UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

ANÁLISIS DE PRONÓSTICO CON TÉCNICAS ESTADÍSTICAS DE SERIES DE TIEMPO Y REDES NEURONALES

TESIS

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

PRESENTA:

CHRISTIAN MENA RUVALCABA

ASESOR DE TESIS: MTRO. IVAN MEJÍA

ACATLÁN, ESTADO DE MÉXICO. DICIEMBRE DE 2005.

Agradecimientos

A mi mama Beatriz, por su apoyo y cariño incondicional y por estar siempre cuando más la necesite.

A mis hermanas, Miriam y Karla por su cariño y apoyo.

A mis amigos Alberto, Gustavo, Gloria, Mayra, Pedro, Alfredo y Marcos por su gran ayuda y amistad.

A mis profesores, por haber compartido un poco de sus conocimientos y paciencia.

Un agradecimiento especial a mis profesores y amigos, Jose Eliud, Carmen, Pablo Pérez Akaki, Víctor Ulloa, Maricarmen Videgaray y desde luego a mi asesor Ivan Mejía.

Índice general

Introducción

Todos sabemos que el poder predecir el futuro es imposible, pero eso no quiere decir que no nos podamos aproximar a lo que pueda suceder y obviamente, entre mejor nos acerquemos a lo que pueda pasar, es posible tomar mejores precauciones. Un ejemplo sería una compañía, en la cual se desea saber cuanto podrían ser sus ventas para el siguiente periodo, y así poder tomar las medidas pertinentes, ya sea para comprar la suficiente materia prima o ver que se puede hacer para mejorar esas ventas. Este tipo de problema puede ser atacado por los métodos clásicos de series de tiempo(siempre y cuando se cuente con historial de los datos a pronosticar), aunque recientemente se ha desarrollado una nueva teoría para poder predecir series de tiempo, las cuales son conocidas como redes neuronales artificiales. Por esta razón, se trabajarán estas dos técnicas de pronóstico para valorar la eficacia de ambas.

De está forma en el primer capítulo se establece una clasificación de los métodos de pronóstico, dividiéndolos en: Subjetivos, causales, de extrapolación y no tradicionales(Redes neuronales), con el fin de poder dar una clasificación inicial tanto a las series de tiempo como a las redes neuronales.

El propósito del segundo capítulo es dar a conocer la teoría para los modelos tradicionales de series de tiempo con el fin de pronosticar, tanto para los métodos de alisamiento(atenuación), los causales(regresión lineal), así como para la metodología de Box y Jenkins. Además, se explican algunos conceptos fundamentales concernientes a series de tiempo, y la forma en que resulta conveniente su uso.

La formalización de las redes neuronales artificiales se hace en el capítulo 3,

el cual nos da un panorama general de las redes neuronales, desde su unidad fundamental, la neurona, hasta la agrupación de muchas de ellas (redes neuronales), así como los tipos de redes neuronales que existen, su clasificación, sus tipos de arquitectura y sus mecanismos de aprendizaje. Finalmente se analiza la estructura de la red y el mecanismo de aprendizaje que se utilizará para predecir.

Toda la teoría expuesta, tanto de las técnicas tradicionales de series de tiempo como de redes neuronales es aplicada finalmente a 5 series, constituyendo esto el 4 capítulo.

0.1. Objetivo

Aplicación de Redes Neuronales al problema de pronóstico y comparación con los resultados obtenidos a través de técnicas de series de tiempo tradicionales.

Objetivos particulares

- Describir los fundamentos teóricos de las redes neuronales artificiales y su aplicación al problema de pronóstico de series de tiempo.
- Describir la metodología estadística tradicional para la solución al problema de pronóstico de series de tiempo.

0.2. Hipótesis

Comparar dos metodologías distintas para la predicción de series de tiempo y demostrar que es posible obtener resultados con redes neuronales comparables a los obtenidos con métodos estadísticos tradicionales.

Capítulo 1

Métodos de Pronóstico

1.1. Situación actual

En las sociedades donde vivimos, el riesgo y la incertidumbre ante el futuro, aparecen como problemas a los que los distintos agentes tratan de adaptarse de diferentes maneras. Para reducir el grado de incertidumbre se suele recurrir a la elaboración de previsiones que tratan de anticipar la evolución de algún fenómeno. El disponer, en el presente, de un conocimiento sobre el futuro, aunque sea de forma aproximada, facilita la toma de decisiones en las que incurrimos cuando pretendemos anticiparnos a una realidad determinada.

Existen en la actualidad muchos métodos de pronóstico que varían en precisión y complejidad y cada uno de ellos con una aplicación especial que hace de su selección un problema de decisión influido por diversos factores, como por ejemplo, la validez y disponibilidad de los datos históricos, la precisión deseada del pronóstico, el costo del procedimiento, los beneficios del resultado, los períodos futuros en que se desee pronosticar y el tiempo disponible para hacer el estudio, entre otros.

Es evidente la importancia que el pronóstico tiene en la toma de decisiones como son problemas de inventario, planeación de la producción, planeación financiera, control de procesos, etc. Por ejemplo Pensemos en un empresario

que quiere diseñar un plan estratégico para su empresa pero que necesita para ello conocer cuales son sus ventas futuras. La decisión no se basará únicamente en los datos actuales disponibles sobre las ventas, sino que también se apoyará en sus valores históricos, independientemente de la consideración de otras técnicas de previsión. Es claro que el riesgo que se tiene al tomar una decisión no puede eliminarse totalmente, pero el propósito del pronóstico es reducir ese riesgo y dependerá del método que se utilice el disminuir o aumentar la magnitud del error que se pueda cometer. Hay que tomar en cuenta que al definir las variables que van a ser analizadas, se determina de hecho la forma que tendrá el pronóstico, mientras que la exactitud deseada dependerá, en gran medida, de la calidad de la información, dicho de otra manera, la certeza en la calidad de los resultados de un pronóstico jamás podrá ser mejor que la calidad de los datos.

La calidad de los datos depende de cuidar tres aspectos principales: su origen, disponibilidad y adecuación al ajuste.

El origen de los datos depende directamente del tipo de variable que va a ser pronosticada, debiendo vigilar que la fuente donde se recolecten sea lo más confiable y observando cuidadosamente cómo son obtenidos, ya que existe la posibilidad de presentarse el caso de escasez de datos o abundancia de los mismos. Si éste es el caso, es conveniente establecer algún criterio para clasificarlos. Asimismo, si se trabaja sólo con una muestra de la población, es importante cuidar la representatividad de dicha muestra y por ende, todos los aspectos inherentes a ello.

Otro aspecto es la disponibilidad de los datos, esto es, el tener la seguridad de que estarán disponibles en el momento que se van a necesitar y si el resultado de su recopilación sigue ajustándose a nuestros requerimientos originales.

Por último, hablemos del *ajuste de los datos*. Es *muy común que los datos* deban modificarse de alguna manera antes de iniciar el análisis para pronosticar con el objeto de que resulten más eficientes. Esta modificación recibe el nombre de *Ajuste* y puede llevarse a cabo de varias formas: una de ellas y quizá la más común es por períodos de tiempo; otra forma es la transformación de los mismos, o sea la aplicación de una transformación simple a los datos originales.

En lo concerniente al factor tiempo debemos conocer tres de sus componentes principales como son: *Período, horizonte e intervalo del pronóstico.*

El período es la unidad básica de tiempo en que el pronóstico es hecho: mensual, trimestral, anual, etc. El horizonte es el número de períodos que en el futuro cubrirá el pronóstico, o sea el alcance del mismo. La frecuencia con la cual el pronóstico debe ser revisado se denomina intervalo del pronóstico y con frecuencia éste y el período son iguales, de tal forma que el pronóstico será revisado cada período.

1.2. Métodos de Pronóstico

Definiremos brevemente un método de pronóstico como un proceso para predecir el futuro, el cual nos dirá lo que es más seguro que ocurra en el futuro.

Los métodos de pronósticos se pueden dividir principalmente en dos tipos: Métodos cualitativos y métodos cuantitativos.

1.2.1. Métodos Cualitativos

También llamados tecnológicos, porque históricamente se usaron primero para pronosticar cambios tecnológicos. Este método es apropiado cuando los datos confiables son escasos o difíciles de emplear o cuando el tiempo para elaborar el pronóstico es escaso, en otras palabras, la posición central en estos métodos no la tienen los datos pasados, sino la experiencia de las personas. Frecuentemente, se usa la experiencia y buen juicio de varios expertos. Existen tres métodos principales, los cuales son:

Método Subjetivo(o Intuitivo). Se basan en el juicio personal y pueden hacer uso de cualidades como la intuición, la opinión de un experto y la experiencia, un ejemplo es el método Delphi(o Delfos).
Método Exploratorio(o Prospectivo). Se parten de las experiencias pasadas y presentes para proyectar al futuro, sopesando las diferentes posibilidades.

Método Normativo(o Deductivo). En éstos se procede al revés, se parte de las metas u objetivos a lograr en el futuro y se analiza qué se necesita para lograrlos y con base en eso poder pensar cuándo sucederán los eventos previstos para el futuro.

1.2.2. Métodos Cuantitativos

Se basan en datos históricos. Esta información pasada se encuentra en forma numérica. Estos métodos hacen una extrapolación del pasado, es decir, el carácter distintivo mostrado por datos relevantes en el pasado, se traslada al futuro basándose en dos principios generales:

- El período siguiente será igual al período presente.
- El patrón que rige las tendencias de las variables del presente período al siguiente será el mismo en relación con el patrón que rigió el período pasado con el presente.

Los pronósticos basados en extrapolación, como un análisis de series de tiempo, recurre a las tendencias pasadas o presentes a fin de proyectar los acontecimientos futuros. También son utilizados cuando se cuenta con suficientes datos estadísticos o confiables para especificar las relaciones existentes entre variables fundamentales. Los métodos más usados son:

- *Método Causal(o estructural).* Aquí se intenta identificar las relaciones entre variables que existieron en el pasado. Luego, se supone que las relaciones continúan siendo validas en el futuro. Generalmente se asume que existe una relación lineal e independiente entre ellas, por ejemplo el modelo de regresión múltiple o los modelos econométricos.
- *El método extrapolativo(Series de Tiempo).* En el cual se efectúan pronósticos para una variable particular, usando únicamente la historia previa de esa variable. Se supone que los patrones identificados en el pasado se extienden hacia el futuro.

- *Redes Neuronales.* Existen , además de los métodos clásicos de pronóstico, diversos métodos de pronóstico no tradicionales, dentro de los cuales se encuentran las redes neuronales artificiales. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos y de abstraer características esenciales a partir de entradas que representan información irrelevante Jose R. Hilera [15]

1.3. ¿Cuándo usarlos?

Analizando la clasificación de los métodos de pronóstico expuesta:- Subjetivos, Causales y de extrapolación. Podemos inferir que:

- Los *Métodos Subjetivos* son apropiados para producir predicciones en áreas no estructuradas y aún más, en áreas no desarrolladas, abarcando largos períodos, un ejemplo ya mencionado son los avances tecnológicos.
- Los *Métodos Causales* son aplicables a predicciones que abarquen períodos cortos de tiempo dado que conforme aumenta el intervalo de predicción, aumentan también los errores en las predicciones disminuyendo en consecuencia la confianza en los eventos pronosticados.
- Los *Métodos extrapolativos* generan predicciones con un alto nivel de confianza para períodos cortos de tiempo.

Los métodos de extrapolación y los causales, no pueden utilizarse bajo las siguientes condiciones:

1. Cuando no exista una teoría formal que permita la formulación de hipótesis.
2. Cuando no existan o no se conozcan datos pasados.
3. Cuando el intervalo de predicción sea a largo plazo.

A continuación presentamos un análisis elemental de cuatro métodos de pronóstico, cada uno de los cuales pertenece a una de las clases expuestas anteriormente.

Métodos subjetivos : Técnica Delphi.

Métodos Causales : Regresión lineal.

Métodos de extrapolación : Series de tiempo de Box y Jenkins.

Métodos no tradicionales : Redes Neuronales

.

Técnica Delphi

Descripción: Se interroga a un panel de expertos por medio de una secuencia de cuestionarios, en los cuales las respuestas al primero se utilizan para producir el siguiente cuestionario. De esta manera, toda la información, de la que disponen algunos expertos, pero otros no, se pasa a éstos últimos y eso permite que todos los expertos tengan acceso a toda la información.
Predicción:

- Corto plazo(máximo un año): Regular a bueno.
- Mediano plazo(de uno a cuatro años): Regular a bueno.
- Largo plazo(de cuatro años en adelante): Regular a bueno.

Datos que se requieren: Un coordinador que emita la secuencia de cuestionarios editando y consolidando las respuestas.

Regresión lineal

Descripción: Se emplean para relacionar una variable dependiente Y con las variables independientes $X_1, X_2, ..., X_t$. Se denominan modelos de regresión lineal porque expresan el valor medio de Y para valores dados de $X_1, X_2, ..., X_t$ como una función lineal de un conjunto de parámetros desconocidos.

Precisión:

- Corto plazo: regular a buena.
- Mediano plazo: Regular.
- Largo plazo: Mala.

Datos que se requieren: Mientras más historial exista, mejor.

Series de tiempo de Box y Jenkins

Descripción: La serie de tiempo se dota de un modelo matemático que es óptimo en el sentido de que asigna menos errores a la historia que los demás modelos. Habrá que identificar el tipo de modelo y entonces estimar sus parámetros. Aparentemente, ésta es la rutina estadística más precisa que se posee en la actualidad, pero también es uno de los métodos más costosos y consumidores de tiempo.
Precisión:

- Corto plazo: Muy buena a excelente.
- Mediano plazo: Regular.
- Largo plazo: Mala.

Datos que se requieren: Tener mucho historial, es muy valioso.

Redes Neuronales

Descripción: Son sistemas dinámicos autoadaptativos. Son adaptables debido a la capacidad de autoajustarse por parte de los elementos procesales(neuronas) que componen el sistema. Son dinámicos, pues son capaces de estar constantemente cambiando sus pesos para adaptarse a las nuevas condiciones, en el proceso de aprendizaje, los enlaces ponderados de las neuronas se ajustan de manera que se obtengan unos resultados específicos [José R. Hilera, 1995]. Precisión:

- Corto plazo: Muy buena a excelente.
- Mediano plazo: Mala a Regular.
- Largo plazo: Muy mala.

Datos que se requieren: Tener mucho historial, es muy valioso.

Capítulo 2

Series de Tiempo

Definición: Una serie de tiempo se puede definir como una sucesión de observaciones correspondientes a una variable en distintos momentos de tiempo. Así las series pueden tener una periodicidad anual, semestral, mensual, trimestral, etc. según los períodos de tiempo en que vengan recogidos los datos que la componen. Un ejemplo gráfico de una serie temporal o de tiempo esta dada en la figura 2.1

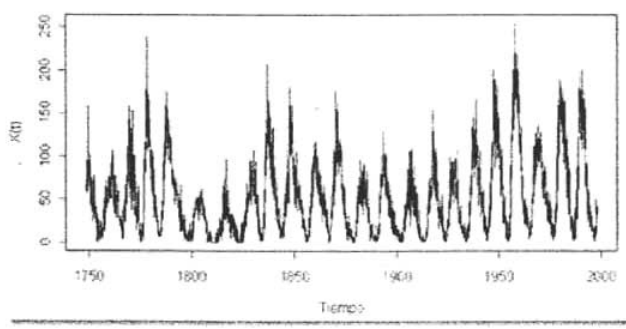


Figura 2.1: Representación de una serie temporal

2.1. Descomposición clásica de una serie de tiempo

Dentro de la descomposición clásica se encuentran 4 componentes principales los cuales son: Tendencia, estacionalidad, ciclo y componente irregular.

2.1.1. Tendencia

Se considera tendencia al movimiento suave y regular de la serie a largo plazo. Es una componente que reviste gran interés dado que refleja la dirección del movimiento de una determinada variable. De esta forma, puede detectarse si a largo plazo la serie adopta una dirección, ya sea de crecimiento, decrecimiento o estabilidad.
La predicción de esta componente suele ser en muchos casos el objetivo del análisis de series de tiempo a través de los modelos de ajuste de tendencia, donde se supone que la serie carece de variaciones estacionales y cíclicas(la cuales veremos más adelante). En resumen, la tendencia es la dirección general de la variable en el período de observación, es decir el cambio a largo plazo de la media de la serie. La figura 2.2 nos muestra una serie con tendencia positiva.

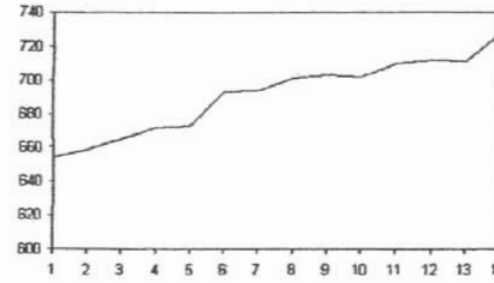


Figura 2.2: Serie temporal con tendencia

2.1.2. Ciclo

Esto se refiere a el tiempo de oscilaciones, cerca de una linea de tendencia o curva. Estos ciclos pueden o no ser periódicos, esto es, que pueden o no, seguir un patrón después de un intervalo igual de tiempo. Los movimientos son considerados cíclicos si ellos recurren en intervalos mayores a un año. La figura 2.3 nos muestra una serie temporal si suponemos que las líneas verticales nos representan 10 años, entonces podemos ver una serie con ciclo cada 10 años.

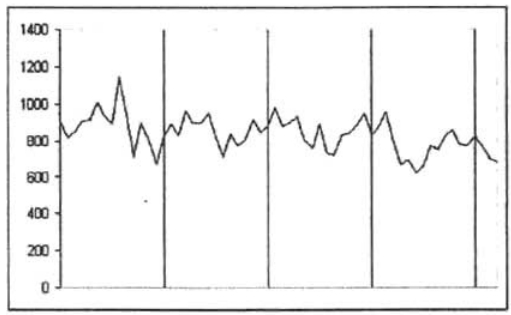


Figura 2.3: Serie temporal con ciclo

2.1.3. Variación estacional (o periódica)

También conocido como movimiento estacional. Esto se refiere al comportamiento idéntico o casi idéntico que una serie de tiempo sigue durante meses, cuatrimestres o semestres de años consecutivos. Tales movimientos se repiten cada año. Por ejemplo las compras en navidad, donde las ventas se incrementan de forma notoria en Diciembre. La figura 2.4 nos muestra una serie con estacionalidad, suponiendo que son ventas trimestrales(cada año termina en cada linea vertical), podemos ver que se repite el comportamiento cada año

y que tanto las ventas del primero como del ultimo trimestre son mas altas que los otros dos.

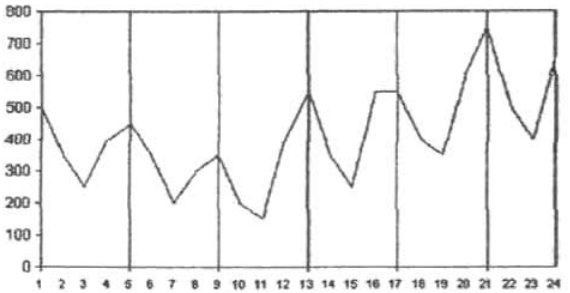


Figura 2.4: Serie temporal con estacionalidad

2.1.4. Componente irregular

Esta componente está constituida por una especie de "caja de Pandora" donde se incluirían las variaciones de la serie cuyo comportamiento desconocemos. Se caracteriza, porque no responde a un *comportamiento sistemático o regular* y en consecuencia no sería posible su predicción.
Aunque el enfoque clásico atribuye esta irregularidad al azar, en los análisis habituales de series de tiempo, algunas veces es posible encontrar la causa que provoca esta irregularidad. Por ejemplo supongamos que estamos analizando el comportamiento de las ventas mensuales en unos grandes almacenes desde 1996 hasta 2005 y observamos gráficamente que existe una disminución en las ventas en el mes de Noviembre de 1999, esta desviación, en principio, la trataríamos como irregular, lo cual no quiere decir que se ignore la causa de la misma. Buscando en los archivos históricos, encontramos que en ese mes hubo una huelga importante de los trabajadores que, naturalmente, afectó a los niveles de venta de estos almacenes.

Muchas veces, esos picos irregulares u observaciones anormales que aparecen en los gráficos de las series y que sabemos que la causa que los provocó no es de esperar que influya en el futuro de la serie, conviene eliminarlos, o no tomarlos en cuenta.

Dentro de esta componente distinguiremos, por tanto, aquellas irregularidades cuyas causas se pueden identificar(factor errático) y aquellas atribuibles al azar(factor aleatorio).

En una serie no tienen por qué estar presentes todas estas componentes. Por ejemplo, las series que tienen una periodicidad anual están desprovistas de estacionalidad, con lo que, en estas series, el interés se centrará en el estudio de la tendencia. La componente irregular, sin embargo, deberá ser incluida siempre, ya que estamos trabajando con series que no son deterministas y por tanto estarán afectadas por alguna perturbación, al menos, de carácter aleatorio.
En la siguiente figura (2.5) vemos un ejemplo de una serie temporal en la que se aprecia la existencia de las distintas componentes comentadas.

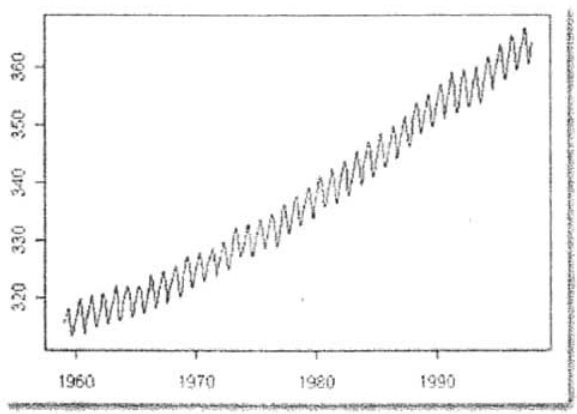


Figura 2.5: Serie temporal con todos sus componentes

A continuación veremos técnicas de pronóstico de diferentes tipos de datos.

Técnicas de pronóstico para datos estacionarios

Una *serie estacionaria* es aquella cuyo valor promedio no varía a través del tiempo(sin tendencia). Estas situaciones se presentan cuando los patrones de demanda que influyen sobre la serie son relativamente estables. En su forma más simple, el pronóstico de series estacionarias comprende el uso de la historia disponible de las series para estimar su valor promedio, el cual se convierte después en el pronóstico de valores futuros.
Las técnicas de pronóstico estacionarias se emplean siempre que:

- *Las fuerzas que generan una serie se han estabilizado y al medio en el que existe la serie permanece relativamente sin cambios.*
- *Se refiere a un modelo muy sencillo debido a la falta de datos o para facilitar su explicación o implementación.* Un ejemplo sería cuando un negocio u organización es nuevo y hay disponible muy poca información histórica.
- *Se puede lograr la estabilidad haciendo correcciones sencillas.*
- *La serie se puede transformar en una serie estable.* Como ejemplo, está la transformación de series mediante logaritmos, raíces cuadradas o diferencias.

Varias técnicas que se podrían considerar al pronosticar series estacionarias son los métodos no formales, los métodos de medias simples, los métodos de medias móviles, atenuación exponencial y de Box-Jenkins

Técnicas de pronóstico para datos con una tendencia

Anteriormente definimos a una serie con *tendencia* como una serie que contiene un componente de largo plazo que representa el crecimiento o declinación de la serie a través de un período amplio, de modo que se espera que

aumente o disminuya durante el período para el que se desea el pronóstico. Es común que las series económicas contengan una tendencia.

Las técnicas de pronóstico para series con tendencia se utilizan siempre que:

- *Una productividad creciente y la nueva tecnología conducen a cambios en el estilo de vida.* Como ejemplos se pueden citar la demanda de componentes electrónicos, que se incrementó con el advenimiento de la computadora; y el uso del ferrocarril que disminuyo con la llegada del avión.
- *El incremento en la población provoca un incremento en la demanda de bienes y servicios.*
- *El poder de compra del dólar afecta las variables económicas por causa de la inflación.* Los ejemplos son salarios, costos de producción y precios.
- *Aumenta la aceptación en el mercado.*

Las técnicas a considerar al pronosticar series con tendencia son: medias móviles dobles, atenuación exponencial lineal de Brown, atenuación exponencial lineal de Holt, atenuación exponencial cuadrática de Brown, regresión simple, modelo de Gompertz, curvas de crecimiento y modelos exponenciales.

Técnicas de pronóstico para datos con estacionalidad

Anteriormente definimos una serie *estacional* como una serie de tiempo con un patrón de cambio que se repite a sí mismo año tras año. Por lo regular, el desarrollo de una técnica de pronóstico estacional comprende la selección de un método multiplicativo o uno de adición.
Las técnicas de pronóstico para datos estacionales se usan siempre que:

- *El clima influye en la variable de interés.* Como ejemplos están el consumo de energía eléctrica, las actividades de verano e invierno(deportes como el patinaje), el guardarropa y las estaciones de desarrollo agrícola.

- *El año calendario influye en la variable de interés.* Ejemplos de ello son las ventas al menudeo influidas por días festivos, fines de semana de tres días y los calendarios escolares.

Las técnicas a considerar al pronosticar series estacionales son descomposición clásica, Census II, atenuación exponencial de Winter, regresión múltiple de series de tiempo y métodos de Box-Jenkins.

Técnicas de pronóstico para series cíclicas

El efecto *cíclico* se definió anteriormente como la fluctuación en forma de onda alrededor de la tendencia. Los patrones cíclicos tienden a repetirse en los datos cada dos, tres o más años. Es difícil establecer un modelo para estos patrones cíclicos, ya que no son estables. Las fluctuaciones en forma de onda hacia arriba y hacia abajo alrededor de la tendencia rara vez se repiten en intervalos fijos de tiempo y también varía la magnitud de la fluctuaciones. Las técnicas de pronóstico para datos cíclicos se utilizan siempre que:

- *El ciclo del negocio influye sobre la variable de interés.* Como ejemplos están los factores económicos del mercado y de la competencia.
- *Se presentan cambios en el gusto popular.* Ejemplos de ello son la moda, la música y la alimentación.
- *Se presentan cambios en la población.* Podemos citar como ejemplos las guerras, escasez, epidemias y desastres naturales.
- *Se presentan cambios en el ciclo de vida del producto.* Ejemplos de ello son la introducción, crecimiento, maduración, saturación y declinación del mercado.

Las técnicas a considerar al pronosticar series cíclicas son la descomposición clásica, los indicadores económicos, los modelos econométricos, la regresión múltiple y los métodos de Box-Jenkins.

Otros factores a considerar

El horizonte en el tiempo para un pronóstico tiene una relación directa con la selección de una técnica de pronóstico. Para los pronósticos de corto (máximo un año), mediano (de dos a cuatro años) y largo (de cuatro en adelante) plazo, se pueden aplicar diversas técnicas cuantitativas. Sin embargo, al aumentar el horizonte del pronóstico, algunas de estas técnicas se hacen menos aplicables. Por ejemplo, las medias móviles, la atenuación exponencial y los modelos de Box-Jenkins no son muy buenos para pronósticos de cambios económicos radicales, mientras que los modelos econométricos son mejores para este fin. Los modelos de regresión son apropiados para períodos corto, mediano y largo. Las proyecciones de medias, medias móviles, descomposición clásica, son técnicas cuantitativas apropiadas para horizontes de corto y mediano plazo. Las técnicas más complejas de Box-Jenkins y los modelos econométricos resultan también apropiados para pronósticos de corto y mediano plazo. Para horizontes mayores de tiempo, se usan con frecuencia los métodos cualitativos [John E. Hanke y Arthur G. Reitsch, 1996].
En general, la aplicabilidad de las técnicas de pronóstico es algo que el pronosticador realiza con base en su experiencia. Es común que los administradores requieran de pronósticos en un tiempo relativamente corto. En esta situación, tienen ventaja los métodos de atenuación exponencial, proyección de tendencia, modelos de regresión y la descomposición clásica.

