainstorming. Ver "Lluvia de ideas".

Causalidad. Perspectiva que supone que la realidad es explicable a través de causas y efectos. Bajo ésta, sería posible formular una ecuación para el comportamiento del cosmos si se contara con una cantidad infinita de información y se pudieran cuantificar todas las causas y todos los efectos existentes. Permite la existencia del determinismo y el libre albedrío.

Casualidad. Perspectiva que supone que la realidad funciona bajo un "orden supremo" y que la realidad se rige a través de leyes estadísticas. Bajo esta perspectiva es imposible la existencia de un libre albedrío, ya que todos hacemos lo que nos corresponde hacer, es decir, somos el resultado de nuestro entorno.

Choques aleatorios. Se refiere al comportamiento "inexplicable" para un cierto enfoque de modelos, como pueden serlo, los modelos ARIMA. La ventaja que trae consigo la consideración de los choques aleatorios dentro del modelo, es la de permitir ver y manejar una realidad **causal** sin dejar completamente de lado las ventajas del enfoque indeterminista de la realidad (casualidad). Se acostumbra manejarlo como **ruido blanco**, ya que siendo éste una representación de la distribución normal, tiene propiedades bien conocidas para la estadística.

Lluvia de ideas. Método en el cual los miembros del grupo proponen tantas ideas como les es posible, sin criticarlas o calificarlas. El objetivo es el incitar al explayamiento, el libre pensamiento y el pensamiento creativo que pueda generar ideas –aunque inesperadas y poco convencionales– que resulten útiles. (Definición proporcionada por la World Future Society).

MA. Nomenclatura con la que se abrevia a los procesos de "Medias Móviles" (en inglés *Moving Averages*). Son procesos estocásticos cuyo estado actual depende de los choques aleatorios posteriores en el tiempo.

Proceso estocástico. Proceso que se considera explicable a través de modelos matemáticos cuyas variables –aunque numéricas– una vez resueltas pueden tomar un valor entre una serie de valores. Esta serie de valores está determinada por una distribución probabilística.

Proceso Markoviano. Sea la secuencia $x_1, \dots, x_n, \dots, x_T$; $x_i \in X$; $i \in \{1, \dots, T\}$. Para que un proceso sea Markoviano de orden n , se tiene que cumplir que:

$$P\{x_{T+1} / x_T, \dots, x_1\} = P\{x_{T+1} / x_T, \dots, x_{T-n+1}\}$$

Es decir, debe haber una cierta periodicidad en el tiempo, de manera que sea posible explicar cualquier estado del proceso en base a una *pequeña* cantidad 'n' de estados anteriores y no en base a toda la historia del proceso.

Ruido blanco. *Ruido:* término que se asigna dado que se trata de un comportamiento impredecible e inrastreado. Esto quiere decir que no se sabe de dónde viene desde un punto de vista causal. *Blanco:* se le denomina así por ser la mezcla de muchas distribuciones a la vez, en clara analogía con el espectro luminoso. Intrínsecamente está

201

relacionado con la **exponencialidad** y la impredecibilidad a largo plazo con los modelos ARIMA. La cuantificación del RB y su consideración dentro de los modelos ARIMA acepta la posibilidad de otros factores que son "en potencia" significativos – que no son significativos en el momento de formular el modelo– pero que empiezan cada una a mostrar su presencia e importancia para el modelado del proceso a través de "pequeñas" contribuciones individuales al ruido blanco.

Sistema. Conjuntos de elementos que guardan estrechas relaciones entre sí, que mantienen al sistema directo o indirectamente unido de modo más o menos estable y cuyo comportamiento global persigue, normalmente, algún tipo de objetivo.

Think-tank. Instituto de investigación u organización que tiene como función resolver problemas complejos o predecir o planear inventos futuros en el área militar, política o social.