2.2. Modelos de tendencia

Cuando la serie a pronosticar varía con el paso del tiempo, esa variación se atribuye a componentes subyacentes, como tendencias, estacionalidad y ciclos. En esta sección describiremos la *tendencia*[1]. La existencia de una tendencia es obvia y su comportamiento muchas veces se asemeja a funciones conocidas, las cuales pueden ser lineales o no lineales. A continuación veremos una descripción de varios tipos de tendencia.

[1]Después estudiaremos series que contengan no solo la tendencia sino estacionalidad.

Tendencia lineal

Una tendencia lineal se representa como:

$$Y_t = \beta_0 + \beta_1 t + e_t \tag{2.1}$$

Es decir, el valor de la serie de tiempo es igual a una constante (llamada ordenada al origen) más una pendiente multiplicada por el valor del tiempo, el cual se incrementa sucesivamente. Así, si la pendiente es negativa, la serie de tiempo será decreciente, mientras que, si la pendiente es positiva, la tendencia será creciente. Es fácil ver que si el valor de β_1 es cero, se trata de una serie de tiempo sin tendencia, paralela al eje del tiempo, este modelo es adecuado cuando los datos se asemejan a una recta.

Tendencia cuadrática

Por otro lado, la tendencia puede ser *no lineal*, de diversas formas: cuadrática, logarítmica, polinómica, exponencial, etc. La más común es la tendencia cuadrática, en la cual los crecimientos o decrecimientos son mucho más significativos. Este modelo se usa cuando los datos se asemejan a una parabola. Esta tendencia se modelaría como sigue:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + e_t \tag{2.2}$$

Tendencia exponencial

Una tendencia exponencial existe cuando el crecimiento(o decrecimiento) es todavía más rápido; los datos se asemejan a la figura de la función exponencial cuya ecuación se escribiría como:

$$Y_t = \beta_0 e_t^{\beta_1 * t} \tag{2.3}$$

La tendencia es una función no lineal (exponencial) de tiempo en escala lineal, pero en logaritmos se tiene:

$$\ln(Y_t) = \ln(\beta_0) + \beta_1 t + e_t \tag{2.4}$$

Así, $\ln(Y_t)$ es función lineal del tiempo. En el caso en el cual la tendencia es no lineal en escala lineal, pero lineal en logaritmos, se llama *tendencia exponencial*, o *tendencia lineal logarítmica*.

A continuación profundizaremos en el caso lineal, no sólo por ser el más simple, sino porque permiten dar respuesta a un gran número de problemas. Además, algunos de los modelos no lineales pueden, mediante adecuadas transformaciones, ser expresados en forma lineal.

2.2.1. Modelos de regresión lineal

Antes de continuar, haremos mención de una definición y algunos teoremas importantes[2] para un mejor entendimiento.

Definición 2.1. Un estimador $\hat{\theta}$ es un estimador insesgado para el parámetro θ si, para cualquier tamaño muestral, su esperanza es igual al parámetro que estima. Esto es $E(\hat{\theta}) = \theta$, para todo valor de θ. El sesgo del estimador es definido como: $Sesgo(\hat{\theta})=E(\hat{\theta} - \theta)$.

Teorema 2.1. *Si se cumplen los supuestos del modelo básico de regresión lineal, entonces $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores insesgados de β_0 y β_1, respectivamente.*

Teorema 2.2. *Si se cumplen los supuestos del modelo básico de regresión lineal, entonces los estimadores mínimo cuadrados de β_0 y β_1 son los mejores estimadores lineales insesgados de estos parámetros.*

El objeto de un análisis de regresión es investigar la relación estadística que existe entre una variable dependiente(Y) o también conocida como variable endógena y una o más variables independientes o exógenas. La forma funcional que más se utiliza en la práctica es la relación lineal, cuando solo existe una variable independiente. En el caso de series de tiempo nuestra variable independiente sera el tiempo y la variable dependiente seguirá siendo Y[3]. El modelo básico de la regresión lineal viene dado por:

[2]Para ver la demostración de los teoremas consultar Calero Vinelo Arístides, 1998 [5]

[3]Cabe señalar que manejaremos la variable independiente como X_t en lugar de t para mostrar un caso mas general y no limitarnos solo al tiempo

$$Y_t = \beta_0 + \beta_1 X_t + e_t \tag{2.5}$$

donde los coeficientes β_0 y β_1 son parámetros[4] de la regresión y son desconocidos, que definen la posición e inclinación de la recta. El parámetro β_0, conocido como la ordenada en el origen, nos indica el valor de Y cuando $X = 0$. El parámetro β_1, conocido como la "pendiente", nos indica cuánto aumenta Y por cada aumento de una unidad en X. Nuestro problema consiste en obtener estimaciones de estos coeficientes a partir de una muestra de observaciones sobre las variables Y y X. Pero antes de aventurarnos en los métodos de estimación para estos parámetros, pondremos los supuestos que debe cumplir este modelo.

- *Supuesto1.* $E[e_t] = 0$ La media de la distribución de probabilidad de e es 0, i.e. la media de los errores a lo largo de una serie infinitamente larga de experimentos es 0 para cada valor de la variable independiente X. Este supuesto implica que el valor medio de Y, E[Y], para un valor dado de X es $E[Y] = \beta_0 + \beta_1 X_t$

- *Supuesto2.* La varianza de la distribución de probabilidad de e es constante para todos los valores de la variable independiente X. En el caso de nuestro modelo de línea recta, este supuesto significa que la varianza de e es igual a una constante, digamos σ^2, y garantiza que hay homoscedasticidad (igual variabilidad) entre las perturbaciones aleatorias. $Var(e_t) = E[e_t - E(e_t)]^2 = E[e_t] = \sigma^2$.

- *Supuesto3.* La distribución de probabilidad de e es normal con media cero y varianza σ^2, es decir, $e \sim N(0, \sigma^2)$.

- *Supuesto4.* Los errores asociados a cualesquiera dos observaciones distintas son independientes. Este supuesto indica que no hay autocorrelación entre las perturbaciones e_t y e_j; es decir, que están incorrelacionadas. $Cov(e_t, e_j) = E(e_t - E(e_t))(e_j - E(e_j)) = E(e_t, e_j) = 0; t \neq j$.

[4]Se utilizará el punto de vista de la teoría del muestreo, que considera a un parámetro como una cantidad fija pero desconocida

Una vez aceptados estos supuestos[5], pasamos a los métodos para estimar los parámetros β_0 y β_1 para el modelo de regresión(2.5) los cuales son:

- El método de momentos.
- El método de mínimos cuadrados.
- El método de máxima verosimilitud.

En el caso del modelo de regresión simple, los tres métodos proporcionan estimaciones idénticas. Cuando se trata de generalizaciones, proporcionan estimaciones diferentes. Pero nosotros solo utilizaremos el método de mínimos cuadrados debido a su sencilla interpretación y calculo (para ver los dos métodos faltantes consultar Maddala G.S. 1996 [23]). Antes de entrar a el método de mínimos cuadrados veremos los tipos de error.

Tipos de error

Se han diseñado diversos métodos para resumir los errores generados por una técnica particular de pronóstico. La mayoría de estas mediciones implican promediar alguna función de la diferencia entre el valor real y su valor de pronóstico. A menudo se denominan *residuales* a estas diferencias entre valores observados y los valores de pronóstico.
EL error o residual para cada periodo

$$e_t = Y_t - \hat{Y}_t \tag{2.6}$$

Un método para la evaluación de un pronóstico consiste en obtener la suma de los errores absolutos. La desviación absoluta de la media (DAM) mide la precisión de un pronóstico mediante el promedio de la magnitud de los errores de pronóstico (valores absolutos de cada error). La DAM resulta de gran utilidad cuando el analista desea medir el error de pronóstico en las mismas unidades de la serie original.

$$DAM = \frac{\sum_{t=1}^{n} |Y_t - \hat{Y}_t|}{n} \tag{2.7}$$

[5]Si se desea ver las técnicas para regresión lineal cuando uno de estos supuestos no se cumple, consultar Maddala G.S. 1996 [23]

Otro método para evaluar una técnica de pronóstico es el Error Medio Cuadrado (MSE). Cada error o residual se eleva al cuadrado; luego, estos valores se suman y se divide entre el numero de observaciones. Este enfoque penaliza los errores mayores de pronóstico ya que eleva cada uno al cuadrado.

$$MSE = \frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{n} \tag{2.8}$$

En ocasiones, resulta mas útil calcular los errores de pronóstico en términos de porcentaje y no en cantidades. El porcentaje de error medio absoluto (PEMA) se calcula encontrando el error absoluto en cada periodo, dividiendo éste entre el valor real observado para ese periodo y después promediando estos errores absolutos de porcentaje. El PEMA proporciona una indicación de qué tan grandes son los errores de pronóstico comparados con los valores reales de la serie.

$$PEMA = \frac{\sum_{t=1}^{n}\frac{|Y_t - \hat{Y}_t|}{Y_t}}{n} * 100 \tag{2.9}$$

A veces resulta necesario determinar si un método de pronóstico está sesgado. En estos casos, se emplea el porcentaje medio de error (PME), que se calcula encontrando el error en cada periodo, dividiendo esto entre el valor real de ese periodo y promediando después estos porcentajes de error. Si un enfoque no esta sesgado la ecuación 2.10 producirá un porcentaje cercano a cero.

$$PME = \frac{\sum_{t=1}^{n}\frac{(Y_t - \hat{Y}_t)}{Y_t}}{n} * 100 \tag{2.10}$$

Una parte importante en la decisión o confiabilidad de un método de pronóstico consiste en que los errores producidos por el modelo se juzguen como suficientemente pequeños, de ahí la importancia de medir los errores.

Método de mínimos cuadrados

Este método consiste en elegir aquellos estimadores que hacen mínima la suma de las diferencias cuadráticas entre los valores observados y los valores estimados de la variable dependiente, es decir, que minimizan la suma de los

errores al cuadrado (EC).

$$EC = \sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{n} e_t^2 \tag{2.11}$$

Aplicando en el modelo (2.5) el método de los mínimos cuadrados se obtienen estimadores lineales insesgados[6] y óptimos. En este texto solo nos limitaremos a poner las ecuaciones resultantes para poder estimar los parámetros óptimos[7].

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t Y_t - n\bar{X}\bar{Y}}{\sum_{t=1}^{n} X_t^2 - n\bar{X}^2} \tag{2.12}$$

y

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{2.13}$$

Donde

- X_t= Es la variable independiente en el tiempo t
- Y_t= Es la variable dependiente en el tiempo t
- n= Es el numero de observaciones de la muestra hasta el tiempo t
- $\bar{X}$= Es la media muestral calculada como $\frac{\sum_{t=1}^{n} X_t}{n}$
- $\bar{Y}$= Es la media muestral calculada como $\frac{\sum_{t=1}^{n} Y_t}{n}$

[6]Lo que quiere decir es que si se calculan los parámetros para cada muestra y se repite este proceso muchas veces hasta el infinito, el promedio de todos estos estimados será igual a β, en otras palabras $E(\hat{\beta}) = \beta$.

[7]para ver el desarrollo de las formulas consultar Calero Vinelo Arístides, 1998 [5]

Coeficiente de correlación

Se necesita una forma de medir la cantidad de relación lineal que existe entre dos variable de interés. Para usar la terminología correcta, se desea una medición de la *correlación* que existe entre dos variables. La medición que se utiliza comúnmente para esta relación es el *coeficiente de correlación*. Dos variables con una relación negativa perfecta tienen un coeficiente de correlación igual a -1. En el otro extremo, dos variables con una relación positiva perfecta tienen un coeficiente de correlación igual a +1. Por lo que el coeficiente de correlación varía -1 y +1. Un valor positivo de r implica que Y aumenta cuando X aumenta; un valor negativo implica que Y disminuye cuando X aumenta. Cuando el coeficiente de correlación es 0, no existe relación lineal, es decir, al aumentar X, Y no parece aumentar o disminuir en forma predecible alguna. Pero hay que tener un par de consideraciones como señalan Hanke John E. y G.Reitsch Arthur [8] "Si fuera el caso en que el coeficiente de correlación fuese bajo no significa que no hay correlación entre las variables, podría ser que si tuvieran una estrecha relación de manera curva o no lineal, solamente que no es aparente que exista una relación lineal o recta entre las variables. Segundo, se debe de tener en cuenta que se esta midiendo la *correlación* y no la *causalidad*. Podría ser perfectamente válido que dos variables estuvieran correlacionadas con base en un coeficiente de correlación alto. Pero podría ser o no válido decir que una variable *causa* el movimiento de la otra; ésta es una cuestión para el juicio del analista". Entonces, la fórmula para obtener el coeficiente de correlación viene dado como:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}_i\bar{Y}_i}{\sqrt{(\sum_{i=1}^{n} X_i^2 - n\overline{X^2})}\sqrt{(\sum_{i=1}^{n} Y_i^2 - n\overline{Y^2})}} \qquad (2.14)$$

Coeficiente de determinación

Otra forma de medir la contribución de X a la predicción de Y o, en otras palabras, saber cuanta variabilidad de la variable dependiente es explicada a través de la variable independiente se conoce como coeficiente de determi-

[8] véase Hanke E. John y Arthur G. Reitsch [12]

nación. Este varia entre 0 y 1, cuanto más próximo esté a 1, mayor valor predictivo tendrá el modelo en el sentido que los valores observables estarán muy próximos a la esperanza estimada por la regresión. A este coeficiente lo llamaremos como r^2 y se calcula de la manera siguiente:

$$r^2 = \frac{\beta_0 \sum_{i=1}^{n} Y_i + \beta_1 \sum_{i=1}^{n} X_i Y_i - n\bar{Y}^2}{\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2} \tag{2.15}$$

La ventaja del coeficiente de determinación (r^2) consiste en que tiene una interpretación muy útil. El valor de r^2 mide el porcentaje de variabilidad en Y que se explica por la variabilidad en X. Esta interpretación es muy útil, lo que hace de r^2 una de las estadísticas consultadas con mayor frecuencia en el análisis de regresión.

Asimismo, si r^2 es el cuadrado de r, el coeficiente de correlación, entonces nos vendría a la mente ¿por qué es importante identificar ambos valores en un análisis de regresión?, y la razón es que uno tiene ventaja sobre otro. La ventaja del coeficiente de correlación (r) es que revela relaciones tanto positivas como negativas. En cambio, nótese que cuando se eleva al cuadrado el coeficiente de correlación, el valor es siempre positivo, perdiéndose la naturaleza de la relación.

2.3. Modelos sencillos de pronostico

2.3.1. Métodos no formales o ingenuos

Los métodos *no formales* se emplean para desarrollar modelos sencillos que suponen que los periodos recientes son los mejores pronosticadores del futuro. El modelo más sencillo es:

$$\hat{Y}_{t+1} = Y_t \tag{2.16}$$

En donde $\hat{Y}_{t+1}$ es el pronóstico realizado en el periodo t para el periodo $t+1$. y Y_t es el valor real de la serie en el tiempo t.

La técnica puede adaptarse para tomar en cuenta la tendencia agregando

la diferencia entre éste y el último periodo. La predicción que obtendríamos estaría suponiendo que el incremento previsto para la variable coincide con el último registrado.

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \tag{2.17}$$

Si la serie tuviera estacionalidad tendríamos la siguiente ecuación

$$\hat{Y}_{t+1} = Y_{t-p-1} \tag{2.18}$$

En donde p seria el periodo en que esta dividido el año, por ejemplo 4 en el caso trimestral. Esta predicción supone que la variable tomará el mismo valor que tenía en el periodo correspondiente al año anterior. El problema de esta ecuación es cuando exista tendencia dado que se considera. Pero podríamos agregar información reciente, un buen ejemplo podría ser:

$$\hat{Y}_{t+1} = Y_{t-p-1} + \frac{(Y_t - Y_{t-1}) + ... + (Y_{t-p-1} - Y_{t-p})}{p} \tag{2.19}$$

Con esto vemos las diferentes combinaciones de métodos no formales que se pueden formar, la única limitante dependerá del ingenio del pronosticador.

2.3.2. Medias Móviles

También llamados como promedios móviles, pero para evitar confusiones con los promedios móviles autorregresivos (que veremos más adelante), en este texto adoptaremos el nombre de medias móviles para evitar dicha confusión. Las medias móviles son un promedio de un número constante de observaciones, este promedio esta basado en un mismo número de observaciones en un lapso de tiempo que mueve sus datos un periodo de tiempo, del inicio del promedio al último ó más reciente dato de la serie. El numero de observaciones usados para el calculo del promedio es llamado el orden de la serie, es decir, si nuestra media móvil es trimestral, nosotros dividiríamos entre 4 y nuestra media móvil serie de orden 4. La siguiente ecuación corresponde a la ecuación de medias móviles simple de orden n.

$$M_t = \hat{Y}_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + ... + Y_{t-n-1})}{n} \tag{2.20}$$

- M_t= Media móvil en el periodo t.
- n= Número de términos que se desean en la media móvil.

Nótese que se asignan ponderaciones(pesos) iguales a cada observación. Al tener el dato mas reciente, este se incluye en el promedio y se desecha el dato más antiguo. El modelo de medias móviles funciona mejor con datos estacionarios. No maneja muy bien la tendencia o la estacionalidad.

Cabe señalar que un valor de n grande, proporciona una serie más alisada que un valor pequeño. Si las variaciones de la serie se deben a la influencia del elemento aleatorio, un n grande sería más conveniente. Por el contrario, si es el nivel medio la causa principal de variación en la serie deberíamos recurrir a un valor de n pequeño, precisamente para no sobrevalorar el cambio en el nivel medio, ya que, en este caso, las predicciones son más sensibles a los valores recientes de la serie.

Medias móviles dobles

Una media móvil doble puede ser utilizada para un alisamiento adicional, dado que uno de los problemas que tienen las medias móviles simples, es que no trabajan bien cuando existe tendencia, en este caso, es mas factible utilizar las medias móviles dobles para una serie con tendencia lineal. El método hace lo que nos dice su nombre: se calcula un conjunto de medias móviles y después se calcula un segundo conjunto de medias móviles del primero. Para ver esto partimos de la ecuación de medias móviles simple:

$$\hat{Y}_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + ... + Y_{t-n+1})}{n} \tag{2.21}$$

Recordando que $M_t = \hat{Y}_{t+1}$ entonces para calcular la segunda media móvil aplicamos nuevamente la formula pero ahora con los pronósticos obtenidos de las medias móviles.

$$M'_t = \frac{(M_t + M_{t-1} + M_{t-2} + ... + M_{t-n+1})}{n} \tag{2.22}$$

- M'_t= Es la segunda media móvil en el tiempo t.

La siguiente ecuación (2.23) calcula la diferencia entre ambas medias móviles.

$$a_t = 2M_t - M'_t \tag{2.23}$$

La ecuación 2.24 es un ajuste adicional, el cual llamaremos coeficiente de la tendencia, es similar a la medición de una pendiente que cambia a través de la serie. En otras palabras, es el promedio de la diferencia entre la media móvil simple y la media móvil doble de un punto a otro.

$$b_t = \frac{2}{n-1}(M_t - M'_t) \tag{2.24}$$

Finalmente, para poder pronosticar se utiliza la siguiente ecuación, para realizar el pronóstico de p periodos en el futuro.

$$\hat{Y}_{t+p} = a_t + b_t \cdot p \tag{2.25}$$

2.4. Métodos de atenuación (alisamiento)

Se llaman métodos de alisamiento o atenuación, porque su objetivo es el disminuir o alisar las fluctuaciones de la serie de tiempo. Son métodos que se retroalimentan a medida que se generan nuevos datos, lo que facilita enormemente la predicción que se realiza, como se verá, con fórmulas recurrentes. Las técnicas de atenuación presentan ciertas peculiaridades respecto a éstos que podemos resumir en los siguientes puntos:

- Cuando la estructura de los datos se muestra inestable, los métodos de atenuación dan mejores resultados, ya que, en general, tratan de combinar en cada momento las observaciones pasadas con el fin de descubrir la estructura del fenómeno. En este sentido, se dice que son modelos de validez local.
- Las técnicas de atenuación son más indicadas cuando queremos predecir, sobre todo a corto plazo, ya que los cambios estructurales se detectan antes.

- Los resultados que se obtienen con ellas son satisfactorios, incluso cuando se dispone de un número limitado de observaciones.

Al describir los distintos métodos de atenuación comenzaremos con los modelos sencillos seguido de una clasificación de los mismos según las componentes que estén presentes en las series de análisis. De esta forma estudiaremos en primer lugar los modelos para series donde las fluctuaciones oscilan alrededor de una constante y qué, por tanto, no están afectados por tendencia ni estacionalidad, seguiremos con series con tendencia y finalizaremos con modelos completos aplicados a series con tendencia y estacionalidad.

2.4.1. Atenuación exponencial simple

La principal limitación que tiene el procedimiento de medias móviles como técnica de predicción, es que pondera de la misma manera los valores que integran la media, i.e, les da la misma importancia. Pero quizás, cuando se trata de predecir la información más cercana al momento, en el cual se realiza la predicción, puede resultar más relevante las observaciones próximas al momento, que las observaciones más alejadas. Para poder entender mejor esté método partiremos de la atenuación exponencial simple para luego proceder con la atenuación exponencial doble o método de Brown.

La atenuación exponencial es un Método, concebido por Robert Macaulay en 1931 y desarrollado por Robert G. Brown durante la segunda guerra mundial. Este método trata de predecir con base en una media ponderada de los valores pasados de la serie, dando ponderaciones decrecientes conforme los datos se alejan del momento actual. Las diferencia por tanto entre la atenuación exponencial y las medias móviles son:

- Se trabajan con medias ponderadas
- La media se aplica, no solo a n observaciones, caso de la media móvil, sino a toda la información muestral hasta el momento t.

La ecuación de atenuación exponencial simple es:

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \tag{2.26}$$

- $\hat{Y}_{t+1}$= Nuevo valor atenuado o pronóstico para el periodo $t+1$.
- α= Constante de atenuación $(0 < \alpha < 1)$ la cual determina cuanto peso sera dado a cada observación.

Frecuentemente nuestro pronostico inicial se toma como: $\hat{Y}_0 = Y_0$ ó puede ser un promedio de varias observaciones. Una constante de atenuación α pequeña da mayor peso a las observaciones más antiguas. Una constante de atenuación grande da mayor peso a las observaciones más recientes y menor a las pasadas. Una forma de saber cual es la mejor α, es la que minimiza el error cuadrático medio(2.8), pero si se utiliza un paquete estadístico el programa nos dará el α óptimo [Hanke John E. y Arthur G. Reitsch, 1996 [12]].

Nótese que en la técnica de atenuación exponencial simple existe la suposición de que los datos son estacionarios. Si existe una tendencia significativa, la atenuación exponencial se ubicará por debajo de los datos reales de la serie de tiempo.

Los pronósticos de medias móviles simples y atenuación exponencial se basan en promedios ponderados de mediciones anteriores. La explicación consiste en que los valores anteriores contienen información de lo que ocurrirá en el futuro. Debido a que los valores anteriores incluyen fluctuaciones aleatorias, así como información relativa al patrón subyacente de una variable, se hace un intento de atenuar estos valores. La atenuación exponencial es una técnica popular para los pronósticos de corto plazo, sus mejores ventajas son un bajo costo y simplicidad.

2.4.2. Método de Brown

También conocido como atenuación exponencial doble. Como su nombre nos indica, primero se atenúan los valores reales y después se atenúa los valores ya atenuados, para después con ayuda de un coeficiente de tendencia poder pronosticar. Esté método se utiliza para poder pronosticar series de tiempo con tendencia lineal.

Sea A_t definida como sigue:

$$A_t = \alpha Y_t + (1-\alpha)A_{t-1} \tag{2.27}$$

La ecuación 2.27 nos representa el valor simple atenuado (A_t). Seguido por:

$$A'_t = \alpha A_t + (1-\alpha)A'_{t-1} \tag{2.28}$$

Esta ecuación (2.28) se trata del valor doblemente atenuado exponencialmente (A'_t). Frecuentemente se utiliza A_0 como valor inicial de A'_0 o se podría hacer un promedio de varios valores atenuados. Una vez tenido esto seguimos como en el caso de las medias móviles dobles, la diferencia entre los dos valores atenuados:

$$a_t = 2A_t - A'_t \tag{2.29}$$

Y también tenemos el coeficiente de la tendencia:

$$b_t = \frac{\alpha}{1-\alpha}(A_t - A'_t) \tag{2.30}$$

Para finalmente llegar a la ecuación 2.31 para poder formular el pronóstico de p periodos en el futuro.

$$\hat{Y}_{t+p} = a_t + b_t \cdot p \tag{2.31}$$

Al igual que en el caso de la atenuación exponencial simple la α que optimiza el método, es el que minimiza los errores medios cuadráticos y además también en el caso de A_0 se podría promediar algunos datos o tomarse Y_0 como valor inicial.

2.4.3. Método de Holt

Otra técnica que se usa para manejar series de tiempo con tendencia lineal, es el *método de dos parámetros de Holt*. El método consiste en dos ecuaciones para las dos componentes de atenuación que son A_t y T_t además de dos constantes de atenuación, el hecho de utilizar dos constantes de atenuación hace que este método se adapte con mayor flexibilidad a los valores de la serie que el modelo de Brown. Las ecuaciones del método son las siguientes:

$$A_t = \alpha Y_t + (1-\alpha)(A_{t-1} + T_{t-1}) \tag{2.32}$$

La ecuación 2.32 nos representa un tipo de serie exponencialmente atenuada, pero esta vez se considera la tendencia, la cual es calculada con la siguiente ecuación(2.33) y la llamaremos como estimación de la tendencia, está ecuación es muy similar a la anterior pero en vez de atenuar los valores reales atenúa la tendencia.

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \tag{2.33}$$

De estas ecuaciones podemos observar que, al igual que en el caso de la atenuación exponencial simple, son fórmulas de actualización mediante las cuales las estimaciones son modificadas a medida que se dispone de nuevas observaciones. Al mismo tiempo, los valores de la tendencia se actualizarían a través de una media ponderada entre el último valor estimado, T_t, y la diferencia entre las dos estimaciones más recientes de los valores atenuados $(A_t - A_{t-1})$. Finalmente, tenemos la ecuación 2.34 que nos permite pronosticar p periodos en el futuro.

$$\ddot{Y}_{t+p} = A_t + p \cdot T_t \tag{2.34}$$

en donde:

- A_t= Nuevo valor atenuado.
- α= Constante de atenuación de los datos $(0 < \alpha < 1)$
- Y_t=Nueva observación o valor real de la serie, en el periodo t.
- β= Constante de atenuación de estimación de la tendencia $(0 < \alpha < 1)$.
- T_t= estimación de la tendencia.

Cabe mencionar que el valor inicial atenuado (A_0) se calcula como se mencionó en el método de Brown, ya sea con un promedio de varios valores reales o como su valor real que le corresponda.En el caso del valor inicial de la estimación de la tendencia frecuentemente se utiliza el valor de cero.

2.4.4. Método de Winter

Los métodos anteriores son buenos cuando solo existe tendencia lineal pero cuando tenemos tendencia y estacionalidad se opta por utilizar el método de Winter a menos que se elimine la estacionalidad y se trabaje con la serie que contiene solo tendencia. Las ecuaciones que emplea el modelo Winter son:

La ecuación 2.35 nos representa a la serie exponencialmente atenuada. Donde S_{t-L} representa el factor estacional pero de un año anterior, está ecuación es muy similar a la utilizada en el método de Holt, sólo que en este caso la actualización se hace ante la llegada de un nuevo dato (Y_t), debemos considerar la observación sin influencia de la estacionalidad. Por tal motivo aparece el valor de la serie desestacionalizado por el factor estacional correspondiente a un periodo inmediatamente anterior S_{t-L}.

$$A_t = \alpha \frac{Y_t}{S_{t-L}} + (1-\alpha)(A_{t-1} + T_{t-1}) \tag{2.35}$$

La estimación de la tendencia. Como podemos ver es exactamente igual que en el método de Holt, con lo que no se necesita ningún tipo de aclaración.

$$T_t = \beta(A_t - A_{t-1}) + (1-\beta)T_{t-1} \tag{2.36}$$

La siguiente ecuación (2.37) nos muestra la estimación de la estacionalidad. Este componente podemos incluirlo de forma multiplicativa o bien de forma aditiva, pero la más usual es la forma multiplicativa[9] y es está la que manejaremos, quedando entonces como:

$$S_t = \gamma \frac{Y_t}{A_t} + (1-\gamma)S_{t-L} \tag{2.37}$$

Las constantes de atenuación óptimas son las que minimizan el MSE, estos parámetros los calculan paquetes estadísticos. Finalmente para poder hacer predicciones tenemos la ecuación (2.38), la cual nos permite pronosticar p periodos en el futuro y además, podemos ver que el factor estacional multiplica (de allí su nombre) a la diferencia del valor atenuado y la tendencia:

$$\hat{Y}_{t+p} = (A_t - pT_t)S_{t-L+p} \tag{2.38}$$

[9]Para mayor referencia sobre el método aditivo ver Bowerman Connell, 1993 [4]

- A_t= Nuevo valor atenuado
- α= constante de atenuación $(0 < \alpha < 1)$
- Y_t= Nueva observación o valor real de la serie en el periodo p.
- β= Constante de atenuación de la tendencia $(0 < \beta < 1)$.
- T_t= Estimación de la tendencia.
- γ= Constante de atenuación de la estacionalidad $(0 < \gamma < 1)$.
- S_t= Factor estacional.
- p= Periodos a estimar a futuro.
- L= Longitud de la estacionalidad, i.e, como esta dividido el año(trimestres, meses, semestres, etc).

Cabe señalar, que si solo tuviéramos estacionalidad en nuestra serie, bien se podría hacer cero el factor de la tendencia y trabajar el resto del modelo de igual forma, la ecuación quedaría como sigue:

$$\hat{Y}_{t+p} = (A_t)S_{t-L+p} \tag{2.39}$$

Por último, debe señalarse que para el valor inicial del valor atenuado (A_0) se puede tratar igual que en los métodos anteriores, o bien se puede utilizar la siguiente formula sí uno tiene muchos datos históricos:

$$A_0 = \bar{Y}_1 - \frac{L}{2} \cdot T_0 \tag{2.40}$$

Mientras que en el caso de la tendencia puede iniciarse con cero o sí se dispone de muchos datos puede utilizar la siguiente formula:

$$T_0 = \frac{Y_m - Y_1}{(m-1) \cdot L} \tag{2.41}$$

Donde $\bar{Y}_m$ nos representa un promedio de los datos hasta el año m (último año disponible) y $\bar{Y}_1$ nos representa el promedio de los valores reales del

primer año. Por último, en el caso el valor inicial del factor estacional sería de la siguiente manera:

$$S_0 = \frac{Y_0}{A_0} \tag{2.42}$$

O una manera menos formal sería darle el valor de 1. Pero es recomendable utilizar la ecuación 2.42.

2.5. Modelos Estadísticos para series de tiempo

Debido a que las series de tiempo constan de datos numericos, es natural usar la herramienta de la estadística para describirlas y analizarlas, así como ocurre con cualquier otro conjunto de información numérica. Recordemos que la estadística emplea dos enfoques básicos: 1) el enfoque descriptivo, que se ocupa esencialmente de resumir y describir en forma concisa, ya sea mediante gráficas o a través de unas cuantas medidas descriptivas y 2) el enfoque inferencial, cuyo objetivo fundamental es utilizar datos muestrales para realizar inferencias, que sean válidas para toda la población de donde se obtuvo la muestra. Una vez visto esto podemos aventurarnos a algunos conceptos importantes que se retomarán más adelante.

Procesos estocásticos

Para describir lo que es una serie de tiempo dentro del contexto de procesos estocásticos, necesariamente debemos definir los procesos estocásticos. Supongamos que tenemos una muestra de tamaño T de alguna variable alcatoria Y_t:

$$\{y_1, y_2, ..., y_T\} \tag{2.43}$$

Consideremos una colección de T variables independientes e identicamente distribuidas(i.i.d.) e_t:

$$\{e_1, e_2, ..., e_T\} \tag{2.44}$$

con

$$e_t \sim N(0, \sigma^2) \tag{2.45}$$

Esto se refiere como una muestra de tamaño T de un proceso Gaussiano de ruido blanco[10].

La muestra 2.43 representa observaciones particulares de T, pero este conjunto de T observaciones es solo una posible muestra de un proceso estocástico, si tuviéramos una muestra de tamaño infinito veríamos que se trata de una realización particular de un proceso de series de tiempo. En pocas palabras un proceso estocástico es una familia de variables aleatorias asociadas a un conjunto índice de números reales, de tal manera que a cada elemento del conjunto le corresponda una y solo una variable aleatoria.

Ruido blanco

La construcción básica de todos los procesos que consideraremos aquí será una sucesión $\{e_t\}_{-\infty}^{\infty}$ cuyos elementos tienen media cero y varianza σ^2,

$$\begin{aligned} E[e_t] &= 0 \\ E[e_t^2] &= \sigma^2 \end{aligned} \tag{2.46}$$

y para los cuales las e_t´s están no correlacionadas a través del tiempo, es decir:

$$E[e_t, e_s] = \begin{cases} \sigma^2 & para \;\; t = s \\ 0 & para \; t \neq s \end{cases} \tag{2.47}$$

y la función de autocorrelación vendría dada como:

$$\rho(\kappa) = \begin{cases} 1 & para \;\; \kappa = 0 \\ 0 & para \; \kappa \geq 0 \end{cases} \tag{2.48}$$

Entonces, un proceso que satisface todas las condiciones expuestas es descrito como un *proceso de ruido blanco.*

A continuación veremos un operador muy útil para la simplificación de expresiones que utilizaremos posteriormente.

[10]Esto quiere decir que $E[e_t] = 0$ y $var[e_t] = \sigma^2$, pero esto lo veremos más adelante.

El operador de rezago

Este operador L es muy sencillo, "opera"sobre una serie retrasándola, es decir:

$$LY_t = Y_{t-1} \tag{2.49}$$

De igual forma,

$$L^2Y_t = L(L(Y_t)) = L(Y_{t-1}) = Y_{t-2} \tag{2.50}$$

Un polinomio de operador de rezago de grado m no es más que una función lineal de potencias de L hasta la m-ésima potencia.

$$B(L) = b_0 + b_1L + b_2L^2 + ... + b_mL^m \tag{2.51}$$

Un caso sencillo y general sería:

$$L^mY_t = Y_{t-m} \tag{2.52}$$

Es decir, al aplicar m-veces el operador L, se obtiene la variable retrasada m periodos y ya que $L^0 = 1$ entonces $L^0Y_t = Y_t$. Hay que tener en cuenta que el operador modifica a toda la sucesión de valores $\{Y_1, Y_2, Y_3, ..., Y_t, ..., Y_n\}$ para transformarla en la nueva sucesión $\{Y_{1-m}, Y_{2-m}, Y_{3-m}, ..., Y_{t-m}, ..., Y_{n-m}\}$. Es decir, si la serie originalmente contaba con n observaciones al aplicar el operador L^m, las observaciones $Y_{1-m}, Y_{2-m}, Y_{3-m}, ..., Y_{m-m}$ no se tendrán, quedando así una serie de n-m observaciones.

El siguiente operador de uso frecuente en el análisis de series de tiempo y que esta ligado con el retraso L es el de *diferencia* Δ. La característica principal de éste es que expresa las relaciones del tipo $Y_t - Y_{t-1}$, es decir, el cambio cuantitativo de la observación, entre un periodo y otro[11], entonces la siguiente igualdad se satisface:

$$\Delta Y_t = (1 - L)Y_t = Y_t - Y_{t-1} \tag{2.53}$$

Entonces $\Delta = 1 - L$

[11] En otras palabras es un polinomio de primer orden en el operador rezago

Generalizando mediante la aplicación sucesiva del operador diferencia y con ayuda del teorema del binomio obtenemos:

$$\Delta^m Y_t = (1-L)^m Y_t = \sum_{k=1}^{m} \frac{m!}{(m-k)!k!}(-1)^k Y_{t-k} \tag{2.54}$$

Donde el termino $L^k Y_t$ de (2.52) se convierte en Y_{t-k}.

Los polinomios de retraso utilizados para el análisis de series de tiempo, son una herramienta que muestra de manera clara y concisa el comportamiento de ciertos modelos que resultan ser de utilidad para representar fenómenos reales, dentro de ellos se encuentran los *promedios móviles* (MA), los *autorregresivos* (AR); así como las combinaciones de estos *autorregresivos de promedios móviles* (ARMA) y, por último, junto con la aplicación del operador diferencia, se tiene a los modelos *autorregresivos integrados de promedios móviles* (ARIMA).

2.5.1. Procesos Estacionarios

Una serie de tiempo es un conjunto ordenado $\{..., y_{-2}, y_{-1}, y_0, y_1, y_2, ...\}$. Casi siempre las observaciones están ordenadas en el tiempo, y de ahí el nombre de series de tiempo. En teoría una serie de tiempo inicia en el pasado infinito y se prolonga hasta el futuro infinito. Esto parece un poco abstracto y con poca aplicabilidad, pero es útil para deducir ciertas propiedades muy importantes de los modelos de pronóstico que más adelante usaremos. En la practica, los datos observados son sólo un conjunto finito de una serie de tiempo $\{y_1, ..., y_t\}$, lo que llamaremos muestra. Si la estructura probabilística básica de la serie cambiara a través del tiempo, sería muy malo porque no habría modo de predecir con exactitud el futuro a partir del pasado. Si queremos pronosticar una serie, lo mínimo es conocer su promedio y la estructura de su covarianza (esto es, las covarianzas entre los valores actuales y pasados) para que sea estable en el tiempo. En ese caso, se dice que la serie es de covarianza estacionaria (que solo la llamaremos como serie estacionaria). El primer requisito para una serie de este tipo será que su promedio sea estable con el tiempo. El promedio de la serie en el tiempo t es:

$$E[Y_t] = \mu \tag{2.55}$$

El segundo requisito para que una serie sea estacionaria es que la estructura de su covarianza sea estable en el tiempo. La cuantificación de la estructura de la covarianza es un poco laboriosa, pero su importancia es inmensa, y se hace recurriendo a la *función de autocovarianza*. La autocovarianza en el desplazamiento κ no es más que la covarianza entre Y_t y $Y_{t-\kappa}$. Naturalmente, que depende de κ y puede depender también de t, por lo que se escribe en general:

$$\gamma(t,\kappa) = Cov[Y_t, Y_{t+\kappa}] = E[(Y_t - \mu)(Y_{t+\kappa} - \mu)] \tag{2.56}$$

Si la estructura de la covarianza es estable en el tiempo, como se requiere, las autocovarianzas sólo dependen del desplazamiento κ y no del tiempo t, entonces la podemos reescribir como:

$$\gamma(t,\kappa) = \gamma_\kappa \tag{2.57}$$

para toda t.

2.5.2. Función de Autocovarianza

Otra implicación de la estacionariedad se deriva del hecho de que la autocovarianza entre dos observaciones cualesquiera depende sólo del número de periodos que separan dichas observaciones la cual definiremos como:

$$\begin{aligned}\gamma_\kappa &= Cov[Y_t, Y_{t+\kappa}] = E[(Y_t - E(Y_t))(Y_{t+\kappa} - E(Y_{t+\kappa}))] \\ &= E[(Y_t - \mu)(Y_{t+\kappa} - \mu)]\end{aligned} \tag{2.58}$$

Si analizamos el producto$(Y_t-\mu)(Y_{t+\kappa}-\mu)$ notaremos que si una observación se encuentra por encima de la media y κ periodos después es seguida por otra observación que también se encuentra por encima de la media, la autocovarianza entre Y_t y $Y_{t+\kappa}$ será positiva.

Sucede lo mismo si tenemos una observación por debajo de la media seguida κ periodos más tarde por otra observación que se encuentra también por debajo de la media.

Pero si una observación que se encuentra por arriba de la media, tiende a ser seguida κ periodos más tarde por una observación que está por debajo de la media, o viceversa, la autocovarianza entre Y_t y $Y_{t+\kappa}$ será negativa

El hecho de que la autocovarianza γ_κ parezca determinar la apariencia de una serie de tiempo, sugiere que un proceso estacionario mostrará el mismo modelo general de comportamiento sin importar cuando sea observado. Esto es, la realización $(Y_n, ..., Y_{n+j})$ no será exactamente la misma que $(Y_{n+j+h}, ..., Y_{n+2j+h})$ pero su apariencia general será la misma.

Puede entonces parecer apropiado caracterizar un proceso simplemente mostrando el conjunto de covarianzas $\gamma_0, \gamma_1, \gamma_2, ...$ Este conjunto será llamado función de autocovarianza.

Para propósitos de comparar series diferentes, sin embargo, no es muy satisfactorio ya que una diferencia en la dispersión de dos procesos, conduciría a autocovarianzas muy diferentes(causada quizá por escalas diferentes de medida). Por ejemplo, si una variable está medida en cientos o miles de pesos en lugar de millones de pesos, todos los segundos momentos estarán aumentados por un factor de 100, porque la varianza es una medida de dispersión. La comparabilidad puede llevarse a cabo si estandarizamos las autocovarianzas dividiéndolas por γ_0, esto es, transformándolas en correlaciones. Lo cual veremos en seguida.

A partir de este momento, siempre que nos refiramos a γ_κ lo haremos como la autocovarianza de periodo κ[12]

2.5.3. Función de autocorrelación y autocorrelación parcial

Cuando se mide una variable a través del tiempo, con frecuencia está correlacionada consigo misma cuando se desfasa uno o más períodos a esto lo llamaremos *Autocorrelación*. La estandarización de la función de autocovarianzas se denomina *función de autocorrelación* (FAC) y a tales correlaciones las denotaremos como ρ, entonces la correlación entre Y_t y $Y_{t+\kappa}$ será denotada

[12]Se usa el prefijo "auto" porque se trata de la covarianza de dos observaciones de la misma serie.

por $\rho(\kappa)$

$$\rho(\kappa) = \frac{E[(Y_t - \mu)(Y_{t+\kappa} - \mu)]}{\sqrt{E[(Y_t - \mu)^2]E[(Y_{t+\kappa} - \mu)^2]}} = \frac{E[(Y_t - \mu)(Y_{t+\kappa} - \mu)]}{\sigma^2} \tag{2.59}$$

ya que para un proceso estacionario $\sigma^2 = \gamma_0$ es la misma al tiempo $t+\kappa$ que al tiempo t. Entonces, la autocorrelación en el periodo κ es:

$$\rho(\kappa) = \frac{\gamma_\kappa}{\gamma_0} \tag{2.60}$$

lo que implicaría que

$$\rho(0) = 1 \tag{2.61}$$

$\rho(\kappa)$ es un número sin unidades o dimensiones, ya que la escala del numerador y del denominador son ambas el producto de las escalas en que se miden Y_t y $Y_{t+\kappa}$.

El conjunto de correlaciones a veces llamado función de autocorrelación estará dado por:

$$\rho(0) = \frac{\gamma_0}{\gamma_0} = 1 \quad \rho(1) = \frac{\gamma_1}{\gamma_0} \quad \ldots \quad \rho(\kappa) = \frac{\gamma_\kappa}{\gamma_0} \tag{2.62}$$

Una gráfica que muestra estas correlaciones a través de κ periodos,es decir una gráfica de la función de autocorrelación se llama correlograma (véase figura 2.6).

La función de autocorrelación además de no dimensional ($-1 \leq \rho(\kappa) \leq 1$) es simétrica, esto es:

$$\rho(\kappa) = \rho(-\kappa)$$

En general, cuando las observaciones en el tiempo t y $t+\kappa$ son similares en valor, $\rho(\kappa)$ tiene un valor cercano a 1. Cuando una observación grande en el tiempo t seguida por una pequeña observación en el tiempo $t+\kappa$, $\rho(\kappa)$ es cercana a -1. Si existen pequeñas relaciones entre las observaciones $\rho(\kappa)$ es aproximadamente 0.

Otra función que mide la correlación entre observaciones, es la ***función de autocorrelación parcial*** (FACP), que cuantifica la correlación entre Y_t y $Y_{t+\kappa}$,

una vez que las observaciones intermedias han sido removidas $Y_{t+1}, ..., Y_{t+\kappa-1}$. En otras palabras, las autocorrelaciones parciales se emplean para ayudar a identificar el grado de relación entre los valores de una variable y valores anteriores de la misma, mientras que se mantienen constantes los efectos de las otras variables (períodos retrasados). Lo anterior se logra con la siguiente expresión:

$$\varphi(\kappa) = \frac{Cov[(Y_t - \widehat{Y}_t), (Y_{t+\kappa} - \widehat{Y}_{t+\kappa})]}{\sqrt{Var(Y_t - \widehat{Y}_t)}\sqrt{Var(Y_{t+\kappa} - \widehat{Y}_{t+\kappa})}} \tag{2.63}$$

donde

$$\widehat{Y}_{t+\kappa} = E\{Y_{t+\kappa} Y_{t+1}, ..., Y_{t+\kappa-1} \tag{2.64}$$

es la esperanza condicional y estima la dependencia lineal de $Y_{t+\kappa}$ con respecto a $Y_{t+1}, ..., Y_{t+\kappa-1}$. Es de resaltar que estas funciones son generalmente estimadas con paquetería así que no hay que preocupase en como calcularse, basta con entender qué representa cada coeficiente. A continuación mostramos unos ejemplos (véase la figura 2.6) de los correlogramas tanto para una FAC como para la FACP, a simple vista no se ve la diferencia, está radica en el hecho de que el correlograma para la FAC sirve muchas veces para ver el grado del modelo MA y la FACP sirve para el modelo AR (pero esto se vera más adelante en la sección de identificación del modelo).

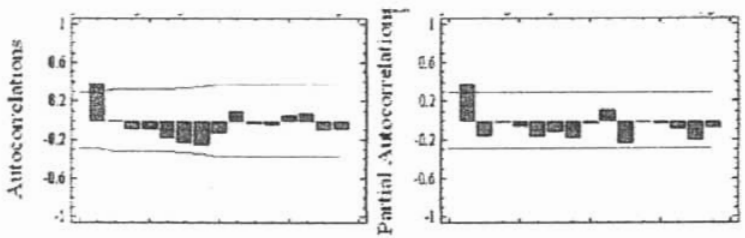


Figura 2.6: Ejemplos de correlogramas

2.5.4. Ecuaciones en diferencia

Las ***ecuaciones en diferencia*** son el equivalente discreto de las ecuaciones diferenciales que involucran variables en función del tiempo, es decir, dada una variable Z(t) y considerando que el tiempo es continuo, entonces el comportamiento de la variable queda determinado por sus derivadas

$$\frac{dZ}{dt}, \frac{d^2Z}{dt^2}, ..., \frac{d^kZ}{dt^k}$$

en cambio, si la misma variable es ahora observada en un tiempo discreto, entonces el comportamiento de Z(t) esta dictado por sus diferencias

$$\nabla Z_t, \nabla^2 Z_t, ..., \nabla^k Z_t, ...$$

La manera correcta de denotar este comportamiento discreto sería $\nabla Z_t / \nabla_t$, sin embargo, ya que t sólo toma valores enteros contiguos se sabe que $\nabla_t = 1$; por lo tanto, ∇Z_t es el equivalente de $\frac{dZ}{dt}$ cuando t toma los valores enteros ...,-2,-1,0,1,2,... La notación común para ecuaciones en diferencia es mediante el ***operador incremento*** definido por $\Delta Z_t = Z_{t+1} - Z_t$, donde se puede utilizar la relación $\nabla_t = \Delta_{t-1}$ que liga a los operadores incremento y diferencia, para conservar el concepto de variable retrasada utilizando simplemente ∇_t.

Es importante mencionar que existe una relación estrecha entre los procesos deterministas (cuyo comportamiento esta determinado por ecuaciones en diferencia) y las series de tiempo que admiten la representación autorregresiva, además el concepto de equilibrio eventual para los primeros esta relacionado con el de estacionariedad para los segundos.

Ecuaciones en diferencia de primer orden

La ecuación en diferencia más simple es la de orden uno, la cual se denota por:

$$Z_t = a_0 + a_1 Z_{t-1} \qquad t = ..., -2, -1, 0, 1, 2, ... \tag{2.65}$$

donde a_0 y $a_1 \neq 0$ son constantes. De manera, simplificada

$$(1 - a_1 L) Z_t = a_0 \qquad t = ..., -2, -1, 0, 1, 2, ... \tag{2.66}$$

El método más sencillo para resolver ecuaciones en diferencia es el llamado método iterativo, el cual parte de algún valor inicial Z_0 y por aplicación sucesiva (iterativas) encuentra los valores para Z_1, Z_2, etc, dando por resultado la posibilidad de inferir el comportamiento de Z_t en general. El método iterativo se emplea de la siguiente manera, considérese a la ecuación de primer orden (2.65) y el valor inicial Z_0, entonces

$$Z_1 = a_0 + a_1 Z_0$$
$$Z_2 = a_0 + a_1 Z_1 = a_0(1 + a_1) + a_1^2 Z_0$$
$$Z_3 = a_0 + a_1 Z_2 = a_0(1 + a_1 + a_1^2) + a_1^3 Z_0$$

En general,

$$Z_t = a_0 \sum_{j=0}^{t-1} a_1^j + a_1^t Z_0 \quad \textit{para } t \geq 1 \tag{2.67}$$

Si $a_1 \neq 1$, se tiene

$$\sum_{j=0}^{t-1} a_1^j = \frac{1 - a_1^t}{1 - a_1} \tag{2.68}$$

y así, sustituyendo 2.68 en 2.67 obtendremos una solución general para este tipo de ecuaciones en diferencia la cual es:

$$Z_t = a_1^t Z_0 + \frac{1 - a_1^t}{1 - a_1} \quad \textit{con } a_1 \neq 1, Z_0 = \textit{constante} \tag{2.69}$$

Se hará la limitación al caso $|a_1| < 1$, debido a que es cuando converge la serie y se requieren procesos estables. Por otro lado, si desglosamos la ecuación 2.69 a la forma

$$Z_t = a_1^t Z_0 + \frac{a_0}{1 - a_1} - \left(\frac{a_0}{1 - a_1}\right) a_1^t \qquad t = ..., -2, -1, 0, 1, 2, ... \tag{2.70}$$

Ahora nótese que debido al supuesto de que $|a_1| < 1$, $\lim_{t \to \infty} a_1^t = 0$, y por tanto,

$$\lim_{t \to \infty} Z_t = \frac{a_0}{1 - a_1}$$

Lo que significa que la serie tiende a estabilizarse en el punto $\frac{a_0}{1-a_1}$ conforme el tiempo crece. Por otro lado, si $|a_1| > 1$ entonces Z_t nunca se estabilizaría, i.e, el proceso no convergerá. Para el caso $|a_1| = 1$ y si observamos la ecuación 2.69 veremos que el proceso será divergente.

La ecuación en diferencia 2.65 se clasifica como lineal, ya que ningún término de Z aparece elevado a alguna potencia distinta de uno; de primer orden, porque interviene a lo más una diferencia y, por tanto, a lo más un sólo retraso para Z_t, en comparación a una ecuación en diferencia de orden $p > 1$, en donde pueden intervenir hasta p retrasos, es decir, $Z_t, Z_{t-1}, ..., Z_{t-p}$. Este último lo estudiaremos más adelante.

Ecuaciones en diferencia de segundo orden

Al igual que se hizo previamente con las ecuaciones de primer orden, a continuación se hará una breve exposición de cómo se resuelve ahora una ecuación de segundo orden y de cuales son las condiciones para que el proceso alcance un equilibrio a largo plazo. Partiendo de la ecuación en diferencia de segundo orden tenemos:

$$Z_t = a_0 + a_1 Z_{t-1} + a_2 Z_{t-2} \qquad t = ..., -2, -1, 0, 1, 2, ... \tag{2.71}$$

La cual también podemos escribir de la siguiente forma gracias a los operadores de retraso:

$$(1 - a_1 L - a_2 L^2) Z_t = a_0 \qquad t = ..., -2, -1, 0, 1, 2, ... \tag{2.72}$$

En donde naturalmente $a_2 \neq 0$, sino sería de orden uno. La solución[13] general de esta ecuación viene dada por:

$$Z_t = \frac{a_0}{1 - a_1 - a_2} + s_1 g_1^t + s_2 g_2^t \tag{2.73}$$

en donde s_1 y s_2 son constantes que se determinan mediante dos condiciones iniciales. Los valores g_1 y g_2 están relacionados con los coeficientes a_1 y a_2 de 2.72, mediante las ecuaciones:

$$a_1 = g_1 + g_2 \;\; y \;\; a_2 = -g_1 g_2 \tag{2.74}$$

[13]Si se desea ver la comprobación ver Guerrero Guzmán Víctor Manuel, 2003 pp 46 [11].

Que surgen de la siguiente factorización:

$$\begin{aligned} 1 - a_1 L - a_2 L^2 &= (1 - g_1 L)(1 - g_2 L) \\ &= 1 - g_1 L - g_2 L + g_1 g_2 L^2 \\ &= 1 - (g_1 + g_2) L + g_1 g_2 L^2 \end{aligned} \tag{2.75}$$

Con lo cual vemos que se cumplen las igualdades 2.74.

Para determinar los valores de g_1 y g_2 para valores dados de a_1 y a_2, partimos del siguiente argumento: nótese que la ecuación

$$(1 - g_1 x)(1 - g_2 x) = 0 \tag{2.76}$$

Tiene como raíces(o ceros) los valores $x_1 = g^{-1}$ y $x_2 = g_2^{-1}$; por tanto, dados los polinomios $1 - a_1 L - a_2 L^2$ y la factorización 2.75, las raíces g_1^{-1} y g_2^{-1} se encuentran al resolver *la ecuación característica* del proceso:

$$1 - a_1 x - a_2 x^2 = 0 \tag{2.77}$$

Entonces por la formula general de segundo grado obtenemos las raíces que vienen dadas como:

$$x = \frac{-a_1 \pm \sqrt{a_1^2 + 4a_2}}{2a_2} \tag{2.78}$$

Que se puede resolver una vez que se conocen a_1 y a_2.

Es importante distinguir tres casos diferentes en relación con las raíces que obtienen como solución de la ecuación característica, estos son:

Caso1. Si $a_1^2 + 4a_2 > 0$ entonces 2.78 proporciona dos valores reales y diferentes, $x_1 = g_1^{-1} \neq x_2 = g_2^{-1}$, con la condición $|x_1^{-1}| < 1$ y $|x_2^{-1}| < 1$. Si se calcula el límite cuando t tiende a infinito a la solución general, se tendría que

$$\lim_{t \to \infty} Z_t = \lim_{t \to \infty} \left\{ \frac{a_0}{1 - a_1 - a_2} + s_1 g_1^t + s_2 g_2^t \right\} = \frac{a_0}{1 - a_1 - a_2} \tag{2.79}$$

Si por el contrario $|g_1| > 1$ y $|g_2| > 1$, entonces se tendrá que el término $s_1 g_1^t + s_2 g_2^t$ tenderá a crecer rápidamente y no existirá convergencia, lo mismo pasa cuando se tiene el caso $|g_1| > 1$ y $|g_2| < 1$, pues $s_1 g_1^t$ tenderá a crecer mientras que $s_2 g_2^t$ tenderá a cero, con lo cual Z_t no convergerá y lo mismo

ocurre para el caso en que $|g_1| < 1$ y $|g_2| > 1$. Cuando $|g_1| = 1$ y $|g_2| = 1$ se estudiara más adelante (ver procesos divergentes).

Caso2. Si $a_1^2 + 4a_2 < 0$, entonces las dos raíces de 2.78 serán complejos.

$$g_1^{-1} = u + iv \quad y \quad g_2^{-1} = u - iv$$

Estos pueden escribirse en coordenadas polares para ver si Z_t converge o no, es decir,

$$\begin{aligned} g_1^{-1} &= s\exp\{i\theta\} = s[\cos\theta + isen\theta] \\ g_2^{-1} &= s\exp\{-i\theta\} = s[\cos\theta - isen\theta] \end{aligned} \tag{2.80}$$

En donde,

$$s = \sqrt{u^2 + v^2} = \sqrt{g_1^{-1} \cdot g_2^{-1}}$$

y θ es el ángulo, en radianes, que cumple con que $\cos\theta = \frac{u}{s}$ y $sen\theta = \frac{v}{s}$ para $\theta\epsilon[0, 2\pi]$. A partir 2.78 se obtienen los valores de g_1 y g_2

$$g_1 = r[\cos\theta - isen\theta] \quad y \quad g_2 = r[\cos\theta + isen\theta]$$

con $r = s^{-1} = \sqrt{g_1 \cdot g_2}$ así la parte que determina la convergencia o no de Z_t, al punto de equilibrio $a_0/1 - a_1 - a_2$, esta dada por:

$$\begin{aligned} s_1 g_1^t + s_2 g_2^t &= s_1 r^t[\cos(\theta t) - isen(\theta t)] + s_2 r^t[\cos(\theta t) + isen(\theta t)] \\ &= r^t[(s_1 + s_2)\cos(\theta t) + i(s_2 - s_1)sen(\theta t)] \end{aligned} \tag{2.81}$$

La cual seguirá un modelo de fluctuaciones cíclicas que tenderán a aumentar o a disminuir dependiendo del factor r^t. *Parar* < 1 el patrón cíclico disminuirá al grado de desaparecer y habrá convergencia; si $r > 1$, Z_t mostrará oscilaciones explosivas y no convergerá (a menos que las condiciones iniciales hayan sido $Z_0 = Z_1 = a_0/[(1 - g_1)(1 - g_2)]$, en cuyo caso el proceso estará en equilibrio); finalmente, si $r = 1$ el patrón oscilatorio no cambiará y se puede decir que el punto de equilibrio es periódico.

Caso3. Si $a_1^2 + 4a_2 = 0$, la ecuación característica tiene dos raíces iguales $g_1^{-1} = g_2^{-1} = -a_1/2a_2$, en cuyo caso si considera $g = g_1 = g_2$, la nueva ecuación en estudio será

$$(1 - gL)^2 Z_t = a_0 \tag{2.82}$$

cuya solución general está dada por

$$\begin{aligned} Z_t &= (1-gL)^{-2}a_0 + s_1 g^t + s_2 t g^t \\ &= \sum_{j=0}^{\infty}(1+j)g^j a_0 + s_1 g^t + s_2 t g^t \end{aligned} \tag{2.83}$$

Una solución particular de la ecuación se obtiene al determinar las constantes s_1 y s_2, lo cual se puede lograrse si se conocen dos condiciones iniciales. Para que el término $\sum_{j=0}^{\infty}(1+j)g^j$ que aparece en 2.83 sea finito, es necesario que $|g| < 1$, y con esta condición se puede observar que Z_t convergerá al punto de equilibrio(2.79) que en esta ocasión se convierte en:

$$\frac{a_0}{(1-g)^2} \tag{2.84}$$

En conclusión se puede observar que la condición para que exista la convergencia en los tres casos anteriores es que los módulos de g_1 y g_2 sean menores que la unidad y s_1 y s_2 se determinan mediante las condiciones iniciales Z_0 y Z_1.

Entonces, la convergencia de un proceso está en términos de las raíces g_1 y g_2, pero también pueden estar expresadas en función de los parámetros originales a_1 y a_2 que determinan a la ecuación en diferencia. La manera es la siguiente dadas las condiciones $|g_1| < 1$ y $|g_2| < 1$ se puede considerar que lo anterior implica los siguientes casos:

$$\begin{aligned} &i) g_1 < 1 \qquad ii) - g_1 < 1 \qquad iii) g_2 < 1 \\ &\qquad\qquad iv) - g_2 < 1 \qquad v) |g_1 g_2| < 1 \end{aligned} \tag{2.85}$$

a su vez de i) y iii) se puede obtener que $g_1(1-g_2) < (1-g_2)$, de igual manera de ii) y iv) puede derivarse el resultado $-g_1(1+g_2) < (1+g_2)$, desarrollando y despejando las restricciones se llega a que

$$g_1 + g_2 - g_1 g_2 < 1 \qquad -g_1 - g_2 - g_1 g_2 < 1 \qquad y \qquad |g_1 g_2| < 1 \tag{2.86}$$

Que en términos de los parámetros originales y por la factorización 2.75 se obtienen las nuevas condiciones para la estabilidad o convergencia del proceso:

$$a_1 + a_2 < 1 \qquad -a_1 + a_2 < 1 \qquad y \qquad |a_2| < 1 \tag{2.87}$$

Estas últimas son más fáciles de verificar que las condiciones 2.85, si es que no hay necesidad de calcular los valores g_1 y g_2, lo cual ocurre en el caso de que no se requiera encontrar la solución para la ecuación en diferencia y simplemente se desee indicar si el proceso alcanzará o no su equilibrio a largo plazo.

Ecuaciones en diferencia de orden p

Ahora se verá a grandes rasgos la metodología para resolver ecuaciones en diferencia de orden $p \geq 2$. Si se considera la ecuación general de la forma

$$(1 - a_1 L - a_2 L^2 - ... - a_p L^p) Z_t = a_0 \qquad \begin{matrix} t = ..., -2, -1, 0, 1, 2, ... \\ a_p \neq 0 \end{matrix} \tag{2.88}$$

y utilizando el teorema fundamental del Algebra[14], el polinomio de retraso involucrado queda como sigue

$$(1 - g_1 L)(1 - g_2 L)...(1 - g_p L) = G(L) \tag{2.89}$$

de tal forma que las raíces de la ecuación característica

$$G(x) = 0 \tag{2.90}$$

son $x = g_1^{-1}, g_2^{-1}, ..., g_p^{-1}$. Si se supone que todas las raíces son distintas, entonces la solución general de 2.88 en términos del polinomio $G(L)$, viene a ser:

$$Z_t = \frac{a_0}{(1 - g_1)(1 - g_2)...(1 - g_p)} + s_1 g_1^t + s_2 g_2^t + ... + s_p g_p^t \tag{2.91}$$

en donde $s_1, s_2, ..., s_p$ son constantes que se determinan con base en p condiciones iniciales. Si se tiene una raíz real m veces repetida con $1 < m \leq p$, es decir, $g_1 = g_2 = ... = g_m = g$, entonces la solución general sería:

$$Z_t = \frac{a_0}{(1 - g)^m (1 - g_{m+1})...(1 - g_p)} + s_1 g^t + s_2 g^t + ... + s_m t^{m-1} g^t + s_{m+1} g_{m+1}^t + ... + s_p g_p^t \tag{2.92}$$

[14]Todo polinomio de grado positivo sobre el campo de los complejos tiene un cero, Hasser Norman B. 1999, pp502 [13]

y bajo el supuesto de que $|g| < 1$ y tomando a $(1-g)^{-1}$ como una función f con argumento g, se tiene el siguiente procedimiento:

$$f(g) = \frac{1}{(1-g)} = \sum_{j=0}^{\infty} g^j \tag{2.93}$$

al obtener la primera derivada de f con respecto a g se obtiene

$$f'(g) = \frac{1}{(1-g)^2} = \sum_{j=0}^{\infty} j g^{j-1} = \sum_{j=0}^{\infty} (j+1) g^j \tag{2.94}$$

tomando la segunda derivada de f, se obtiene la expresión

$$f''(g) = \frac{2}{(1-g)^3} = \sum_{j=0}^{\infty} j(j-1) g^{j-2} \tag{2.95}$$

de donde se puede deducir que

$$\frac{1}{(1-g)^3} = \sum_{j=0}^{\infty} \frac{(j+2)(j+1)}{2} g^j \tag{2.96}$$

Al considerar el cociente $1/(1-g)^p$ para cualquier $p \geq 2$ se tiene que la expresión quedaría como

$$\frac{1}{(1-g)^p} = \sum_{j=0}^{\infty} \frac{(p-1+j)(p-2+j)...(j+2)(j+1)}{(p-1)!} g^j \tag{2.97}$$

por lo tanto, dada la condición $|g| < 1$, la solución general de 2.92 con m=p queda expresada como

$$Z_t = a_0 \sum_{j=0}^{\infty} \frac{(p-1+j)(p-2+j)...(j+2)(j+1)}{(p-1)!} g^j + g^t \sum_{i=1}^{p} s_i t^{i-1} \tag{2.98}$$

En el caso de las ecuaciones de primer y segundo orden apreciamos que, el que un proceso (descrito por una ecuación en diferencia) alcance su punto de equilibrio, depende de que el recíproco del modulo de cada una de las raíces de la ecuación característica, sea menor que la unidad. Así pues, para el caso general 2.88, se tiene que para que dicha condición se cumpla es equivalente a

que se verifique el siguiente resultado, al cual se le conoce como el ***Teorema de Schur***[15]

Los módulos de las raíces de la ecuación

$$g^p - a_1 g^{p-1} - a_2 g^{p-2} - ... - a_{p-1} g - a_p = 0 \tag{2.99}$$

Sean menores que la unidad, si y solo si los p determinantes que se muestran a continuación son positivos.

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

$$\mathbf{D_2} = \begin{vmatrix} -1 & 0 & a_p & a_{p-1} \\ a_1 & -1 & 0 & a_p \\ a_p & 0 & -1 & a_1 \\ a_{p-1} & a_p & 0 & -1 \end{vmatrix}$$

$$\vdots$$

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

La utilidad de este teorema se ve clarificada si se considera la ecuación característica del proceso representado por 2.88, la cual es

$$1 - a_1 x - a_2 x^2 - ... - a_{p-1} x^{p-1} - a_p x^p = 0 \tag{2.100}$$

que al multiplicarse por x^{-p} se convierte en

$$(x^{-1})^p - a_1 (x^{-1})^{p-1} - a_2 (x^{-1})^{p-2} - ... - a_{p-1}(x^{-1}) - a_p = 0 \tag{2.101}$$

[15]Ver Lawrence E. Stephen Arnold y H. Friedberg Insel Spence, 1989, 338pp. [18].

en donde, al identificar a x^{-1} con la g de 2.99, sobresale la relación existente entre el teorema de Schur y el criterio para determinar la convergencia del proceso. Es decir, lo único que se deberá hacer para verificar si una cierta ecuación en diferencia de orden p representa un proceso convergente, es calcular los p determinantes (si $p = 2$ se deberá calcular D_1 y D_2 definidos previamente y así para cada caso), si **todos** éstos son positivos podrá concluirse que el proceso tenderá a equilibrarse en el largo plazo; por el contrario, bastara que uno de los determinantes no sea positivo para concluir que el proceso no se estabilizará jamás. Es interesante ver que gracias a esto también nos arroja las mismas condiciones de 2.87 cuando $p = 2$, al igual que cuando $p = 1$ que debe ser $|a_1| < 1$

Procesos divergentes

Aquí veremos el caso especial de procesos divergentes en el que intervienen raíces unitarias de la ecuación característica, asunto que quede pendiente en el caso 1 en que $a_1^2 + 4a_2 > 0$. Este tipo de procesos puede ser representado por una ecuación en diferencia de orden p+d dada por:

$$\begin{aligned} A(L)Z_t &= a_0 \\ &= G(L)(1-L)^d \qquad con\ d \geq 0 \\ &= (1 - a_1L - a_2L^2 - ... - a_{p+d}L^{p+d}) \qquad y \\ G(L) &= (1-g_1L)(1-g_2L)...(1-g_pL) \end{aligned} \tag{2.102}$$

Y debido a que el comportamiento de un proceso representado por una ecuación en diferencia, depende del tipo de raíces de su ecuación característica entonces esta última se puede denotar como:

$$(1-g_1x)(1-g_2x)...(1-g_px)(1-x)^d = 0 \tag{2.103}$$

de donde se sigue que existen d raíces unitarias, además de las p raíces no unitarias $g_1^{-1}, g_2^{-1}, ..., g_p^{-1}$. En particular, nótese que las raíces de la ecuación (2.103) se obtienen como solución de las dos ecuaciones características

$$(1-g_1x)(1-g_2x)...(1-g_px) = 0, \qquad (1-x)^d = 0 \tag{2.104}$$

que se podrían asociar con los procesos

$$G(L)Z_t = a_0 \qquad y \qquad \nabla^d Z_t = 0 \tag{2.105}$$

respectivamente. El primero de estos procesos ya lo hemos estudiado en lo que toca a su convergencia en el punto de equilibrio con el hecho de que $|g_i| < 1$ para $i = 1, 2, ..., p$. En cuanto al proceso $\nabla^d Z_t = 0$ es divergente , ya que las raíces de su ecuación característica son todas unitarias, pero es interesante conocer el comportamiento que sigue dicho proceso. Se tomarán los casos más sencillos y de ahí generalizaremos. Primero consideremos la ecuación

$$\nabla Z_t = 0, \qquad t = 1, 2, ... \tag{2.106}$$

del cual obtenemos $\nabla Z_t = Z_t - Z_{t-1} = 0 \Longleftrightarrow Z_t = Z_{t-1}$ Luego observemos

$$\begin{aligned}
\nabla Z_1 = Z_1 - Z_0 = 0 &\Longrightarrow Z_1 = Z_0 \\
\nabla Z_2 = Z_2 - Z_1 = 0 &\Longrightarrow Z_2 = Z_1 \\
&... \\
\nabla Z_t = Z_t - Z_{t-1} = 0 &\Longrightarrow Z_t = Z_{t-1} \\
\Longrightarrow Z_t = Z_{t-1} = ... = Z_4 = Z_3 = Z_2 = Z_1 = Z_0 &
\end{aligned}$$

en general $Z_t = Z_{t-1}$ para $i = 1, 2, ..., t$ que es lo mismo que poner $Z_t = Z_0$ $t = 1, 2, ...$ Por lo tanto, el proceso queda completamente determinado al conocer la condición inicial. Ahora considérese a la ecuación

$$\nabla^2 Z_t = 0, \qquad t = 2, 3, ... \tag{2.107}$$

donde desarrollando el lado derecho de la igualdad se tiene que

$$\begin{aligned}
\nabla(\nabla Z_t) = \nabla(Z_t - Z_{t-1}) = \nabla Z_t - \nabla Z_{t-1} = Z_t - Z_{t-1} - (Z_{t-1} - Z_{t-2}) = \\
= Z_t - 2Z_{t-1} + Z_{t-2}
\end{aligned} \tag{2.108}$$

igualando a cero se tiene que:

$$\begin{aligned}
Z_t - 2Z_{t-1} + Z_{t-2} = 0 \\
Z_t = 2Z_{t-1} - Z_{t-2}
\end{aligned} \tag{2.109}$$

y resolviendo de manera iterativa obtenemos:

$$\begin{aligned} Z_t &= 2Z_{t-1} - Z_{t-2} \\ &= 2(2Z_{t-2} - Z_{t-3}) - Z_{t-2} \\ &= 3Z_{t-2} - 2Z_{t-3} \\ &= 3(2Z_{t-3} - Z_{t-4}) - 2Z_{t-2} \\ &= 4Z_{t-3} - 3Z_{t-4} \\ &= 4(2Z_{t-4} - Z_{t-5}) - 3Z_{t-4} \\ &= 5Z_{t-4} - 4Z_{t-5} \\ &\cdots \\ &= nZ_{t-(n-1)} - (n-1)Z_{t-n} \quad n \geq 2 \end{aligned} \tag{2.110}$$

Ahora bien, $t-(n-1)=0$ ó $t-n=0$.Lo anterior implica $t=n$ ó $t=n-1$ y se elige como máximo valor posible de n, en consecuencia.

$$\begin{aligned} Z_t &= tZ_{t-(t-1)} - (t-1)Z_0 \\ &= tZ_1 - tZ_0 + Z_0 \end{aligned} \tag{2.111}$$

Retomando estos dos casos se puede decir que para la ecuación en diferencia de primer orden (2.106) se obtiene como solución general una constante, es decir, un polinomio de grado cero, necesitando una condición inicial para la solución particular; mientras que para la ecuación en diferencia de segundo orden(2.107), la solución general es un polinomio de grado uno y requiere dos condiciones iniciales para la solución particular. En consecuencia para el caso general de la ecuación en diferencia de orden d se tendrá como solución general, a un polinomio de grado d-1 y requerirá de d condiciones iniciales para la solución especifica.

Del resultado anterior, se puede afirmar que el proceso representado por (2.102) con $|g_i| < 1$ para $i = 1, 2, ..., p$, es divergente debido a que existe una tendencia polinomial, la cual al ser eliminada, el proceso que resulta es convergente. De manera que si se considera una nueva variable

$$W_t = \nabla^d Z_t$$

el proceso queda entonces definido por:

$$G(L)W_t = a_0$$

el cual tendera a estabilizarse eventualmente. Cabe señalar que teniendo la expresión:

$$A(L)Z_t = G(L)\nabla^d Z_t = G(L)W_t = a_0$$

El proceso en términos de la variable original Z es divergente (por las d raíces unitarias de $A(x) = 0$), pero en términos de la variable W sí puede ser convergente (cuando $|g_1| < 1, |g_2| < 1, ..., |g_p| < 1$).

2.5.5. Modelos autorregresivos (AR)

Recordando las ecuaciones en diferencia en las cuales teníamos modelos del tipo:

$$A(L)Y_t = constante \tag{2.112}$$

en donde A(L) es un polinomio de retraso. Una generalización de este tipo de ecuaciones consiste en introducir una variable aleatoria en el lado derecho de la expresión y así obtener:

$$A(L)Y_t = constante + e_t \tag{2.113}$$

en donde por simplicidad, se supone que $\{e_t\}$ es un proceso de ruido blanco, las ecuaciones en diferencia del tipo 2.113 permiten representar los *procesos autorregresivos* que desarrollando el polinomio, tenemos:

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)Y_t = constante + e_t \tag{2.114}$$

en donde la constante es igual a $(1 - \phi_1 - \phi_2 - ... - \phi_p)\mu$, en caso de que el proceso Y_t tenga un nivel medio constante dado por $E[Y_t] = \mu$ para toda t; de esta forma, la ecuación 2.114 se convierte en:

$$\phi(L)\widetilde{Y}_t = e_t \quad con \quad \widetilde{Y}_t = Y_t - \mu \tag{2.115}$$

El término autorregresivo(AR) que se le da al proceso representado por 2.115 se refiere al hecho de que también puede expresarse como

$$Y_t = (1 - \phi_1 - \phi_2 - ... - \phi_p)\mu + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + e_t \tag{2.116}$$

la cual si nos damos cuenta es una ecuación de regresión lineal, con la característica especial de que el valor de la variable dependiente Y en el periodo t

depende, no de los valores de un cierto conjunto de variables independientes, como sucede en el modelo de regresión, sino de sus propios valores, observados en periodos anteriores a t y ponderados de acuerdo con los coeficientes autorregresivos $\phi_1, ..., \phi_p$.

Como vimos, es importante saber si el proceso asociado con una ecuación en diferencia alcanzará en el largo plazo su punto de equilibrio. Al referirse a ecuaciones en diferencia en las que interviene algún elemento aleatorio, no es estrictamente válido hablar de convergencia debido precisamente a las fluctuaciones aleatorias que siempre existirán aun cuando éstas ocurran alrededor del punto de equilibrio. Por tal razón, es necesario utilizar el concepto de equilibrio estocástico, mejor conocido como *estacionariedad*, esto es mientras que en un proceso determinista se habla de equilibrio, cuando se tiene un proceso estocástico se habla de estacionaricdad; de esta manera, un proceso AR será estacionario o no estacionario, dependiendo de los valores que tomen las raíces de la ecuación característica:

$$\phi(x) = 0 \tag{2.117}$$

la cual rige el comportamiento del proceso autorregresivo.
Si recordamos el caso general de una ecuación en diferencia, se sabe que $\phi(L)$ se escribe como:

$$\phi(L) = (1 - g_1 L)(1 - g_2 L)...(1 - g_p L) \tag{2.118}$$

de tal manera que el proceso AR definido por $\phi(L)$ será estacionario, si y solo si,

$$|\, g_i| < 1 \qquad para \;\; i = 1, 2, ..., p \tag{2.119}$$

o dicho de otra forma, si las raíces de 2.117, que son $g_1^{-1}, g_2^{-1}, ..., g_p^{-1}$, se encuentran fuera del circulo unitario (en el plano complejo).

Modelos autorregresivos de primer orden AR(1)

Iniciaremos con el caso mas simple de un modelo autorregresivo de orden uno, i.e, un AR(1), que se representa como:

$$\widetilde{Y}_t = \phi \widetilde{Y}_{t-1} + e_t \tag{2.120}$$

para que este modelo sea estacionario se requiere que la raíz de la ecuación

$$1 - \phi x = 0 \tag{2.121}$$

se encuentra fuera del círculo unitario; es decir, se requiere que $|\phi| < 1$ para asegurar la estacionariedad del proceso AR(1) descrito por 2.120. La media y la varianza de este modelo vienen dadas como:

$$E[\tilde{Y}_t] = 0 \quad y \quad \gamma_0 = Var(\tilde{Y}_t) = \sigma^2(1 + \phi^2 + \phi^4 + ...) \tag{2.122}$$

de tal forma que tanto la media como la varianza del modelo son constantes y además se obtiene:

$$\gamma_0 = \sigma^2/(1 - \phi^2) \tag{2.123}$$

Asimismo, las autocovarianzas vienen dadas como:

$$\begin{aligned} \gamma_\kappa &= \sigma^2(\sum_{i=1}^{\infty} \phi^i \phi^{\kappa+i} + \phi^\kappa) \\ &= \sigma^2 \phi^\kappa \sum_{i=0}^{\infty} \phi^{2i}, \quad \kappa = 1, 2, ... \end{aligned} \tag{2.124}$$

si $|\phi| < 1$ se reduce a:

$$\gamma_\kappa = \sigma^2 \phi^\kappa/(1 - \phi^2), \quad \kappa = 1, 2, ... \tag{2.125}$$

y recordando que $\gamma_\kappa = \gamma_{-\kappa}$, se obtiene la fórmula general:

$$\gamma_\kappa = \sigma^2 \phi^{|\kappa|}/(1 - \phi^2), \quad \kappa = \pm 1, \pm 2, ... \tag{2.126}$$

de donde se sigue que las autocorrelaciones deben ser de la forma:

$$\rho(\kappa) = \frac{\gamma_\kappa}{\gamma_0} = \phi^{|\kappa|}, \quad \kappa = \pm 1, \pm 2, ... \tag{2.127}$$

lo cual nos indica que, conforme $\kappa > 0$ crece, la función de autocorrelación (FAC) tiende a cero, con decaimiento de tipo exponencial cuando $0 < \phi < 1$ y con signos alternados cuando $-1 < \phi < 0$. Además debido a la condición 2.127 la condición de estacionariedad del proceso AR(1), $|\phi| < 1$, se convierte en términos de las autocorrelaciones como:

$$|\rho(1)| < 1 \tag{2.128}$$

Modelos autorregresivos de orden p AR(p)

Como caso general de un proceso autorregresivo, se procede a considerar el proceso AR(p) que se describe 2.116, la cual es equivalente a:

$$\widetilde{Y}_t = \phi_1\widetilde{Y}_{t-1} + \phi_2\widetilde{Y}_{t-2} + ... + \phi_p\widetilde{Y}_{t-p} + e_t \qquad (2.129)$$

en donde $\widetilde{Y}_t = Y_t - \mu$.

Un proceso AR(p) será estacionario, si y solo si, las raíces de la ecuación característica:

$$1 - \phi_1 x - \phi_2 x^2 - ... - \phi_p x^p = 0 \qquad (2.130)$$

se encuentran fuera del circulo unitario. En la practica, el teorema de Schur es útil para encontrar las condiciones de estacionariedad en términos de los parámetros $\phi_1, ..., \phi_p$. Dichas condiciones surgen del requisito de que sean positivos los p determinantes que se muestran a continuación:

$$\mathbf{D_1} = \begin{vmatrix} -1 & \phi_p \\ \phi_p & -1 \end{vmatrix}$$

$$\mathbf{D_2} = \begin{vmatrix} -1 & 0 & \phi_p & \phi_{p-1} \\ \phi_1 & -1 & 0 & \phi_p \\ \phi_p & 0 & -1 & \phi_1 \\ \phi_{p-1} & \phi_p & 0 & -1 \end{vmatrix}$$

$$\vdots$$

$$\mathbf{D_p} = \begin{vmatrix} -1 & 0 & ... & 0 & \phi_p & \phi_{p-1} & ... & \phi_1 \\ \phi_1 & -1 & ... & 0 & 0 & \phi_p & ... & \phi_2 \\ ... & & ... & & ... & & ... & \\ \phi_{p-1} & \phi_{p-2} & ... & -1 & 0 & 0 & ... & \phi_p \\ \phi_p & 0 & ... & 0 & -1 & \phi_1 & ... & \phi_{p-1} \\ \phi_{p-1} & \phi_p & ... & 0 & 0 & -1 & ... & \phi_{p-2} \\ ... & & ... & & ... & & ... & \\ \phi_1 & \phi_2 & ... & \phi_p & 0 & 0 & ... & -1 \end{vmatrix}$$

Por otro lado, es conveniente apreciar que un proceso AR(p) estacionario tiene asociada una FAC que decae rápidamente a cero, de manera exponencial. Si el proceso AR(p) resulta ser estacionario, será posible representarlo como una suma ponderada de choques aleatorios con ponderaciones absolutamente convergentes; es decir, debe poderse escribir como:

$$\widetilde{Y}_t = e_t - \psi_1 e_{t-1} - \psi_2 e_{t-2} - ... \tag{2.131}$$

con $\sum_{i=1}^{\infty} |\psi_i| = constante < \infty$. Los coeficientes ψ_i, $i = 1, 2, 3, ...$ se obtienen a partir del hecho de que un proceso AR(p) estacionario se debe expresar como:

$$\phi(L)\widetilde{Y}_t = e_t \quad y \quad \widetilde{Y}_t = \psi(L)e_t \tag{2.132}$$

es decir,

$$\psi(L) = 1/\phi(L) \quad o \quad 1 = \phi(L)\psi(L) \tag{2.133}$$

donde $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_{p-1}L^{p-1} - \phi_p L^p)$, de manera que:

$$\begin{aligned}
1 &= (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_{p-1}L^{p-1} - \phi_p L^p)(1 - \psi_1 L - \psi_2 L^2 - ...) \\
&= (1 - \psi_1 L - \psi_2 L^2 - ...) - \phi_1(L - \psi_1 L^2 - \psi_2 L^3 - ...) - ... \\
&\quad - \phi_{p-1}(L^{p-1} - \psi_1 L^p - \psi_2 L^{p+1} - ...) \\
&\quad - \phi_p(L^p - \psi_1 L^{p+1} - \psi_2 L^{p+2} - ...) \\
&= 1 - (\psi_1 + \phi_1)L - (\psi_2 - \phi_1\psi_1 + \phi_2)L^2 - ... \\
&\quad - (\psi_{p-1} - \phi_1\psi_{p-2} - \phi_2\psi_{p-3} - ... + \phi_{p-1})L^{p-1} \\
&\quad - (\psi_p - \phi_1\psi_{p-1} - \phi_2\psi_{p-2} - ... - \phi_{p-1}\psi_1 + \phi_p)L^p - ...
\end{aligned} \tag{2.134}$$

así que, para que se cumpla esta última relación, todos los coeficientes que aparecen multiplicando a L^i deben de ser cero, para toda $i \geq 1$, lo cual implica que:

$$\begin{aligned}
\psi_1 &= -\phi_1 \\
\psi_2 &= \phi_1\psi_1 - \phi_2 \\
&... \\
\psi_{p-1} &= \phi_1\psi_{p-2} + \phi_2\psi_{p-3} + ... - \phi_{p-1} \\
\psi_p &= \phi_1\psi_{p-1} + \phi_2\psi_{p-2} + ... + \phi_{p-1}\psi_1 - \phi_p
\end{aligned} \tag{2.135}$$

y en general,

$$\psi_i = \phi_1\psi_{i-1} + \phi_2\psi_{i-2} + ... + \phi_{i-1}\psi_1 - \phi_i, \quad para \ \ i \geq 2 \tag{2.136}$$

con $\phi_i = 0$ para $i > p$. Nótese cómo el mismo proceso AR(p) puede representarse con p parámetros autorregresivos o mediante un número infinito de coeficientes ψ asociados con el proceso $\{e_t\}$. Con fines prácticos, resulta preferible trabajar con un número finito de parámetros, sobre todo si la explicación que se tiene del fenómeno es la misma. A esta idea de ahorro en el número de parámetros se le conoce con el nombre de ***principio de parsimonia***. Por consiguiente, la expresión 2.131 no brindará mucha utilidad si todos los coeficientes $\psi_1, \psi_2, ...$, fuesen distintos de cero, como ocurre con el caso de un proceso AR(1) estacionario, en el cual $(1 - \phi L)\widetilde{Y}_t = e_t$ implica:

$$\widetilde{Y}_t = (1 - \phi L)^{-1}e_t = (1 + \phi L + \phi L^2 + ...)e_t \tag{2.137}$$

de tal forma que $\psi_i = -\phi^i$, para $i = 1, 2, ...$ y, por tanto, $\psi_i \neq 0$ para toda i. En muchas ocasiones, sin embargo, dichos coeficientes son distintos de cero número, a partir del cual todos son cero, o sea que se tiene:

$$\psi_1 \neq 0, \psi_2 \neq 0, ..., \psi_q \neq 0, \psi_{q+1} = 0, \psi_{q+2} = 0, ... \tag{2.138}$$

Como ejemplo de esto, pensemos en el caso extremo en que p fuese infinito y que el proceso estacionario tuviese la representación:

$$(1 - \phi L + \phi^2 L^2 - ...)\widetilde{Y}_t = e_t \quad con \ \ |\phi| < 1 \tag{2.139}$$

la cual también se escribe como:

$$\begin{aligned} \widetilde{Y}_t &= (1 - \phi L + \phi^2 L^2 - ...)^{-1}e_t \\ &= (1 + \phi L)e_t \\ &= (1 - \psi_1 L - \psi_2 L^2 - ...)e_t \end{aligned} \tag{2.140}$$

en donde $\psi_1 = -\phi$ y $\psi_i = 0$ para $i \geq 2$. El ahorro en parámetros en este ejemplo sería infinito, por lo cual la representación del proceso en 2.140 es superior a la representación 2.139, aunque ambas sean equivalentes. El argumento anterior conduce a pensar en la existencia de procesos que puedan representarse mejor en términos de los choques aleatorios, que en términos autorregresivos. A este tipo de procesos se les denomina de promedios móviles y se les denota por MA.

2.5.6. Modelos de promedios móviles (MA)

La característica que define el proceso MA en general es que el valor actual de la serie observada se expresa como función de choques actuales y rezagados inobservables, como si fuera un modelo de regresión sólo con perturbaciones actual y retrasadas en el lado derecho.

El modelo MA de orden 1 MA(1)

El proceso de promedios móviles de orden uno, es el más simple y se expresa como sigue:

$$\widetilde{Y}_t = (1 - \theta L)e_t = e_t - \theta e_{t-1} \tag{2.141}$$

Del cual podemos obtener de manera inmediata:

$$E[\widetilde{Y}_t] = E[(e_t - \theta e_{t-1})] = 0 \tag{2.142}$$

y

$$\gamma(0) = Var(\widetilde{Y}_t) = Var(e_t - \theta e_{t-1}) = \sigma^2(1 + \theta^2) \tag{2.143}$$

Observemos que para un valor fijo de σ, cuando θ aumenta en valor absoluto, también lo hace la varianza. Luego la función de autocovarianza esta dada por:

$$\gamma(\kappa) = E[Y_t, Y_{t-\kappa}] = E[(e_t - \theta e_{t-1})(e_{t-\kappa} - \theta e_{t-\kappa-1})] = \begin{cases} -\theta\sigma^2, & \kappa = 1 \\ 0, & \kappa \geq 2 \end{cases} \tag{2.144}$$

Por lo que la función de autocorrelación esta dada como:

$$\rho(\kappa) = \frac{\gamma(\kappa)}{\gamma(0)} = \begin{cases} \frac{-\theta}{1+\theta^2}, & si\ \kappa = 1 \\ 0, & \kappa \geq 2 \end{cases} \tag{2.145}$$

Se puede concluir, del hecho de que las autocorrelaciones para retrasos mayores que un periodo sean cero, que el proceso MA(1) no tiene memoria para más allá de lo ocurrido en un periodo anterior. Pero, aunque la primera autocorrelación sea distinta de cero, no puede ser muy grande, puesto que este hecho indicaría que existiría una fuerte dependencia de la observación actual con la anterior y, así sucesivamente.Entonces, sería más adecuado pensar en

un modelo autorregresivo para esta situación. De la ecuación (2.145) y debido a que $|\theta| < 1$, se obtiene que:

$$|\rho_1| \leq 0{,}5 \tag{2.146}$$

Así, se puede afirmar que aún cuando una cierta función de autocorrelación muestre al retraso 1, ésta representará a un proceso MA(1) solo si se satisface la restricción (2.146).

Cabe señalar, que los procesos autorregresivos estacionarios también pueden representados a través de modelos de promedios móviles, en particular un modelo $AR(\infty)$ tiene una representación equivalente en el modelo MA(1), una conclusión que surge de esto es que el proceso descrito por 2.141 puede representarse en forma autorregresiva, con ponderaciones absolutamente convergentes si $|\theta| < 1$.

En general, cuando un proceso se expresa apropiadamente mediante un modelo AR, se dirá que dicho proceso es *invertible*, lo cual significa que se puede representar como:

$$\pi(L)\widetilde{Y}_t = e_t \tag{2.147}$$

en donde:

$$\pi(L) = 1 - \pi_1 L - \pi_2 L^2 - ... \tag{2.148}$$

es un polinomio de retraso que cumple con que la suma

$$\pi(x) = 1 - \sum_{i=1}^{\infty} \pi_i x^i \tag{2.149}$$

converge dentro o sobre el círculo unitario. Este requisito podría interpretarse como una restricción sobre los coeficientes $\pi_1, \pi_2, ...$ (que se encuentran asociados con las variables retrasadas $\widetilde{Y}_{t-1}, \widetilde{Y}_{t-2}, ...$) de tal manera que mientras mayor sea el retraso de la variable, menor deberá ser el valor de la π correspondiente, es decir, es menor la influencia de dicha variable retrasada sobre las variables más recientes. Todo proceso MA es estacionario, mientras que todo proceso AR es invertible. Además, las condiciones para invertibilidad de un proceso MA se obtienen de manera similar a las condiciones de estacionariedad para procesos AR, como se aprecia en la condición de invertibilidad del proceso 2.141, la cual surge del requerimiento de que la suma:

$$\pi(x) = \theta^{-1}(x) = 1 + \sum_{i=1}^{\infty} \theta^i x^i \tag{2.150}$$

converja dentro o sobre el circulo unitario, lo cual recordando sucede cuando $|\theta| < 1$.

En consecuencia, la condición de invertibilidad para un proceso MA también se expresa en términos del polinomio $\theta(L)$ por el requerimiento de que las raíces de la ecuación

$$\theta(x) = 0 \tag{2.151}$$

se encuentran fuera del circulo unitario. La importancia del concepto de invertibilidad radica en que *todo proceso invertible esta determinado de manera única por su FAC*, esto es que cumplen con las condiciones antes mencionadas, lo cual no ocurre cuando los procesos son no invertibles.

El modelo MA de orden q MA(q)

Generalizando, decimos que un proceso estocástico sigue un esquema de promedios móviles de orden $q \geq 1$ si se puede representar como:

$$\widetilde{Y}_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} \tag{2.152}$$

con $\widetilde{Y}_t = Y_t - \mu$, en donde μ es el nivel del proceso, $\theta_1, \theta_2, ..., \theta_q$ son los parámetros de promedios móviles y e_t es un proceso de ruido blanco con media cero y varianza constante σ^2. Recordando que todo proceso MA es estacionario y, en particular, se observa en las formulas siguientes que ni la media, ni la varianza, ni las covarianzas del proceso MA(q), dependen del tiempo:

$$\begin{aligned} E[\widetilde{Y}_t] &= E[Y_t] - \mu = 0 \\ \gamma_0 &= (1 + \theta_1^2 + \theta_2^2 + ..., \theta_q^2)\sigma^2 \\ \gamma_\kappa &= \begin{cases} (-\theta_\kappa + \theta_1\theta_{\kappa+1} + ... + \theta_{q-\kappa}\theta_q)\sigma^2, & si\ \kappa = 1, 2, ..., q \\ 0, & si\ \kappa \geq q+1 \end{cases} \end{aligned} \tag{2.153}$$

para que esta expresión de γ_κ sea valida en general, se define $\theta_0 = \theta_{q+1} = \theta_{q+2} = ... = 0$. De aquí es inmediato obtener la función de autocorrelación:

$$\rho(\kappa) = \begin{cases} \frac{-\theta_\kappa + \theta_1\theta_{\kappa+1} + ... + \theta_{q-\kappa}\theta_q}{1 + \theta_1^2 + \theta_2^2 + ..., \theta_q^2}, & si\ \kappa = 1, 2, ..., q \\ 0, & si\ \kappa \geq q+1 \end{cases} \tag{2.154}$$

la cual muestra que el proceso MA(q) tiene una memoria limitada a q periodos (lo cual se ve claramente en el correlograma de la FAC). Además, la expresión 2.153 permite apreciar que si se deseara obtener un conjunto de ecuaciones que permitan expresar a los parámetros de promedios móviles en términos de las autocovarianzas, dichas ecuaciones serán no lineales y no tendrían una solución única, a menos que se impusieran algunas restricciones sobre los valores de las θ son precisamente las condiciones de invertibilidad, las cuales permiten asociar un solo proceso MA a una FAC. Las condiciones de invertibilidad del proceso 2.152 se obtienen con las condiciones de estacionariedad para un proceso AR(q), de tal manera que, según el teorema de Schur, para que el proceso MA(q) sea invertible, se requiere que los q determinantes:

$$\mathbf{D_1} = \begin{vmatrix} -1 & \theta_q \\ \theta_q & -1 \end{vmatrix}$$

$$\mathbf{D_2} = \begin{vmatrix} -1 & 0 & \theta_q & \theta_{q-1} \\ \theta_1 & -1 & 0 & \theta_q \\ \theta_q & 0 & -1 & \theta_1 \\ \theta_{q-1} & \theta_q & 0 & -1 \end{vmatrix}$$

$$\vdots$$

$$\mathbf{D_q} = \begin{vmatrix} -1 & 0 & \dots & 0 & \theta_q & \theta_{q-1} & \dots & \theta_1 \\ \theta_1 & -1 & \dots & 0 & 0 & \theta_q & \dots & \theta_2 \\ \dots & & \dots & & \dots & & \dots & \\ \theta_{q-1} & \theta_{q-2} & \dots & -1 & 0 & 0 & \dots & \theta_q \\ \theta_q & 0 & \dots & 0 & -1 & \theta_1 & \dots & \theta_{q-1} \\ \theta_{q-1} & \theta_q & \dots & 0 & 0 & -1 & \dots & \theta_{q-2} \\ \dots & & \dots & & \dots & & \dots & \\ \theta_1 & \theta_2 & \dots & \theta_q & 0 & 0 & \dots & -1 \end{vmatrix}$$

sean todos positivos (con un solo determinante que no sea positivo bastara para concluir que es un proceso no invertible). Si el proceso resulta ser invertible, entonces se podría escribir como:

$$\pi(L)\widetilde{Y}_t = \widetilde{Y}_t - \pi_1\widetilde{Y}_{t-1} - \pi_2\widetilde{Y}_{t-2} - \dots = e_t \qquad (2.155)$$

con $\sum_{j=1}^{\infty} |\pi_j| < \infty$, donde los coeficientes $\pi_1, \pi_2, \ldots$ podrían obtenerse de la relación

$$\pi(L)\theta(L) = 1 \tag{2.156}$$

la cual implica que:

$$\begin{aligned} 1 &= (1 - \pi_1 L - \pi_2 L^2 - \ldots)(1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q) \\ &= (1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q) - \pi_1(L - \theta_1 L^2 - \ldots - \theta_q L^{q+1}) \\ &- \pi_2(L^2 - \theta_1 L^3 - \ldots - \theta_q L^{q+2}) - \ldots \\ &= 1 - (\theta_1 + \pi_1)L - (\theta_2 - \pi_1\theta_1 + \pi_2)L^2 - (\theta_3 - \pi_1\theta_2 - \pi_2\theta_1 + \pi_3)L^3 - \ldots \end{aligned} \tag{2.157}$$

de tal manera obtenemos las siguientes ecuaciones:

$$\begin{aligned} \pi_1 &= -\theta_1 \\ \pi_2 &= \pi_1\theta_1 - \theta_2 \\ \pi_2 &= \pi_2\theta_1 + \pi_1\theta_2 - \theta_3 \\ &\ldots \\ \pi_j &= \pi_{j-1}\theta_1 + \pi_{j-2}\theta_2 + \ldots + \pi_1\theta_{j-1} - \theta_j, \quad j > q \end{aligned} \tag{2.158}$$

en donde $\theta_j = 0$ para $j > q$.

2.5.7. Modelos de promedios móviles autorregresivos (ARMA)

En la construcción de modelos, ocasionalmente encontramos algunos que incluyen términos tanto autorregresivos como de promedios móviles y dan como resultado un modelo autorregresivo de promedios móviles (ARMA) de orden(p, q). El cual se puede representar mediante:

$$\phi(L)\widetilde{Y}_t = \theta(L)e_t \tag{2.159}$$

en donde $\phi(L)$ y $\theta(L)$ son polinomios de retraso de orden p y q, respectivamente, $\{e_t\}$ es un proceso de ruido blanco y $\widetilde{Y}_t$ es la serie de desviaciones de la variable Y_t respecto a su nivel medio μ. Esta serie desarrollándola tenemos:

$$\widetilde{Y}_t = \phi_1\widetilde{Y}_{t-1} + \ldots + \phi_p\widetilde{Y}_{t-p} + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \tag{2.160}$$

Las condiciones de estacionariedad e invertibilidad de los procesos AR(p) y MA(q) establecen las propiedades para los procesos $ARMA(p,q)$.

Esto es, un proceso $ARMA(p,q)$ es estacionario si las raíces de $\phi(L) = 0$ están fuera del círculo unitario e invertible si las raíces de $\theta(L) = 0$ están fuera del círculo unitario.

Modelo ARMA(1,1)

El proceso autorregresivo y de promedios móviles de orden (1,1), aun siendo el más sencillo de los procesos ARMA, es de gran interés desde el punto de vista práctico porque proporciona representaciones adecuadas para muchas series de fenómenos reales. El modelo ARMA(1,1) está definido por

$$(1-\phi L)\widetilde{Y}_t = (1-\theta L)e_t \tag{2.161}$$

y puesto que contiene tanto características autorregresivas como de promedios móviles, no tiene por qué ser invertible ni estacionario, pero las condiciones de invertibilidad y estacionariedad se obtienen fácilmente de las condiciones respectivas para procesos AR(1) y MA(1). Si el proceso resulta ser estacionario e invertible, entonces las representaciones,

$$\begin{aligned} \widetilde{Y}_t = \psi(L)e_t = e_t - \psi_1 e_{t-1} - \psi_2 e_{t-2} - \ldots \quad & y \\ \pi(L)\widetilde{Y}_t = \widetilde{Y}_t - \pi_1 \widetilde{Y}_{t-1} - \pi_2 \widetilde{Y}_{t-2} - \ldots = e_t & \end{aligned} \tag{2.162}$$

son tales que las sumas $\sum_{i=1}^{\infty} |\psi_i|$ y $\sum_{j=1}^{\infty} |\pi_j|$ son convergentes. Esto lo podemos verificar al escribir 2.161 como:

$$\begin{aligned} \widetilde{Y}_t &= (\frac{1-\theta L}{1-\phi L})e_t \\ &= (1-\theta L)[1+\phi L+(\phi L)^2+\ldots]e_t \\ &= [1-(\theta-\phi)L-\phi(\theta-\phi)L^2-\phi^2(\theta-\phi)L^3-\ldots]e_t \end{aligned} \tag{2.163}$$

lo cual es válido siempre y cuando se cumpla la condición de estacionariedad $|\phi| < 1$, y en este caso las ponderaciones ψ_i vienen dadas por:

$$\psi_i = \phi^{i-1}(\theta-\phi) \quad i = 1,2,\ldots \tag{2.164}$$

En esta última expresión se observa que, conforme el índice i crece, ψ_i tiende a cero y la suma $\sum_{i=1}^{\infty} |\psi_i|$ será convergente. De igual manera, la representación del proceso ARMA(1,1) como

$$\begin{aligned} e_t &= (\frac{1-\phi L}{1-\theta L})\widetilde{Y}_t \\ &= (1-\phi L)[1+\theta L+(\theta L)^2+...]\widetilde{Y}_t \\ &= [1-(\phi-\theta)L-\theta(\phi-\theta)L^2-\theta^2(\phi-\theta)L^3-...]\widetilde{Y}_t \end{aligned} \tag{2.165}$$

es válida si se satisface que $|\theta| < 1$, la cual es de hecho la condición de invertibilidad, en ese caso tiene:

$$\pi_j = \theta^{j-1}(\phi-\theta) \quad j = 1,2,... \tag{2.166}$$

de donde se sigue que:

$$\sum_{j=1}^{\infty} \pi_j = (\phi-\theta)\sum_{j=1}^{\infty} \theta^{j-1} = \frac{(\phi-\theta)}{(1-\theta)} < \infty. \tag{2.167}$$

La varianza viene dada como:

$$\gamma_0 = \phi\gamma_1 + [1-\theta(\phi-\theta)]\sigma^2 \tag{2.168}$$

Las autocovarianzas vienen dadas como:

$$\gamma_\kappa = \begin{cases} \phi\gamma_0 - \theta\sigma^2, & si \ \kappa = 1 \\ \phi\gamma_{\kappa-1}, & si \ \kappa \geq 2 \end{cases} \tag{2.169}$$

las autocovarianzas para $\kappa \geq 2$ son idénticas a las del proceso AR(1), esto se debe a que parte de promedios móviles es de orden 1 y por tanto, sólo la primera autocovarianza refleja la inclusión de parámetros del tipo MA. Para obtener la autocovarianza, dados los parámetros del proceso, las ecuaciones anteriores se pueden resolver para γ_0 y γ_1, de la manera siguiente:

$$\begin{aligned} \gamma_0 &= \phi\gamma_1 + [1-\theta(\phi-\theta)]\sigma^2 \\ \gamma_1 &= \phi\gamma_0 - \theta\sigma^2 \end{aligned} \tag{2.170}$$

que tiene por solución

$$\begin{aligned} \gamma_0 &= \frac{(1-2\phi\theta+\theta^2)\sigma^2}{1-\phi^2} \\ \gamma_1 &= \frac{(1-\phi\theta)(\phi-\theta)\sigma^2}{1-\phi^2} \end{aligned} \tag{2.171}$$

De 2.170 y 2.171 podemos concluir que la función de autocovarianzas viene dada como:

$$\gamma_\kappa = \frac{\phi^{\kappa-1}(1-\phi\theta)(\phi-\theta)\sigma^2}{1-\phi^2}, \qquad \kappa = 1, 2, ... \tag{2.172}$$

Con lo cual podemos construir la función de autocorrelación(FAC).

$$\rho(\kappa) = \frac{\phi^{\kappa-1}(1-\phi\theta)(\phi-\theta)}{1-2\phi\theta+\theta^2}, \qquad \kappa = 1, 2, ... \tag{2.173}$$

de aquí, debido al factor $\phi^{\kappa-1}$ y la condición de estacionariedad $\mid \phi \mid < 1$, se sigue que $\rho(\kappa)$ tiene un decaimiento exponencial a cero a partir de $\rho(1)$.

Modelo ARMA(p,q)

Un proceso ARMA(p,q) incluye ambos componentes, tanto el autorregresivo como el de promedios móviles, es decir;

$$\widetilde{Y}_t = \phi_1\widetilde{Y}_{t-1} + \phi_2\widetilde{Y}_{t-2} + ... + \phi_p\widetilde{Y}_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{2.174}$$

o la podemos poner en la forma de operador de rezago;

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)\widetilde{Y}_t = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)e_t \tag{2.175}$$

La cual representaremos como:

$$\phi(L)\widetilde{Y}_t = \theta(L)e_t \tag{2.176}$$

En donde los polinomios $\phi(L)$ y $\theta(L)$ son de orden p y q respectivamente. Para que el proceso sea estacionario[16] se requiere que las raíces de $\phi(x) = 0$, estén fuera del círculo unitario y para que sea invertible la condición es que las raíces de la ecuación $\theta(x) = 0$, se encuentren también fuera del círculo unitario; si ocurre $\phi(x) = 0$, admite la siguiente representación:

$$\widetilde{Y}_t = \frac{\theta(L)}{\phi(L)}e_t = \psi(L)e_t \tag{2.177}$$

[16]En el caso del proceso ARMA se requiere comprobar tanto la estabilidad como la invertibilidad, porque están presentes al mismo tiempo los componentes autorregresivos y los promedios móviles

con $\sum_{j=1}^{\infty} |\phi_j| < \infty$. Así, un proceso estacionario de un ARMA depende enteramente de los parámetros autorregresivos $(\phi_1, \phi_2, ..., \phi_p)$ y no de los parámetros de promedios móviles $(\theta_1, \theta_2, ..., \theta_q)$. De la expresión 2.174 se sigue que si el proceso es estacionario, la media es cero; además para $\kappa \geq 0$

$$\begin{aligned}\gamma_\kappa &= E[\widetilde{Y}_t \widetilde{Y}_{t-\kappa}] \\ &= \phi_1 E[\widetilde{Y}_{t-1}\widetilde{Y}_{t-\kappa}] + \phi_2 E[\widetilde{Y}_{t-2}\widetilde{Y}_{t-\kappa}] + ... + \phi_p E[\widetilde{Y}_{t-p}\widetilde{Y}_{t-\kappa}] \\ &+ E[e_t \widetilde{Y}_{t-\kappa}] - \theta_1 E[e_{t-1}\widetilde{Y}_{t-\kappa}] - ... - \theta_q E[e_{t-q}\widetilde{Y}_{t-\kappa}]\end{aligned} \tag{2.178}$$

en donde, como $\widetilde{Y}_{t-\kappa}$ está afectada por los choques aleatorios $e_{t-\kappa}, e_{t-\kappa-1}$, pero es independiente de $e_{t-\kappa+1}, e_{t-\kappa+2}, ...$, y se tiene

$$E[e_{t-i}\widetilde{Y}_{t-\kappa}] = 0 \quad si\ \kappa > i \tag{2.179}$$

y por consiguiente la ecuación 2.178 da como resultado:

$$\gamma_\kappa = \phi_1\gamma_{\kappa-1} + \phi_2\gamma_{\kappa-2} + ... + \phi_p\gamma_{\kappa-p}, \quad para\ \kappa > q \tag{2.180}$$

mientras que para $\kappa \leq q, \gamma_\kappa$ involucrará los parámetros $(\theta_\kappa, \theta_{\kappa+1}, ..., \theta_q)$. Entonces la varianza viene dada como:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + ... + \phi_p\gamma_p + \sigma^2 - \theta_1 E[e_{t-1}\widetilde{Y}_t] - ... - \theta_q E[e_{t-q}\widetilde{Y}_t] \tag{2.181}$$

la cual, debido a $\gamma_1, \gamma_2, ..., \gamma_p$, tiene que ser resuelta simultáneamente con las ecuaciones para estas p autocovarianzas.

En lo que respecta a las autocorrelaciones para procesos ARMA(p,q), están obtienen para retrasos mayores a q, de la relación

$$\rho(\kappa) = \frac{\gamma_\kappa}{\gamma_0} = \phi_1\rho_{\kappa-1} + \phi_2\rho_{\kappa-2} + ... + \phi_p\rho_{\kappa-p}, \quad \kappa > q \tag{2.182}$$

y para retrasos menores a q, la autocorrelación $\rho(\kappa)$ involucrará los parámetros $(\theta_\kappa, \theta_{\kappa+1}, ..., \theta_q)$. Con frecuencia los modelos ARMA son muy exactos y muy parsimoniosos a la vez. Por ejemplo, en un caso particular, se puede recurrir a un AR(5) para obtener la misma aproximación que la que podría obtenerse con un ARMA(2,1), así el AR(5) tiene cinco parámetros que hay que estimar, mientras que el ARMA(2,1) sólo tiene tres. Ahora veremos los métodos más poderosos y más generales para series de tiempo.

2.6. Modelos ARIMA

Hay que notar que hasta ahora solo hemos trabajado con series estacionarias, pero lamentablemente en la realidad la mayoría de las veces no pasa así, debido a que puede existir algún tipo de tendencia o porque este influenciada por un factor semideterminista como lo es la estacionalidad o ambas. Si el problema es la apreciación de una tendencia en el comportamiento de la serie, es posible que sea de carácter polinomial adaptivo y que por tanto pueda eliminarse con la aplicación del *operador diferencia*. Existen otros métodos pero solo nos limitaremos a esté por su sencilles y aplicabilidad, lo cual nos da origen a los modelos ARIMA. Primero iniciaremos el estudio de series de tiempo o procesos no estacionarios causados por una tendencia para seguir con el estudio de series con estacionalidad.

2.6.1. Modelos ARIMA para series con tendencia

Los modelos autorregresivos e integrados de promedios móviles(ARIMA) pueden ser vistos como una generalización de los modelos ARMA. Yaglom (1955) sugirió la posibilidad de que un cierto tipo de no estacionariedad mostrado por algunas series de tiempo, podían representarse como una simple toma sucesiva de diferencias de la serie original. Esto da mucha flexibilidad a los modelos ARMA, puesto que en realidad lo que se hace es aplicar el operador diferencia ∇^d para eliminar una posible tendencia polinomial de orden d, presente en la serie que se analice. Entonces, si un proceso $\{\widetilde{Y}_t\}$ tuviera una tendencia polinomial no determinista es posible construir el proceso estacionario $\{W_t\}$, en donde:

$$W_t = \nabla^d \widetilde{Y}_t \quad \textit{para toda } t \tag{2.183}$$

para esta serie ya sería posible obtener un modelo ARMA; $\phi(L)W_t = \theta(L)e_t$, lo cual sería equivalente a considerar el modelo ARIMA

$$\phi(L)\nabla^d \widetilde{Y}_t = \theta(L)e_t \quad d \geq 1 \tag{2.184}$$

para $\{\widetilde{Y}_t\}$, en donde $\{e_t\}$ es un proceso de ruido blanco. El término "integrado"se refiere a que $\widetilde{Y}_t$ se obtiene de la relación 2.183 por inversión del

operador ∇^d, dando como resultado una suma infinita de términos W_t. Entonces, si aplicamos la serie de Maclaurin a $(1-L)^{-n}$, con n un número entero, se obtiene:

$$(1-L)^{-n} = 1 + nL + \frac{n(n+1)}{2!}L^2 + \frac{n(n+1)(n+2)}{3!}L^3 + ..., \quad (2.185)$$

de este resultado se sigue que, por ejemplo, el inverso del operador ∇ es

$$\nabla^{-1} = (1-L)^{-1} = 1 + L + L^2 + L^3 + ... \quad (2.186)$$

y así, si $W_t = \nabla\widetilde{Y}_t$, se tendría:

$$\widetilde{Y}_t = \nabla^{-1}W_t = W_t + W_{t-1} + W_{t-2} + ... \quad (2.187)$$

El orden del polinomio de retraso $\phi(L)$, del orden del exponente en el operador diferencia y el orden del polinomio de retraso $\theta(L)$, se acostumbra mencionarlos en ese orden, de manera que un modelo ARIMA(p,d,q) indica que consta de un polinomio autorregresivo de orden p, de una diferencia de orden d y de un polinomio de promedios móviles de orden q. De esta manera el modelo 2.184 se escribe como:

$$W_t - \phi_1 W_{t-1} - ... - \phi_p W_{t-p} = e_t - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} \quad con\ W_t = \nabla^d \widetilde{Y}_t \quad (2.188)$$

Ya que el interés primordial se centra en los modelos estacionarios e invertibles, se requiere que las raíces de $\phi(x) = 0$ y las raíces de $\theta(x) = 0$ se encuentren fuera del círculo unitario, o bien, si se considera como operador autorregresivo generalizado a:

$$\varphi(L) = \phi(L)\nabla^d \quad (2.189)$$

la condición es que d de las raíces de $\varphi(x) = 0$ sean unitarias, mientras que las raíces restantes deben estar fuera del círculo unitario. Con el uso de la ecuación 2.189, la expresión 2.188 se convierte en:

$$\widetilde{Y}_t = \varphi_1\widetilde{Y}_{t-1} + \varphi_2\widetilde{Y}_{t-2} + ... + \varphi_{p+d}\widetilde{Y}_{t-p-d} + e_t - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} \quad (2.190)$$

y, la representación en términos de suma ponderada de choques aleatorios viene dada como:

$$\widetilde{Y}_t = \varphi^{-1}(L)\theta(L)e_t = \psi(L)e_t \quad (2.191)$$

en donde, debido a que

$$(1-\varphi_1 L-\varphi_2 L^2-...-\varphi_{p+d}L^{p+d})(1-\psi_1 L-\psi_2 L^2-...) = (1-\theta_1 L-\theta_2 L^2-...-\theta_q L^q) \tag{2.192}$$

en la cual las ponderaciones ψ deben satisfacer la ecuación en diferencia siguiente (donde L opera sobre j)

$$\varphi(L)\psi_j = \phi(L)\nabla^d\psi_j \quad para\ j > max\{p+d-1, q\} \tag{2.193}$$

Asimismo, si el proceso que indica 2.190 es invertible, se tiene:

$$\pi(L)\widetilde{Y}_t = \theta^{-1}(L)\varphi(L)\widetilde{Y}_t = e_t \tag{2.194}$$

con las ponderaciones π que satisfacen:

$$\theta(L)\pi_j = 0 \quad para\ j > max\{p+d, q\} \tag{2.195}$$

Parece entonces razonable usar diferencias sucesivas para transformar una serie de tiempo no estacionaria en una estacionaria. Es de notar que un proceso Y_t de longitud n, es decir $\{Y_1, Y_2, ..., Y_n\}$ al hacer la transformación 2.183 se tiene una serie de longitud n-d. Si una serie de tiempo no estacionaria puede ser transformada a una serie de tiempo estacionaria aplicando un grado adecuado de diferenciación, se dice que la serie original es homogéneamente no estacionaria. Por último si la no estacionariedad se debe también a que la varianza no es constante, quizá la causa sea que en cada punto de observación t, la variable Y_t tiene varianza σ_t^2 la cual es función de su media μ_t; de ocurrir esto, un argumento que se obtiene del trabajo de Bartlett (1947), conduce a determinar una transformación potencia del tipo.

$$T(Y_t) = \begin{cases} Y_t^{\lambda}, & si\ \lambda \neq 0 \\ \log(Y_t), & si\ \lambda = 0 \end{cases} \tag{2.196}$$

Esta transformación es útil para estabilizar la varianza de la serie, antes de cancelar la posible tendencia.

2.6.2. Modelos ARIMA para series con estacionalidad

Hasta ahora hemos abordado modelos para series de tiempo estacionarias y no estacionarias (con tendencia), ahora consideraremos el caso cuando existe

estacionalidad en la serie la cual es muy frecuente encontrarse en la practica el modelo que utilizaremos es el *operador diferencia estacional*, que es muy parecido al operador diferencia que utilizamos en el caso de los modelos no estacionarios por tendencia, solo que se hacen algunas modificaciones debido a que las diferencias se hacen según la periodicidad en la que venga la estacionalidad (E).

Así como se introdujo la notación de operadores para series no estacionarias por tendencia, ahora introduciremos la siguiente notación para series estacionales. El operador diferencia estacional ∇_E^k se define como:

$$\begin{aligned}\nabla_E^k Y_t &= (1-L^E)^k Y_t \\ &= \sum_{j=0}^{k} \frac{k!}{j!(k-j)!}(-1)^j Y_{t-jE} \quad para\ k=0,1,...\ y\ E=1,2,...\end{aligned} \tag{2.197}$$

mientras que un polinomio de retraso estacional de orden p con coeficientes constantes $g_1, ..., g_k$ viene dado por

$$\begin{aligned}G(L^E) &= 1 - g_1 L^E - g_2 L^{2E} - ... - g_k L^{kE} \\ &= 1 - \sum_{j=1}^{k} g_j L^{jE}\end{aligned} \tag{2.198}$$

Para ejemplificar esto sea E=12 y k=2, entonces nos quedaría como:

$$\nabla_{12}^2 Y_t = (1-L^{12})^2 Y_t = Y_t - 2Y_{t-12} + Y_{t-24}$$

y

$$G(L^{12})Y_t = (1 - g_1 L^{12} - g_2 L^{24})Y_t = Y_t - g_1 Y_{t-12} - g_2 Y_{t-24}$$

Es de advertir que al aplicar el operador ∇_E^D se pierden $E \cdot D$ observaciones automáticamente. Con la notación recién introducida es posible obtener representaciones puramente estacionales del tipo $ARIMA(P,D,Q)_E$ como:

$$\Phi(L^E)\nabla_E^D(Y_t - \mu) = \Theta(L^E)e_t \tag{2.199}$$

donde μ es el nivel de $\{Y_t\}$, $\Phi(L^E)$ es la representación de un polinomio autorregresivo estacional de orden P, $\Theta(L^E)$ denota a un polinomio de promedios móviles de orden Q y la sucesión $\{e_t\}$ es ruido blanco. Ahora antes de ver

un ARIMA estacional veamos sus partes empezando por el elemento autorregresivo.

Sea un proceso estacional autorregresivo con E observaciones por período estacional y donde sólo los parámetros cuyo subíndice es un entero múltiplo de E, son diferentes de cero.

$$\widetilde{Y}_t = \Phi_1\widetilde{Y}_{t-E} + \Phi_2\widetilde{Y}_{t-2E} + \ldots + \Phi_P\widetilde{Y}_{t-PE} + e_t \quad con\ \widetilde{Y}_t = Y_t - \mu \tag{2.200}$$

donde P es el múltiplo mayor de E presente en el proceso el cual nos da el orden del mismo, entonces para un $AR(1)_E$ quedaría como:

$$\widetilde{Y}_t = \Phi_1\widetilde{Y}_{t-E} + e_t \tag{2.201}$$

para esto se tiene la varianza como:

$$\begin{aligned}\gamma_0 &= E[(\Phi_1\widetilde{Y}_{t-E} + e_t)^2] \\ &= \Phi_1^2 E[\widetilde{Y}^2_{t-E}] + 2\Phi_1 E[\widetilde{Y}_{t-E}e_t] + E[e_t^2]\end{aligned} \tag{2.202}$$

en donde, suponiendo estacionariedad para $\{\widetilde{Y}\}$, se obtiene

$$\gamma_0 = \frac{\sigma^2}{(1-\Phi_1^2)}$$

y de las autocovarianzas vienen dadas como:

$$\gamma_{iE} = \frac{\Phi_1^i\sigma^2}{(1-\Phi_1^2)} \quad para\ i \geq 0$$

mientras que:

$$\gamma_\kappa = 0 \quad para\ \kappa \neq iE$$

Entonces la FAC con correspondencia a 2.201 viene a ser

$$\rho(\kappa) = \begin{cases} \Phi^i & para\ \ \kappa = iE \quad con\ i = 0, 1, 2, \ldots \\ 0 & para\ \kappa \neq iE \end{cases} \tag{2.203}$$

Ahora analizando el modelo de promedios móviles estacionales tenemos:

$$\widetilde{Y}_t = e_t - \Theta_1 e_{t-E} - \Theta_2 e_{t-2E} - \ldots - \Theta_Q e_{t-QE} \tag{2.204}$$

donde Q es el múltiplo más grande de E que se encuentra en el modelo. Nuevamente, tomando el caso más sencillo el $MA(1)_E$ representado como:

$$\widetilde{Y}_t = e_t - \Theta_1 e_{t-E} \tag{2.205}$$

la cual tiene como la función de autocovarianza,

$$\begin{aligned} \gamma_0 &= (1+\Theta^2)\sigma^2 \\ \gamma_E &= -\Theta^2\sigma^2 \\ \gamma_\kappa &= 0 \quad para\ \kappa \neq E \end{aligned} \tag{2.206}$$

que, en términos de la FAC, implica

$$\rho(\kappa) = \begin{cases} \frac{-\Theta}{(1+\Theta^2)} & si\ \ \kappa = E \\ 0 & para\ \kappa \geq 1\ con\ \kappa \neq E \end{cases} \tag{2.207}$$

o sea, que existirá únicamente una correlación diferente de cero, correspondiente al retraso $\kappa = E$. Como consecuencia un modelo ARMA estacional se vera así

$$\widetilde{Y}_t = \Phi_1\widetilde{Y}_{t-E} + \Phi_2\widetilde{Y}_{t-2E} + ... + \Phi_P\widetilde{Y}_{t-PE} + e_t - \Theta_1 e_{t-E} - \Theta_2 e_{t-2E} - ... - \Theta_Q e_{t-QE} \tag{2.208}$$

Por último, es de mencionar que la estructura de la función de autocorrelación de un proceso estacional ARMA es análoga a la de un proceso no estacional, con correlaciones diferentes de cero sólo en los períodos $E, 2E, 3E....$ hay que notar que la suposición de este método es que la serie siempre es estacionaria cosa que muchas veces no pasa por tal motivo Box y Jenkins (1970) propusieron el siguiente modelo conocido como modelo multiplicativo estacional.

Modelo multiplicativo estacional

Este modelo permite estudiar series de tiempo con efectos estacionales y no estacionales. Este modelo tiene la forma

$$\Phi(L^E)\nabla_E^D(Y_t - \mu) = \Theta(L^E)\alpha_t \tag{2.209}$$

donde las variables $\{\alpha_t\}$ no se suponen ruido blanco, sino generada por un proceso ARIMA(p,d,q), es decir,

$$\phi(L)\nabla^d\alpha_t = \theta(L)e_t \tag{2.210}$$

con $\{e_t\}$ un proceso de ruido blanco de estas dos últimas expresiones se obtiene el modelo multiplicativo estacional:

$$\phi(L)\Phi(L^E)\nabla^d\nabla_E^D(Y_t-\mu)=\theta(L)\Theta(L^E)e_t \tag{2.211}$$

el cual es denotado como $ARIMA(p,d,q)x(P,D,Q)_E$[17].

Para finalizar podemos ver que con este tipo de modelos podemos modelar series que contengan tanto tendencia como estacionalidad debido a los operadores diferencia tanto normal como estacional respectivamente pero es de suma importancia decir que cuando una serie es completamente estacional se utiliza el operador diferencia estacional, pero que pasa cuando una serie tiene partes estacionales y otras no, pues para eso existen los otros parámetros P y D del modelo $ARIMA(p,d,q)x(P,D,Q)_E$ para agregarle más poder al modelo sin necesidad de diferenciar, y para saber el grado o cual parámetro tomar dependerá de como se ven los correlogramas y la gráfica de la serie para poder determinar si es conveniente meter estos parámetros al modelo o no.
A continuación veremos como identificar un modelo adecuado a las series de tiempo, y que por cierto no es una tarea fácil de hacer de hecho es una de las partes más complicadas al analizar series de tiempo.

2.6.3. Identificación del modelo

Hasta ahora hemos mostrado la teoría de las técnicas empleadas en series de tiempo, el como hacerlo, pero ahora nos falta que modelo aceptar y cuales son algunas reglas que deben de cumplir para poder decir que nuestro modelo es el adecuado.

Primeramente debemos saber si la serie es estacionaria, es decir, si el valor de la media no varía a través del tiempo o es no estacionaria. Para esto se recurre además de la examinación de la gráfica, el comportamiento del correlograma (véase la figura 2.7), en la figura podemos notar que cuando las autocorrelaciones decienden muy lentamente se tienen indicios de que existe

[17] Es importante aprenderse esta nomenclatura debido a que cuando uno usa un software para series de tiempo estacionales y utiliza un modelo ARIMA para modelarlas, siempre aparece en esta forma.

tendencia, hay que notar que no importa como esta el otro correlograma con que exista uno que presente este comportamiento con eso bastara.

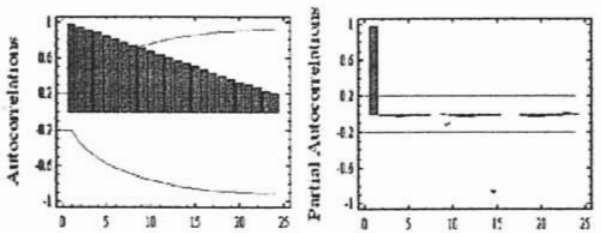


Figura 2.7: correlograma con tendencia

Si la serie no es estacionaria, se toman las diferencias de los datos originales concentrando el análisis en esta serie diferenciada, revisándose nuevamente el comportamiento de la gráfica de los datos y del correlograma y si existiera aún alguna tendencia se procederá a una nueva diferenciación[18]. Obsérvese que cada vez que se diferencia una serie se pierde un dato, aunque esto es irrelevante cuando el número total de datos es grande.

Una vez obtenida una serie estacionaria, debemos identificar la forma del modelo a utilizar. Este paso se logra mediante la comparación de los coeficientes de autocorrelación y de autocorrelación parcial de los datos a ajustar con las correspondientes distribuciones de los diversos modelos ARIMA.

Si trabajamos con una serie W_t, como la serie original y que denota a $\nabla^d Y_t$ entonces procedemos a lo siguiente:

1. Calcular las primeras κ autocorrelaciones muestrales y las autocorrelaciones parciales muestrales de W_t. Las 24 autocorrelaciones explican en buena medida el comportamiento de la serie en general.

2. Graficar las autocorrelaciones y autocorrelaciones parciales.

[18]En la practica es raro encontrarse con una serie que necesite ser diferenciada más de 3 veces, y además siempre hay que cuidar de no sobrediferenciar la serie

3. Asociar el comportamiento de los valores a las funciones de autocorrelación teórica y autocorrelación parcial teórica de un proceso ARMA(p,q).

Antes de continuar, hay que establecer cuáles autocorrelaciones son significativamente diferentes de cero. Para llevar este paso importante hay que advertir que la FAC muestral está afectada por variaciones muestrales, que desvirtúan la apariencia real de las autocorrelaciones; por este motivo se requiere de un criterio para distinguir lo verdadero de lo artificial. Dicho criterio lo proporcionó Bartlett(1946) al obtener expresiones aproximadas (estimaciones r_κ) para las varianzas y covarianzas de las autocorrelaciones muestrales, en caso de que el proceso sea generado a partir de un ruido blanco como distribución normal. Las cuales son:

$$Var(r_\kappa) = \frac{1}{N}\sum_{j=-\infty}^{\infty}(\rho_j^2 + \rho_{j+\kappa}\rho_{j-\kappa} - 4\rho_\kappa\rho_j\rho_{j-\kappa} + 2\rho_\kappa^2\rho_j^2)$$

$$Cov(r_\kappa, r_{\kappa+s}) = \frac{1}{N}\sum_{j=-\infty}^{\infty}\rho_j\rho_{j+s}$$

Arderson(1942) demostró que bajo la hipótesis de que las autocorrelaciones teóricas de $\rho(\kappa)$ sean cero, las estimaciones divididas entre su desviación estándar, se distribuyen aproximadamente a una función normal. Por tal motivo, los valores $\pm 1{,}96x$ (desviación estándar) constituyen los límites del intervalo que con un 95 % de confianza, asegura que las autocorrelaciones son cero.

Entonces Para un modelo MA(1), tenemos un posible correlograma que lo describe(véase la figura 2.8);

Podemos ver que en la función de autocorrelación solo hay una correlación que es significativamente grande y la función de autocorrelación parcial desciende a cero, por tal motivo el mejor modelo para este tipo de serie seria un MA(1), esto se extiende para los modelos MA(q) los cuales para toda k mayor a q la FAC serán cero o aproximadamente cero, es decir, no pasan del intervalo de confianza para decir que son estadísticamente insignificantes (esto en las imágenes se ve como las líneas horizontales y paralelas al eje de las abscisas).

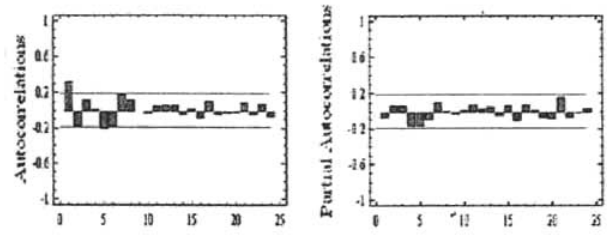


Figura 2.8: correlograma de la FAC y FACP para un modelo MA(1)

Ahora bien consideremos el caso de un AR(1) para ver su correlograma correspondiente tanto a la FAC como a la FACP (véase la figura 2.9);

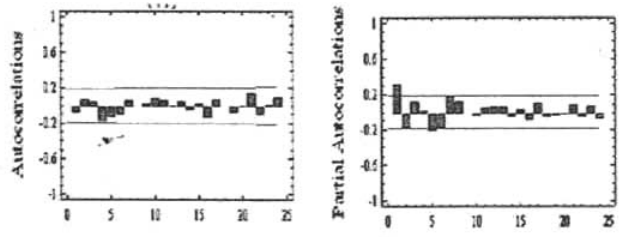


Figura 2.9: correlograma de la FAC y FACP para un modelo AR(1)

Como vemos ahora la correlación significativamente grande se encuentra en el correlograma correspondiente a la FACP y no en la FAC y esta última desciende a cero, de aquí podemos ver que para identificar un modelo AR(p) dependen del numero de correlaciones significativamente grandes en la FACP.

Finalmente, para un modelo ARMA(1,1) tenemos un posible correlograma para la FAC y la FACP (véase la figura 2.10).

Como podemos ver una vez que pasa el primer período las correlaciones van decayendo exponencialmente en ambos casos por tal razón es buena idea el utilizar una ARMA(1,1), en general este comportamiento se mantiene

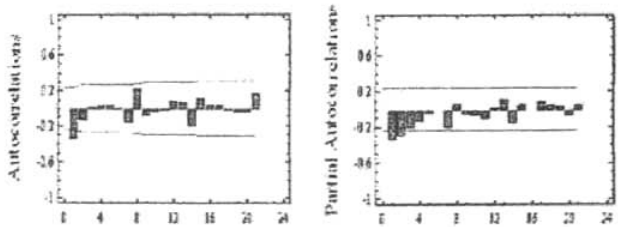


Figura 2.10: correlograma de la FAC y FACP para un modelo ARMA(1,1)

para un modelo ARMA(p,q), el cual dependiendo de las correlaciones más significativas dependerá el grado tanto de p como de q (al final de esté capítulo se dan algunos ejemplos de FAC y FACP teóricas para determinar el orden de p y q, y así poder saber sí se trata de un AR, MA o un ARMA (véase cuadros 2.1 y 2.2)).

Resumiendo lo que hemos visto hasta ahora en cuestión de identificar el modelo, Verificar si la serie es o no estacionaria, en caso de no serlo, aplicar el operador diferencia las veces que sean necesaria para que desaparezca[19] la tendencia una vez que tenemos una serie estacionaria, debemos identificar las autocorrelaciones que caen exponencialmente a cero. Si las autocorrelaciones descienden exponencialmente a cero, el proceso indicado es el AR; si son las autocorrelaciones parciales las que descienden a cero, entonces el proceso indicado es el MA; y si tanto los coeficientes de autocorrelación y autocorrelación parcial descienden a cero, el indicado es un proceso mixto ARIMA, se puede determinar el orden de los procesos AR *y/o* MA *contando* el numero de coeficientes de autocorrelación y de autocorrelación parcial que son diferentes de cero en forma significativa. La tabla 2.11 permite resumir las pautas de comportamiento más frecuentes, de la FAC y FACP en los modelos AR(p), MA(q) y ARMA(p,q)

Por lo que respecta a la identificación de los componentes estacionales, se sigue el mismo criterio que en la parte regular, pero fijándose en los retardos correspondientes a la periodicidad estacional (4,8,12,... para datos trimes-

[19]Hay que recordar que en raras ocasiones el grado de diferenciación excede a 2.

Modelo-Función	FAC	FACP
AR(P)	Decrecimiento rápido de tipo exponencial y/o sinusoidal.	Se anulan para retrasos superiores a p.
MA(q)	Se anula para retardos superiores a q.	Decrecimiento rápido de tipo exponencial y/o sinusoidal.
ARMA(p,q)	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales amortiguadas a partir de q.	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales amortiguadas a partir de p.

Figura 2.11: Resumen del comportamiento de las FAC y FACP

trales o 12,24,36,... para datos mensuales), además hay que recordar que el operador diferencia estacional solo se aplica cuando la serie tiene estacionalidad en toda la serie y no parcialmente dado que para este último caso se ocupan los modelos AR(P) y MA(Q) para el modelo estacional.

Cabe señalar que existen más formas que ayudan a escoger el modelo adecuado, como lo son los criterios de Akaike y Schwarz, los cuales proponen un estimador para ayudar a escoger el mejor modelo, pero nosotros solo utilizaremos los supuestos para los residuos y los tipos de error (vistos en la sección de modelos de regresión) para verificar que nuestro modelo es el más adecuado, pero en términos generales se escoge el modelo que tenga el numero

menor tanto para el criterio de Akaike como con el de Schwarz[20].

Verificación del modelo

Una vez que escogimos el modelo debemos verificar que dicho modelo no viole ciertos supuestos los cuales, utilizan a los residuos. Un residuo $\widehat{e}_t$, se define como la diferencia entre W_t y el pronóstico $\widehat{W}_t$, es decir, $\widehat{e}_t = W_t - \widehat{W}_t$ es el residuo al tiempo t. Cuando el tamaño de la muestra es grande, los errores aleatorios $\{e_t\}$ y los residuos $\widehat{e}_t$ son esencialmente iguales, por tal razón al analizar los residuos estaríamos analizando lo que deben cumplir los errores aleatorios. Entonces, con base en la teoría expuesta sabemos que los errores aleatorios $\{e_t\}$ deben ser una realización de un proceso de ruido blanco. Por ello, la verificación del modelo consistirá principalmente en comprobar si esto se cumple con las características de un ruido blanco, es decir, que su media sea cero y su varianza constante.

Primer supuesto $\mu = 0$

Para probar la hipótesis $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ se puede utilizar el hecho de que.

$$\left(m(\widehat{e}_t) - 1{,}96\sqrt{\frac{\sigma^2}{N}}, m(\widehat{e}_t) + 1{,}96\sqrt{\frac{\sigma^2}{N}}\right) \tag{2.212}$$

donde,

$$m(\widehat{e}) = \sum_{t=t'}^{N} \widehat{e}_t/(N - d - p) \quad con\ t' = d + p + 1 \tag{2.213}$$

y

$$\widehat{\sigma} = \sqrt{\sum_{t=t'}^{N} [\widehat{e}_t - m(\widehat{e})^2/(N - d - p - q)]} \tag{2.214}$$

y representan tanto la media aritmética y la desviación estándar muestral de los residuos respectivamente. Entonces 2.212 constituye un intervalo de

[20]Sí se desea ver más acerca de los criterios de Akaike y Schwarz véase Diebold Francis X. 71-76 pp [8] ó E.P. Box George, Gwilym M. Jenkins y Gregory C. Reinsel 200-202 pp [10].

confianza al 95 % de la media de una normal. Se rechaza la hipótesis nula si el cero no está dentro de este intervalo o bien, se rechaza la hipótesis nula si $|\frac{\sqrt{N'}m(\widehat{e})}{\widehat{\sigma}}| \geq 1{,}96$, donde N'=N-d-p. Por el contrario si $|\frac{\sqrt{N'}m(\widehat{e})}{\widehat{\sigma}}| < 1{,}96$ se dirá que no hay evidencia de que la media del proceso de ruido blanco sea distinta de cero y por lo mismo no se rechaza el supuesto.

Si al concluir la prueba se obtiene que la media de los residuos no es cero, puede ser debido a dos causas:

1. Existe una parte semideterminista en la serie que podría corregirse con una diferencia más.
2. El problema es determinista y se corrige añadiendo una constante μ al modelo que se calcula conjuntamente con los otros parámetros y cuyo valor inicial es $m(\widehat{e})$.

Segundo supuesto $Var(\widehat{e}) = \sigma^2$

Una vez cubierto este problema continuamos con revisar que la varianza de los residuos es constante, y consiste en hacer una gráfica de los residuos contra el tiempo para observar, visualmente, si la varianza parece o no ser constante. Esta verificación visual podría parecer un poco absurda pero la idea es que solamente las violaciones muy notorias de este supuesto son las que realmente llegan a causar problemas. Entonces sólo en caso de que la varianza parezca seguir algún patrón de crecimiento o de decrecimiento, es posible aplicar una transformación potencia[21] para estabilizar la varianza de la serie.

Tercer supuesto. No estén autocorrelacionados

Otra condición que deben de cumplir los residuos es que no estén autocorrelacionados, es decir, que las autocorrelaciones para $\kappa > 1$ sean cero, esto es

[21] La cual hicimos mención al final del modelo ARIMA para series con tendencia, la cual de la teoría de Bartlett (1946) nos da una posible transformación para estabilizar la varianza.

porque si existiese autocorrelación entre los residuos implicaría que se podría pronosticar los residuos y se estaría perdiendo información para el modelo, por tal razón es importante que no estén autocorrelacionados. Para ello debemos calcular la función de autocorrelación de los residuos, que en caso de tener media cero está dada por

$$r_e(\kappa) = \frac{\sum_{t=t'}^{N-\kappa} \widehat{e}_t \widehat{e}_{t+\kappa}}{\sum_{t=t'}^{N} \widehat{e}_t^2} \qquad \kappa = 1, 2, \ldots \tag{2.215}$$

entonces dado que para un proceso de ruido blanco $\rho(\kappa) = 0$ para $\kappa > 1$ las autocorrelaciones serán significativamente diferentes de cero si $|r_e(\kappa)| \geq \frac{1.96}{\sqrt{N-d-p}}$.

Box y Ljung (1978) encontraron que una prueba para verificar que los errores son ruido blanco, puede hacerse a través del estadístico Q'[22] dado por

$$Q' = (N-d-p)(N-d-p+2) \sum_{k=1}^{K} \frac{r_e^2(k)}{(N-d-p-k)} \tag{2.216}$$

Si $K > 20$, $Q' \sim \chi^2_{K-p-q}$, se dirá que el modelo es inadecuado si $Q' > \chi^2_{K-p-q}$.

Cuando las autocorrelaciones no sean las de un ruido blanco, hay que revisar las FAC de los residuos para tratar de identificar algún patrón conocido e incorporar los parámetros que el patrón determine al modelo original.

Cuarto supuesto. Tienen distribución normal, para toda t

La teoría de series de tiempo para efectuar pronósticos supone que los errores $\{e_t\}$ tienen distribución normal para toda t, entonces aproximadamente el 95 % de las observaciones deben de estar dentro de un intervalo que se extienda dos desviaciones estándar por debajo y por arriba de la media, entonces si se cumple que $\mu = 0$, se esperaría que a lo más un total de (N-d-p)/20 observaciones estarán fuera del intervalo $(-2\sigma, 2\sigma)$.

Para verificar este supuesto, a través de los residuos, existen dos posibles caminos.

[22] No hay que confundirlo con el estadístico Q creado por Box y Pierce (1970), debido a que Q' es una modificación al estadístico de Box y Pierce.

1. Graficar el histograma de los residuos y por inspección revisar que exista simetría en los mismos.

2. Graficar los residuos contra el tiempo(al igual que en el supuesto 2)

Es de advertir que este supuesto se debe de cumplir para los errores aleatorios (e_t), pero no tienen por qué ser satisfecho exactamente por los residuos ($\widehat{e}_t$), por tal razón, cabe esperar pequeñas violaciones que no causen problemas en absoluto. En caso de que se viole el supuesto notoriamente, podría aplicarse una transformación normalizadora como las que se hizo mención en el supuesto 2.

Quinto Supuesto. No existen observaciones aberrantes

Son aquellos residuos que están fuera del intervalo $(-3\sigma, 3\sigma)$, dado que esto implicaría, o bien sucedió un evento cuya probabilidad de ocurrencia es muy baja o el residuo en cuestión no fue generado por el mismo proceso generador del resto de la serie, entonces sería conveniente verificar las observaciones que dan lugar a esos residuos y verificar las razones, esto debido a que no es bueno quitar información sin antes haber hecho un análisis de las causas por las cuales paso eso, esto con el fin de no desechar información valiosa.

Es importante tener en cuenta estos supuestos[23] debido a que los estaremos utilizando en el capítulo 4 para escoger el mejor modelo que capte la mayor información. Por otro lado, se da la definición de un estadístico que utilizaremos más adelante y el cual es el **P-value**, el cual sirve como una medida de significancia, cuando el P-value es menor que .01, quiere decir que hay una relación estadísticamente significativa entre dos variables con un 99 % nivel de confianza, si el P-value es mayor o igual que .10 se dice que la relación entre las variables es estadísticamente no significativa con un nivel de confianza de 90 % y así el grado de significancia que se desee ,pero más adelante veremos lo importante de esté estadístico.

Por último damos algunas funciones teóricas de las funciones de autocorrelación y autocorrelación parcial para modelos AR , MA y ARMA (véase

[23]Si se desea ver otros supuestos ver Guerrero Guzmán Víctor Manuel, 2003 [11].

cuadros 2.1 y 2.2).

2.6.4. Pronósticos a través de modelos ARIMA

La última etapa de la construcción de un modelo es el pronóstico y de hecho es el fin que se pretende al analizar una serie de tiempo. Una vez que el modelo ha sido identificado se procede a la generación de un pronóstico de observaciones futuras que minimizando el error cuadrático medio, serán insesgadas y no correlacionadas. Nos limitaremos a los pronósticos para procesos estacionarios[24] debido a que dan la idea general, además que daremos por hecho que ya se utilizaron los métodos antes mencionados para poder hacer de una serie no estacionaria una serie estacionaria.

Sea $\{Y_t\}$ una serie estacionaria que admite una representación del tipo ARMA(p,q), es decir

$$\phi(L)Y_t = \theta(L)c_t \tag{2.217}$$

Ya que estamos suponiendo que es un proceso estacionario, se puede representar como

$$Y_t = \psi(L)c_t \tag{2.218}$$

donde

$$\begin{aligned} \psi(L) &= \frac{\theta(L)}{\phi(L)} \\ &= \sum_{j=0}^{\infty} \psi_j(L) \quad con\ \psi_0 = 1 \end{aligned} \tag{2.219}$$

sustituyendo esta última expresión en 2.218 tenemos

[24]Si se desea ver la teoría para pronósticos para series de tiempo estacionales véase E.P. Box George, Gwilym M. Jenkins y Gregory C. Reinsel cap 5 y 334-341pp [10].

$$Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} \tag{2.220}$$

donde para t+h

$$Y_{t+h} = \sum_{j=0}^{\infty} \psi_j e_{t+h-j} \tag{2.221}$$

supóngase que en el tiempo t se tienen las observaciones $Y_t, Y_{t-1}, \ldots$ y se quiere saber el valor de Y_{t+h} a partir de la combinación lineal de las observaciones mencionadas. Entonces,

$$\widehat{Y}_{t+h} = c_h e_t + c_{h+1} e_{t-1} + c_{h+2} e_{t-2} + \ldots \tag{2.222}$$

que es un pronóstico para Y_{t+h}.

Dado que existe un gran número de posibles combinaciones de valores $c_h, c_{h-1}, \ldots$ el criterio que se ocupa para decidir qué valores seleccionar es el error cuadrático medio, entonces el error para la ecuación 2.222 esta dada como:

$$E[Y_{t+h} - \widehat{Y}_{t+h}]^2 \tag{2.223}$$

Sustituyendo el valor de las expresiones y desarrollando el binomio tenemos:

$$\begin{aligned} E[Y_{t+h} - \widehat{Y}_{t+h}]^2 &= E\Big[\sum_{j=0}^{h-1} \psi_j^2 e_{t+h-j}^2 + \sum_{j=h}^{\infty} (\psi_j - c_j)^2 e_{t+h-j}^2 \\ &\quad + 2 \sum_{j=0, i \neq j}^{\infty} \psi_i (\psi_h - c_h) e_j e_i \Big] \\ &= \sum_{j=0}^{h-1} \psi_j^2 E[e_{t+h-j}^2] + \sum_{j=h}^{\infty} (\psi_j - c_j)^2 E[e_{t+h-j}^2] \\ &= \sigma^2 \Big[\sum_{j=0}^{h-1} \psi_j^2 + \sum_{j=h}^{\infty} (\psi_j - c_j)^2\Big] \end{aligned} \tag{2.224}$$

el mínimo de esta expresión se obtiene haciendo $\psi_j = c_j$. Por tanto.

$$\widehat{Y}_{t+h} = \psi_h e_t + \psi_{h+1} e_{t-1} + \psi_{h+2} e_{t-2} + \ldots \tag{2.225}$$

ya que las observaciones están hechas hasta t, tenemos

$$E[e_{t-j}/Y_t, Y_{t-1}, ...] = \begin{cases} 0 & para \quad j > 0 \\ e_{t-j} & para \quad j \leq 0 \end{cases} \tag{2.226}$$

Por otro lado utilizando la expresión 2.221

$$\begin{aligned} E[Z_{t+h}/Y_t, Y_{t-1}, ...] &= E[e_{t+h} + \psi_h e_{t+h-1} + \psi_h e_{t+h-2} + ... \\ &\quad + \psi_h e_t + \psi_{h+1} e_{t-1} + ...] \\ &= \psi_h e_t + \psi_{h+1} e_{t-1} + ... \end{aligned} \tag{2.227}$$

De aquí se concluye que el pronóstico que proporciona el mínimo error cuadrático medio para Y_{t+h} está dado por la esperanza condicional. Entonces el error del pronóstico esta dado como:

$$\begin{aligned} \widehat{e}_{t+h} &= Y_{t+h} - \widehat{Y}_{t+h} \\ &= \sum_{j=0}^{\infty} \psi_j e_{t+h-j} - \sum_{j=h}^{\infty} \psi_j e_{t+h-j} \\ &= \sum_{j=0}^{h-1} \psi_j e_{t+h-j} \end{aligned} \tag{2.228}$$

Calculando el valor esperado se tiene

$$\begin{aligned} E[\widehat{e}_{t+h}] &= E\Big[\sum_{j=0}^{h-1} \psi_j e_{t+h-j}\Big] \\ &= \sum_{j=0}^{h-1} \psi_j E[e_{t+h-j}] \\ &= 0 \end{aligned} \tag{2.229}$$

Con este último resultado se puede apreciar que los pronósticos son insesgados dado que

$$E[Y_{t+h} - \widehat{Y}_{t+h}] = 0 \tag{2.230}$$

y además implica que

$$E[Y_{t+h}] = E[\widehat{Y}_{t+h}] \tag{2.231}$$

La varianza esta dada por

$$Var(\widehat{e}_{t+h}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \tag{2.232}$$

Expresión que si analizamos un poco veremos que la varianza del pronóstico aumenta a medida que aumenta h.

Una vez dados los pronósticos, es conveniente dar un intervalo que represente la confiabilidad de los mismos. Para ello, bajo el supuesto de que $e_t \sim N(0, \sigma^2)$ para toda t y dadas las ecuaciones 2.229 y 2.232 tenemos

$$\widehat{e}_{t+h}/(Y_t, Y_{t-1}, ...) \sim N(0, Var(\widehat{e}_t)) \tag{2.233}$$

Está última expresión equivale a escribir

$$Y_{t+h} - \widehat{Y}_{t+h}/(Y_t, Y_{t-1}, ...) \sim N(0, Var(\widehat{e}_t)) \tag{2.234}$$

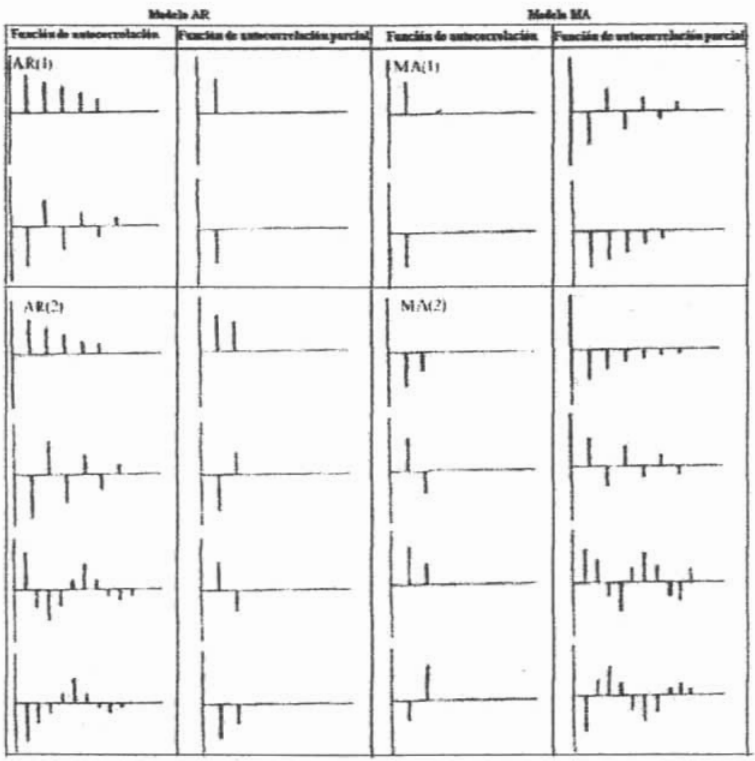
Por lo cual, un intervalo de confianza al $100(1-\alpha)\,\%$ para Y_{t+h}, condicionado por el conocimiento de las observaciones $Y_t, Y_{t-1}, Y_{t-2}, \ldots$ es

$$\widehat{Y}_{t+h} \pm z_{\frac{\alpha}{2}} \Big(\sum_{j=0}^{h-1} \psi_j^2 \Big)^{\frac{1}{2}} \sigma \tag{2.235}$$

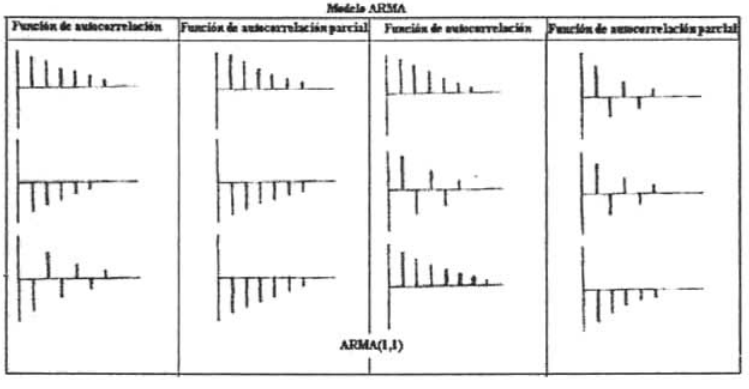
en donde $z_{\frac{\alpha}{2}}$ es el punto porcentual, de modo que $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ si $Z \sim N(0,1)$.

Es evidente que el uso del procedimiento anterior para obtener pronósticos[25] es poco práctico para hacerlo manualmente, pero no para emplearlo en una computadora(gracias a dios), cosa que se hace en la actualidad y que haremos nosotros con un software estadístico conocido como statgraphics versión 4.

[25] Si se está interesado en profundizar en estos temas véase Guerrero Guzmán Víctor Manuel, 2003 [11].



Cuadro 2.1: Funciones teóricas de autocorrelación y autocorrelación parcial para diferentes tipos de modelos AR y MA.



Cuadro 2.2: Funciones teóricas de autocorrelación y autocorrelación parcial para diferentes modelos de ARMA.

Capítulo 3

Redes Neuronales

3.1. Neurona biológica

El cerebro humano continuamente recibe señales (que llamaremos entrada(s)) de muchas fuentes y las procesa a manera de crear una apropiada respuesta (que llamaremos salida(s)). Nuestros cerebros cuentan con millones de neuronas (en un milímetro existen cerca de 50000 neuronas) que se interconectan para elaborar *Redes Neuronales Biológicas*. La neurona biológica (véase figura 3.1) está compuesta principalmente de cuatro partes:

- El *soma* o cuerpo de la neurona. Es la parte central de la neurona, en él se localiza el núcleo y es el encargado de realizar las actividades metabólicas de la neurona.
- *Dendritas*. Son ramificaciones que se originan en el cuerpo de la neurona, sirven para recibir señales provenientes de otra(s) neurona(s).
- *Axón*. Es una gran ramificación; está encargada de trasmitir las señales generadas por el cuerpo de la célula. A su vez, se ramifica varias veces hasta convertirse en pequeños filamentos.
- La *sinapsis*. Se encuentra en las terminales de los filamentos del axón. Es el encargado de trasmitir la respuesta de la neurona a la siguiente

a través de sustancias químicas llamadas neurotransmisores.

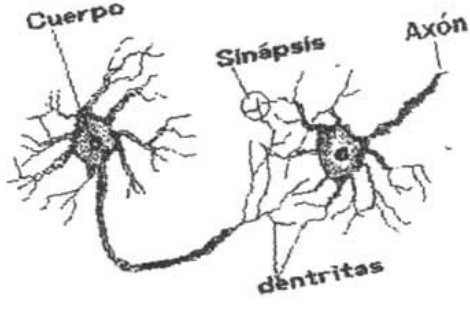


Figura 3.1: Esquema de la neurona biológica

3.2. Neurona artificial

Una neurona artificial es la unidad básica de procesamiento de información sobre la que se fundamenta la operación de una red neuronal artificial (RNA). De una manera más formal podemos decir que una neurona artificial se denomina procesador elemental, el cual es un dispositivo simple de cálculo que, a partir de un vector de entrada procedente del exterior o de otras neuronas, proporciona una única respuesta o salida. Los elementos que constituyen la neurona de etiqueta i son los siguientes (para entender mejor ver la figura 3.2).

- Conjunto de **entradas**, $x_j(t)$.

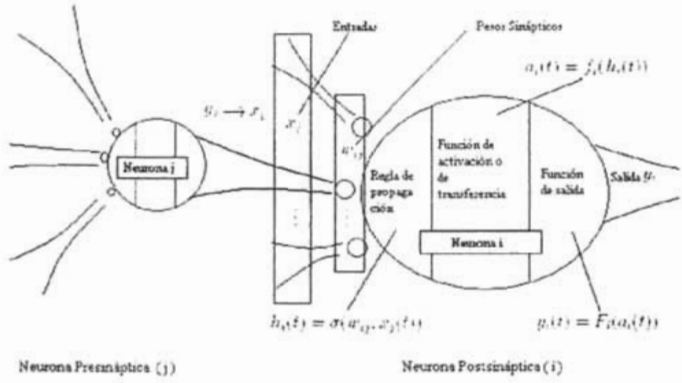


Figura 3.2: Modelo general de la Neurona artificial

- **Pesos sinápticos** de la neurona i, w_{ij} que representan la intensidad de interacción entre cada neurona presináptica j y la neurona postsináptica i.
- **Regla de propagación** $\sigma(w_{ij}, x_j(t))$ que proporciona el valor del potencial postsináptico $h_i(t) = \sigma(w_{ij}, x_j(t))$ de la neurona i en función de sus pesos y entradas.
- **Función de activación** $f_i(a_i(t-1), h_i(t))$, que proporciona el estado de activación actual $a_i(t) = f_i(a_i(t-1), h_i(t))$ de la neurona j, en función de su estado anterior $a_i(t-1)$ y de su potencial postsináptico actual.
- **Función de salida** $F_i(a_i(t))$, que proporciona la salida actual $y_i(t) = F_i(a_i(t))$ de la neurona i en función de su estado de activación.

De este modo, la operación de la neurona i puede expresarse como

$$y_i(t) = F_i\Big(f_i\Big[a_i(t-1), \sigma(w_{ij}, x_j(t))\Big]\Big) \tag{3.1}$$

Este modelo de neurona formal integra una serie de entradas y proporcionar cierta respuesta, que se propaga por el axón. A continuación detallaremos un poco más cada punto para su mayor entendimiento.

3.2.1. Entradas y Salidas

Las variables de entrada y salida pueden ser del tipo binario(0 ó 1) o continuas, dependiendo del modelo y aplicación. En ocasiones, el rango de los valores de la neurona de salida continua se suele limitar a un intervalo definido, por ejemplo, [0,1] o [-1,1].

3.2.2. Regla de propagación

La regla de propagación permite obtener, a partir de las entradas y los pesos, el valor del potencial postsináptico h_i de la neurona,

$$h_i(t) = \sigma(w_{ij}, x_j(t)) \tag{3.2}$$

la función más habitual es de tipo lineal, y se basa en la suma ponderada de las entradas con los pesos sinápticos.

$$h_i(t) = \sum_j w_{ij} x_j \tag{3.3}$$

que también se puede interpretar como el producto escalar de los vectores de entradas y pesos. El *peso sináptico* w_{ij} define en este caso la intensidad de interacción entre la neurona presináptica j y la postsináptica i. Dada una entrada positiva (procedente de un agente externo o simplemente la salida de otra neurona), si el peso es positivo tenderá a excitar a la neurona postsináptica, si el peso es negativo tenderá a inhibirla. Así se habla de sinapsis excitadoras (de peso positivo) e inhibidoras (de peso negativo).

3.2.3. Función de activación o de transferencia

La función de activación o transferencia proporciona el estado de activación actual $a_i(t)$ a partir del potencial postsináptico $h_i(t)$ y el propio estado de activación anterior $a_i(t-1)$

$$a_i(t) = f_i(a_i(t-1)), h_i(t)) \tag{3.4}$$

la función de activación se suele considerar determinista, y en la mayor parte de los modelos es monótona creciente y continua. La forma y=f(x) de las funciones de activación más empleadas en las RNA's las podemos ver en el cuadro 3.1 , y para poder entenderlo mejor, en ella designamos con x al potencial postsináptico ($h_i(t)$) y con y el estado de activación ($a_i(t)$), aquí observamos que la función más simple es la función identidad, otro caso muy simple también es la función escalón, empleada en el perceptrón simple que más adelante veremos. El objetivo de esta función (que generalmente presenta un comportamiento no lineal), es limitar la amplitud de la señal de salida dentro de un rango de valores normalizados (por ejemplo, dentro del intervalo [0,1] o [-1,1]).

En ocasiones los algoritmos de aprendizaje requieren que la función de activación cumpla la condición de ser derivable(como el Algoritmo "Backpropagation"que también analizaremos más adelante), las más empleadas en este sentido son las funciones de tipo sigmoideo.

3.2.4. Función de salida

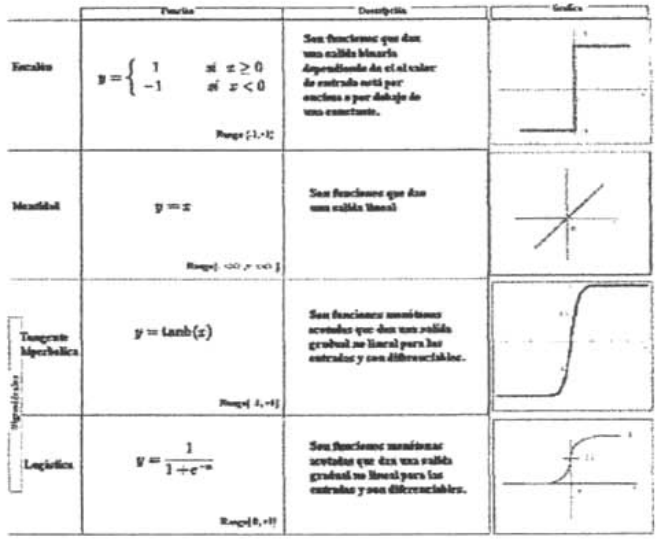
Esta función proporciona la salida global de la neurona $y_i(t)$ en función de su estado de activación actual $a_i(t)$, un ejemplo de esto es cuando la función de salida es la identidad F(x)=x, de modo que el estado de activación de la neurona se considera como la propia salida.

$$y_i(t) = F_i(a_i(t)) = a_i(t) \tag{3.5}$$

Como es lógico pensar las funciones más frecuentes para la función de activación lo son también para la función de salida.

Para resumir todo lo anterior y para el fin de este trabajo, en el modelo(red) que se utiliza para fines de predicción la función de salida es simplemente la función identidad o función lineal, esto implica que la función de salida $F(\cdot)$ pasa a ser la función de activación o transferencia $f(\cdot)$, entonces nuestra neurona resultante se muestra en la figura 3.3 la cual nos muestra una forma abreviada de una neurona artificial.

En varios casos existe otro elemento el cual se denomina umbral (θ), para determinar la activación o inhibición de la neurona.



	Función	Descripción	Gráfica
Escalón	$y = \begin{cases} 1 & si \ x \geq 0 \\ -1 & si \ x < 0 \end{cases}$ Rango:{-1,+1}	Son funciones que dan una salida binaria dependiendo de si el valor de entrada está por encima o por debajo de una constante.	
Identidad	$y = x$ Rango:(-∞, +∞)	Son funciones que dan una salida lineal	
Sigmoideas: Tangente hiperbólica	$y = \tanh(x)$ Rango:(-1,+1)	Son funciones monótonas acotadas que dan una salida gradual no lineal para las entradas y son diferenciables.	
Sigmoideas: Logística	$y = \frac{1}{1+e^{-x}}$ Rango:(0,+1)	Son funciones monótonas acotadas que dan una salida gradual no lineal para las entradas y son diferenciables.	

Cuadro 3.1: Funciones de activación y de salida habituales

3.3. Red Neuronal Artificial

Una RNA puede verse como un conjunto de neuronas interconectadas, de tal manera que la salida de una neurona generalmente sirve como entrada de otras neuronas. Las RNA's están compuestas de un gran número de elementos de procesamiento altamente interconectados (Neuronas) trabajando al mismo tiempo para la solución de problemas específicos. En cualquier caso, se trata de una nueva forma de computación que es capaz de manejar las imprecisiones e incertidumbres que aparecen cuando se trata de resolver problemas relacionados con el mundo real (reconocimiento de formas, toma de decisiones, predicciones, reconocimiento del habla y de caracteres, entre otros). Por último las RNA's se pueden clasificar gracias a ciertas características como son: el numero de capas de la red (monocapa o multicapa), de la

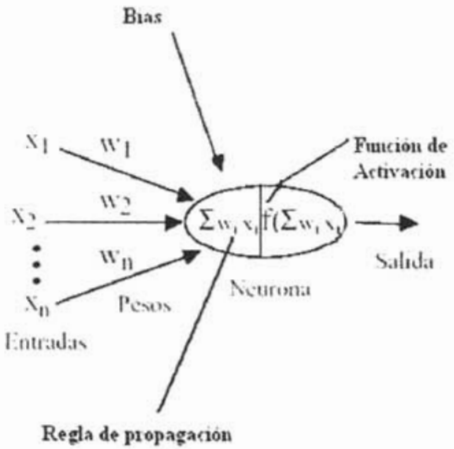


Figura 3.3: Esquema básico de una neurona artificial

forma en como cambia sus pesos, que es gracias a una regla de aprendizaje entre otros que veremos más adelante.

3.4. Ventajas que ofrecen las redes neuronales

Debido a su constitución y a sus fundamentos, las redes neuronales artificiales son capaces de aprender de la experiencia, de generalizar casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas y que este tipo de tecnología se esté aplicando en múltiples áreas. Entre las ventajas se incluyen:

Aprendizaje Adaptativo: capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial.

Auto-organización: una red neuronal puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje.

Tolerancia a fallos: la destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo un gran daño.

Operación en tiempo real: los cómputos neuronales pueden ser realizados en paralelo; para esto se diseñan y fabrican máquinas con hardware especial para obtener esta capacidad.

Aprendizaje adaptativo

La capacidad de aprendizaje adaptativo es una de las características más atractivas de las redes neuronales. Esto es, aprenden a llevar a cabo ciertas tareas mediante un entrenamiento con ejemplos ilustrativos. Entonces debido a que las redes neuronales pueden aprender a diferenciar patrones mediante ejemplos y entrenamientos, no es necesario elaborar modelos a priori ni necesidad de especificar funciones de distribución de probabilidad. Las redes neuronales son sistemas dinámicos autoadaptativos. Son adaptables debido a la capacidad de autoajuste de los elementos procesales (neuronas) que componen el sistema. Son dinámicos, pues son capaces de estar constantemente cambiando para adaptarse a las nuevas condiciones. En el proceso de aprendizaje, los enlaces ponderados de las neuronas se ajustan de manera que se obtengan ciertos resultados específicos. Una red neuronal no necesita un algoritmo[1] para resolver un problema, ya que ella puede generar su propia distribución de pesos en los enlaces mediante el aprendizaje. También existen redes que continúan aprendiendo a lo largo de su vida, después de completado su período de entrenamiento. La función del diseñador es únicamente la obtención de la arquitectura apropiada. No es problema del diseñador el cómo la red aprenderá a discriminar. Sin embargo, sí es necesario que desarrolle un buen algoritmo de aprendizaje que le proporcione a la red la capacidad de discriminar, mediante un entrenamiento con patrones (observaciones).

[1] Un algoritmo, por definición, define una función recursiva, por tal razón las RNA no pueden ser algorítmicas.

Auto-organización

Las redes neuronales emplean su capacidad de aprendizaje adaptativo para auto-organizar la información que reciben durante el aprendizaje y/o la operación. Mientras que el aprendizaje es la modificación de cada elemento procesal, la auto-organización consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico. Cuando las redes neuronales se usan para reconocer ciertas clases de patrones, ellas auto-organizan la información usada. Por ejemplo, la red llamada backpropagation, creará su propia representación característica, mediante la cual puede reconocer ciertos patrones. Esta auto-organización provoca la generalización: facultad de las redes neuronales de responder apropiadamente cuando se les presentan datos o situaciones a las que no había sido expuesta anteriormente. El sistema puede generalizar la entrada para obtener una respuesta. Esta característica es muy importante cuando se tiene que solucionar problemas en los cuales la información de entrada no es muy clara; además permite que el sistema dé una solución, incluso cuando la información de entrada está especificada de forma incompleta.

Tolerancia a fallos

Las redes neuronales fueron los primeros métodos computacionales con la capacidad inherente de tolerancia a fallos. Comparados con los sistemas computacionales tradicionales, los cuales pierden su funcionalidad cuando sufren un pequeño error de memoria, en las redes neuronales, si se produce un fallo en un número no muy grande de neuronas y aunque el comportamiento del sistema se ve influenciado, no sufre una caída repentina. Hay dos aspectos distintos respecto a la tolerancia a fallos:

- Las redes pueden aprender a reconocer patrones con ruido, distorsionados o incompletos. Esta es una tolerancia a fallos respecto a los datos.
- Las redes pueden seguir realizando su función (con cierta degradación) aunque se destruya parte de la red.

La razón por la que las redes neuronales son tolerantes a los fallos es que tienen su información distribuida en las conexiones entre neuronas, existiendo cierto grado de redundancia en este tipo de almacenamiento. La mayoría de los ordenadores algorítmicos y sistemas de recuperación de datos almacenan cada pieza de información en un espacio único, localizado y direccionable. En cambio, las redes neuronales almacenan información no localizada. Por lo tanto, la mayoría de las interconexiones entre los nodos de la red tendrán sus valores en función de los estímulos recibidos, y se generará un patrón de salida que represente la información almacenada.

Operación en tiempo real

Una de las mayores prioridades, casi en la totalidad de las áreas de aplicación, es la necesidad de realizar procesos con datos de forma muy rápida. Las redes neuronales se adaptan bien a esto debido a su implementación paralela. Para que la mayoría de las redes puedan operar en un entorno de tiempo real, la necesidad de cambio en los pesos de las conexiones o entrenamiento es mínimo.

Los inconvenientes que presentan las redes neuronales son: Realizan un complejo procesamiento que involucra millones de operaciones, por lo que es imposible darle seguimiento al razonamiento que han seguido.

3.5. Elementos de una Red Neuronal Artificial

Generalmente se pueden encontrar tres tipos de neuronas: de entrada, ocultas o intermedias y de salida. Aquellas que reciben estímulos externos, que tomarán la información de entrada. Dicha información se transmite a ciertos elementos internos que se ocupan de su procesamiento. Es en las sinapsis y neuronas correspondientes a este segundo nivel donde se genera cualquier tipo de representación interna de información. Como no tienen relación directa con la información de entrada ni con la salida, estos elementos se denominan unidades ocultas. Una vez finalizado el período de procesado, la información

llega a las unidades de salida, cuya misión es dar la respuesta al sistema. En la figura 3.4 se puede ver, un esquema de una red neuronal:

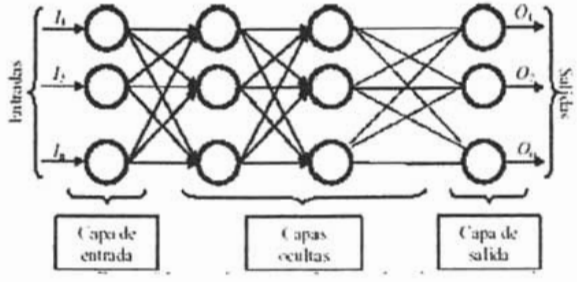


Figura 3.4: Red Neuronal

Los datos ingresan por medio de la "*capa de entrada*", pasan a través de la "*capa oculta*", después salen por la "*capa de salida*". Cabe mencionar que la capa oculta puede estar constituida por varias capas.

3.6. Topología de las Redes Neuronales Artificiales (RNA's)

Consiste en la organización de las neuronas en la red formando capas o agrupaciones de neuronas más o menos alejadas de la entrada y salida de la red. Los parámetros fundamentales de la red son: el número de capas, el número de neuronas por capa, el grado de conectividad y el tipo de conexiones entre neuronas. En términos topológicos podemos clasificar las redes entre: redes de una sola capa y las redes con múltiples capas.

3.6.1. Redes monocapa

Las redes Monocapa sólo cuentan con una capa de neuronas, que intercambian señales con el exterior y que constituyen a un tiempo la entrada y salida del sistema. En las redes Monocapa (red de *Hopfield* o red *Brain-State-in-Box*, *máquina de Boltzman*, *máquina de Cauchy*), se establecen conexiones laterales entre las neuronas, pudiendo existir también conexiones auto recurrentes (la salida de una neurona se conecta con su propia entrada), como en el caso del modelo *Brain-State-in Box*.

3.6.2. Redes Multicapa

Son aquellas que disponen de neuronas agrupadas en varias capas. Una forma para distinguir la capa a la que pertenece una neurona consistiría en observar el origen de las señales que recibe a la entrada y el destino de la señal de salida, las cuales están constituidas por una capa de entrada, una o varias capas ocultas y una capa de salida.

3.6.3. Formas de Conexión entre neuronas

Normalmente, todas las neuronas de una capa reciben señales de entrada de otra capa anterior, más cercana a la entrada de la red, y envían su señal de salida a una capa posterior, más cercana a la salida de la red. A estas conexiones se les denominan conexiones hacia delante o *feedforward*. Sin embargo en un gran número de estas redes también existe la posibilidad de conectar las salidas de las neuronas de capas posteriores a las entradas de capas anteriores, a estas conexiones se les denomina conexiones hacia atrás o *feedback*. Estas dos posibilidades permiten distinguir entre dos tipos de redes: las redes con conexiones hacia adelante (redes feedforward), y las redes que disponen de conexiones tanto hacia delante como hacia atrás (redes feedforward/feedback). En la figura 3.5 se muestran ejemplos de conexiones.

a) Conexiones hacia adelante

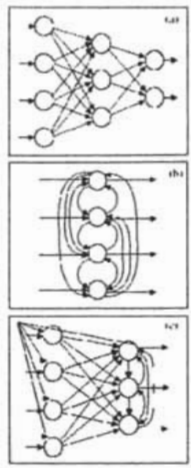


Figura 3.5: Diferentes tipos de RNA

b) Conexiones Laterales

c) Conexiones hacia atrás

3.7. Mecanismo de Aprendizaje

Es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. Los cambios que se producen durante el proceso de aprendizaje se reducen a destrucción, modificación y creación de conexiones entre las neuronas. La creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero; una conexión se destruye cuando su peso pasa a ser cero. El proceso de aprendizaje ha terminado (la red ha aprendido) cuando los valores de los pesos permanecen estables .

Un aspecto importante respecto al aprendizaje es conocer cómo se modifican los valores de los pesos; cuáles son los criterios para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información. Estos criterios determinan la regla de aprendizaje. Se suelen considerar dos tipos de reglas, las que responden a lo que se conoce como ***aprendizaje supervisado*** y ***aprendizaje no supervisado.*** La diferencia fundamental entre ambos tipos es la existencia o no de un agente externo (supervisor o maestro) que controle el proceso de aprendizaje.

3.7.1. Aprendizaje supervisado

El proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor o maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor comprueba la salida de la red y en caso de que ésta no coincida con la deseada, se procederá a modificar los pesos de las conexiones, con el fin de que la salida obtenida se aproxime a la deseada. Se suelen considerar tres formas de llevar a cabo el aprendizaje:

- Aprendizaje por corrección de error
- Aprendizaje por refuerzo
- Aprendizaje estocástico.

Una definición más formal sería. Sea E[W] un funcional que representa el error esperado de la operación de la red, expresado en función de pesos sinápticos W. En este tipo de aprendizaje se pretende estimar una cierta función multivariable desconocida $f : \Re^n \rightarrow \Re^m$ (la que representa la red neuronal) a partir de muestras (x,y),($x\epsilon\Re^n, y\epsilon\Re^m$) tomadas aleatoriamente, por medio de la minimización iterativa de E[W]. En nuestro caso, en el que deseamos usar una red neuronal para poder pronosticar valores futuros de una serie, se empleará el mecanismo de aprendizaje supervisado[2] debido a que le tenemos que dar ejemplos o valores deseados para que la red se ajuste a ellos.

[2]Si se esta interesado en estudiar estos temas véase Hilera José R. y Víctor J. Martínez, 1995 [15].

3.7.2. Aprendizaje no supervisado

Las redes con dicho aprendizaje no requieren de influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta de una entrada es o no correcta. Suele decirse que estas redes son capaces de autoorganizarse y como consecuencia la red debe ser capaz de reconocer regularidades en el conjunto de entradas, extraer rasgos, o agrupar patrones según su similitud. Un ejemplo de este tipo es el aprendizaje competitivo.

3.7.3. Reglas de aprendizaje

Una regla de aprendizaje define la manera en que los pesos deben de ser modificados en la red neuronal. Mediante estos cambios es que ésta se adapta paulatinamente durante el proceso de entrenamiento, y produce la mayor parte de las veces los resultados deseados. Entonces, se podría decir que el objetivo del aprendizaje de una RNA consiste esencialmente en encontrar los pesos correctos que producen la funcionalidad deseada. De aquí que se hayan inventado diversas reglas para lograr los distintos objetivos con mayor facilidad, entre las que se pueden destacar las siguientes:

- *Regla de aprendizaje de Hebb.* Derivada del postulado de Hebb sobre aprendizaje, la regla consiste en incrementar el factor de conexión entre dos neuronas que participan en la misma actividad[3].
- *Regla para aprendizaje competitivo.* En este tipo de regla, las neuronas compiten entre ellas mismas para disparar. Por lo tanto, en cualquier momento sólo la neurona ganadora está activa y es la única que consigue tener sus pesos actualizados. Este fenómeno se conoce como el ganador toma todo. En esta regla, se modifican los pesos en forma proporcional a la diferencia entre el patrón entrada y la suma de los pesos de la neurona ganadora. Con esto se consigue que el patrón almacenado en al neurona ganadora (los pesos) se acerque al patrón de entrada

[3]Si se desea ver más a fondo esta regla consultar Martín del Brío Bonifacio y Alfredo Sanz Molina, 1997, pp 44-45 [26].

- *Regla de aprendizaje de Boltzmann.* Con base en principios de mecánica estadística, una RNA tiene un cierto nivel de energía proporcional a sus pesos; empezando con altas temperaturas, conforme va transcurriendo el entrenamiento, la temperatura va disminuyendo. Permitiendo a la red alcanzar un equilibrio en algún momento. El aprendizaje consiste en reforzar los pesos de la red cuando existe alta probabilidad de que éstos participen en la generación de un patrón de salida.

- *Reglas de corrección por error.* En el aprendizaje supervisado, la RNA es entrenada para producir los valores de salida que corresponden a los patrones de entrada. El error o diferencia entre el valor correcto y el conseguido por la red se utiliza para modificar los pesos y así disminuir gradualmente el error.

Para ir introduciendo en la red que utilizaremos podemos ver que la regla de corrección por error es la más adecuada y gracias a Widrow y Hoff (1960) los cuales definieron una función que permite cuantificar el error global cometido en cualquier momento durante el proceso de entrenamiento de la red, lo cual es importante, ya que cuanto más conocimiento se tenga sobre el error cometido, más rápido se podrá aprender. Además, otro algoritmo de aprendizaje por corrección de error lo constituye la denominada "regla delta generalizada" (que veremos más adelante) o algoritmo "backpropagation" (propagación del error hacia atrás), se trata de una generalización de la regla delta para poder aplicarla a redes con conexiones hacia adelante (feedforward) con capas ocultas, esto es, son redes con capa de entrada, capas ocultas y capa de salida.

3.8. El perceptrón simple

Este modelo fue introducido por Rosenblatt a finales de los años cincuenta, es el modelo más sencillo de todas las redes y la base de partida para la construcción de muchas otras, el perceptrón simple es un modelo unidireccional(feedforward), compuesto por dos capas de neuronas, una de entradas y otra de salida. La operación de una red con n neuronas de entrada y m de

salida, se puede expresar como:

$$y_i(t) = f(\sum_{j=1}^{n} w_{ij}x_j - \theta_i), \quad \forall i, \quad 1 \leq i \leq m \tag{3.6}$$

Donde θ es el umbral[4].

Las neuronas de entrada no realizan ningún calculo, únicamente envían la información a las neuronas de salida, por ejemplo en el caso de que fuese solo una neurona de salida y dos de entrada, la única neurona de salida realiza la suma ponderada de las entradas (regla de propagación), resta el umbral y pasa el resultado a una función de transferencia que es de tipo escalón. La regla de decisión es responder +1 si el patrón presentado pertenece a la clase A y con -1 si el patron pertenece a la clase B. La salida dependerá de la entrada neta y el valor umbral.

Una técnica utilizada para analizar el comportamiento de redes como el perceptrón es representar en un mapa las regiones de decisión creadas en el espacio multidimensional de entradas de la red. En estas regiones se visualiza que patrones pertenecen a una clase y cuáles a otra. El perceptrón separa las regiones por un hiperplano cuya ecuación queda determinada por los pesos de las conexiones y el valor umbral de la función de activación de la neurona.

Por tanto, pese a su gran interés, el perceptrón presenta serias limitaciones, pues solamente puede representar funciones linealmente separables (Aquellas que solo pueden ser separadas por una linea recta). Así, aunque pueda aprender automáticamente a representar complejas funciones booleanas[5] o resolver con éxito muchos problemas de clasificación mediante su algoritmo de aprendizaje. Un ejemplo[6] de una función la cual no se puede tratar con un

[4]Se aplica este valor dado que normalmente la función de activación no está centrada en el origen del eje que representa el valor de la entrada neta (suma de las entradas x_i ponderadas), sino que existe cierto desplazamiento debido a las características de la propia neurona y no es igual en todas ellas.

[5]Son expresiones formadas por variables binarias y cuyo rango puede ser 0 ó 1, si se desea informarse sobre este tema ver Mano M. Morris, 1982, 45-46pp [25].

[6]Si se desea ver un ejemplo paso a paso de un perceptrón simple con su respectiva regla de aprendizaje para una función OR véase Hilera Jose R. y Víctor J. Martínez, 1995, 107-110 pp [15].

perceptrón simple es la función XOR[7]. Por tal razón se construyeron redes más poderosas, lo cual se logro al agregarles una capa (o más). A este tipo de redes se les llama perceptrón multicapa (o multinivel), el cual veremos más adelante.

3.8.1. Regla de aprendizaje del perceptrón simple

A continuación mostramos el algoritmo de aprendizaje del perceptrón simple, que tiene su aprendizaje supervisado y su regla es de corrección por error. Entonces, los pasos son los siguientes:

1. **Inicialización de los pesos y del umbral.** Inicialmente se asignan valores aleatorios a cada uno de los pesos (w_i) y al umbral (θ).

2. **Presentación de un nuevo par (Entrada, Salida esperada).** Consiste en presentar un nuevo patrón de entrada $X_p = (x_1, x_2, ..., x_N)$ junto con la salida esperada d(t).

3. **Cálculo de la salida actual**

$$y(t) = f\Big[\sum_{i=1}^{n} w_i(t)x_i(t) - \theta\Big] \tag{3.7}$$

   Siendo f(x) la función de transferencia escalón.

4. **Adaptación de los pesos.**

$$w_i(t+1) = w_i(t) + \alpha[d(t) - y(t)]x_i(t), \qquad 0 \leq i \leq N \tag{3.8}$$

   Donde d(t) es la salida esperada , y será 1 si el patrón pertenece a la clase A y -1 si pertenece a la clase B. α representa un factor de aprendizaje en el rango de [0,1]. Este factor debe ser ajustado de forma que satisfaga tanto los requerimientos de aprendizaje rápido como la estabilidad de las estimaciones de los pesos. Este proceso se repite hasta

[7]función OX-exclusiva se denota por $\bigoplus$ y es una operación binaria de la forma, $X \bigoplus Y = X\overline{Y} + \overline{X}Y$ si se desea saber más de esta función ver Mano M. Morris, 1982, 148-152pp [25].

que el error que se produzca para cada uno de los patrones sea cero o menor que un valor preestablecido, es decir, la diferencia entre el valor de la salida deseada y el valor obtenido sea mínima.

5. **Regresar al paso 2.** Repetir los pasos 2-4 hasta que la respuesta sea satisfactoria (el error entre la salida esperada y la obtenida sea mínima), o al alcanzar un numero determinado de ciclos.

3.9. Red Backpropagation o BP

Si añadimos capas ocultas a un perceptrón simple, obtendremos un perceptrón multicapa (véase la figura 3.6). El perceptrón multicapa (Rumelhart 1986) es el exponente más típico de las RNA's con aprendizaje supervisado. El entrenamiento de este tipo de redes se basa en la presentación sucesiva y de forma reiterada de pares de vectores en las capas de entrada y salida. La red crea un modelo a fin de ajustar sus pesos en función de los vectores de entrenamiento (o ejemplos) de forma que a medida que se pasan estos patrones para cada vector de entrada, la red producirá un valor de salida más similar al vector de salida esperado. Esta red suele entrenarse con el algoritmo denominado backpropagation de errores, motivo por el cual a una arquitectura del tipo perceptrón multicapa con aprendizaje backpropagation suele denominarse **Red backpropagation** o simplemente BP (backpropagation).

Se ha demostrado que un perceptrón multicapa de cuatro capas (1 de entradas, 2 capas ocultas y una de salida) puede aproximar con un grado de exactitud dado cualquier conjunto de funciones. Cuando las funciones son continuas, es suficiente una única capa oculta, además de que muchas veces una red con una capa oculta puede generar mejores resultados que otra con dos o más capas, ya que cuanto menor sea la variación, mayor es la capacidad de generalización de la red, y debido a que la red backpropagation solicita que las funciones de las capas ocultas sean continuas (como ya se dijo las más comunes son la tanh y la logística) se utilizaran redes con una sola capa oculta. A continuación, mostramos el teorema de aproximación universal que formaliza este resultado.

Teorema 3.1 (Funahashi 1989). . *Sea f(x) una función no constante, acotada y monótona creciente. Sea K un subconjunto compacto (acotado y*

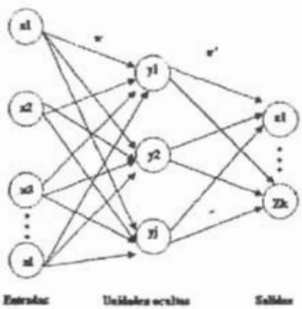


Figura 3.6: Perceptrón multicapa de tres capas

cerrado) de $\Re^n$. Sea un número real $\epsilon\Re$, y sea un entero $k\epsilon Z$, tal que $k \geq 3$, que fijamos. En estas condiciones se tiene que:

Cualquier mapeo $g : x\epsilon K \rightarrow (g_1(x), g_2(x), ..., g_m(x))\epsilon\Re^m$, con $g_i(x)$ sumables en K, puede ser aproximado en el sentido de la topología L_2 en K por el mapeo entrada-salida representado por una red neuronal unidireccional (perceptrón multicapa) de k capas (k-2 ocultas), con f(x) como función de transferencia de las neuronas ocultas, y funciones lineales para las capas de entrada y salida.

hay que observar que si k=3 entonces se obtiene un perceptrón de tres capas (1 de entradas, 1 capa oculta y una de salida).

Con esto resolvemos parcialmente el problema de la estructura de la red para un problema dado. Ahora, la cuestión se reduce a elegir un número apropiado de neuronas ocultas para ajustar el modelo evitando el problema de sobreajuste (que veremos más adelante). La solución a esta cuestión se logra a prueba y error en la mayoría de los casos.

Entonces como en el perceptrón simple y basándonos en la figura 3.6 denominaremos a x_i a las entradas de la red, y_j a las salidas de la capa oculta y z_k a las de la capa final; d_k serán las salidas deseadas. Por otro lado w_{ji} son los pesos de la capa oculta y θ_j sus umbrales, w'_{kj} los pesos de la capa de

salida y θ'_k sus umbrales. La operación de un perceptrón multicapa con una capa oculta(esto por el teorema de Funahashi) y neuronas de salida lineal se expresa matemáticamente de la siguiente manera:

$$z_k = \sum_j w'_{kj} y_j - \theta'_i = \sum_j w'_{kj} f(\sum_i w_{ji} x_i - \theta_j) - \theta'_i \qquad (3.9)$$

Siendo $f(\cdot)$ de tipo sigmoideo (como se menciono en las funciones de activación) siendo las más comunes la tanh y la logística (como se muestra en la figura 3.1.) y como ya sabemos proporcionan un rango entre [-1,1] y [0,1], respectivamente.

3.9.1. Algoritmo Backpropagation o regla delta generalizada

El procedimiento general[8] del algoritmo backpropagation es el siguiente:

Paso 1

Inicializar los pesos de la red con los valores aleatorios de preferencia entre el intervalo [0,1].

Paso 2

Presentar un patrón (puede ser un vector de observaciones) de entrada X_p : $x_{p1}, x_{p2}, ..., x_{pN}$ con sus respectivas salidas deseadas que debe generar la red; $d_1, d_2, ..., d_M$.

[8]Si se desea ver el desarrollo matemático véase Olmeda Ignacio y Sergio Barta Romero, 1993, 73-76pp [29] o Corchado Juan Manuel, 2000, 92-99pp [7].

Paso 3

Calcular la salida actual de la red, para ello presentamos las entradas a la red y vamos calculando la salida que presenta cada capa hasta llegar a la capa de salida ésta será la salida de la red $Z_1, Z_2, ..., Z_M$. Los pasos son los siguientes:

- Se calculan las entradas netas (o reglas de propagación) para las neuronas ocultas procedentes de las neuronas de entrada. Para una neurona j oculta tenemos:

$$y_{pj}^h = \sum_{i=1}^{N} w_{ji}^h x_{pi} + \theta_j^h$$

en donde el índice h se refiere a el número de la capa oculta, el subíndice p, al p-ésimo vector de entrenamiento, y j a la j-ésima neurona oculta. El término θ puede ser opcional, pues actúa como una entrada más, por tal motivo lo omitiremos.

- Se calculan las salidas de las neuronas ocultas

$$Y_{pj} = f_j^h(y_{pj}^h)$$

Siendo $f(\cdot)$ una función[9] diferenciable y monótona creciente (se recomiendan las funciones sigmoidales ya mencionadas).

- Una vez hecho esto se realizan los mismo cálculos para obtener la salida de las neuronas de salida.

$$z_{pk}^s = \sum_{j=1}^{L} w_{kj}^s Y_{pj}$$

y

$$Z_{pk}^s = f_k^s(z_{pk}^s)$$

donde L denota el número de neuronas en la capa oculta y k es el número de neuronas en la capa de salida

[9]Cuando la función de activación es lineal se le llama también como regla delta aunque solo en el caso de dos capas, (perceptrón, la red ADALINE entre otros) y cuando se utilizan más de dos capas (perceptrón multicapa) con funciones de activación no lineales se denomina regla delta generalizada. Si se desea profundizar sobre la regla delta ver Olmeda Ignacio y Sergio Barta Romero, 1993, 72-73pp [29].

Paso 4

Calcular el error de las capas de salida. Si la neurona k es una neurona de la capa de salida el valor delta se define como:

$$\delta_{pk}^{s} = (d_{pk} - Z_{pk}) f_k^{s'}(z_{pk}^{s})$$

Siendo $f'(\cdot)$ la derivada parcial respecto de z_{pk}^{s}

- Calcular el error de las capas ocultas

$$\delta_{pj}^{h} = f_j^{h'}(y_{pj}^{h}) \sum_{kj}^{M} w_{kj} \delta_{pk}^{s}$$

En este punto se manifiesta la idea fundamental que yace debajo del método backpropagation. Este método se basa en distribuir el error (delta) apreciado en cada una de las neuronas de salida hacia todas las neuronas ocultas en proporción de como dichas neuronas contribuyen en la salida.

Paso 5

Actualización de los pesos. Comenzando por las neuronas de salida y trabajando hacia atrás hasta llegar a la capa de entrada, ajustando los pesos de la manera siguiente:

- Para actualizar los pesos en la capa de salida

$$w_{kj}^{s}(t+1) = w_{kj}^{s}(t) + \alpha \delta_{pk}^{s} Y_{pj'} + \mu\big(w_{kj}^{s}(t) - w_{kj}^{s}(t-1)\big)$$

  donde α es el factor (o tasa) de aprendizaje y μ, el momento, pero esto dos términos los veremos más adelante, por último t es el número de iteración.

- Para actualizar los pesos en la capa oculta

$$w_{ji}^{h}(t+1) = w_{ji}^{h}(t) + \alpha \delta_{pj}^{h} x_{pi} + \mu\big(w_{ji}^{h}(t) - w_{ji}^{h}(t-1)\big)$$

Por último, nos hace falta un termino que nos vaya diciendo la magnitud del error y ese término es el siguiente:

Paso 6

Calcular la función de error.

$$E_p = \frac{1}{2}\sum_{k=1}^{M}(d_{pk} - Z_{pk})^2 \tag{3.10}$$

Esta función refleja la capacidad de adaptación de la red y al igual que como vimos en series de tiempo, debe de disminuir conforme la red vaya aprendiendo hasta una cota deseada.

Tasa de aprendizaje(α)

Durante el entrenamiento, la tasa de aprendizaje α juega un papel fundamental, ya que la velocidad del aprendizaje y el tiempo que la red requiere para entrenarse dependen en gran medida de este parámetro. Normalmente, α debe ser un número pequeño entre [0.05,0.25], para asegurar que la red llegue a asentarse en una solución. Un valor pequeño significa que la red tendrá que hacer un gran número de iteraciones. Si es muy grande, los cambios son muy grandes, avanzando muy rápidamente por la superficie de error (función 3.10), con el riesgo de saltar el mínimo y estar oscilando alrededor de él, pero sin poder alcanzarlo, por tal razón y para mejorar las posibilidades de generalización de la red es conveniente que, a medida que la red, vaya aprendiendo, el valor de α vaya disminuyendo hasta alcanzar un valor muy próximo a cero.

Momento(μ)

Como vimos anteriormente para un α grande se produce un rápido aprendizaje, pero con una gran oscilación. Para incrementar el aprendizaje sin que se produzca la oscilación, Rumelhart, Hinton y Williams (1986) sugirieron que para filtrar estas oscilaciones se incluye un nuevo término, el momento(μ). Donde μ es una constante que determina el efecto en t+1 del cambio de los pesos en el instante t.

Con este momento se consigue la convergencia de la red en un menor número de iteraciones, ya que si en t el incremento de un peso era positivo y en t+1 también , entonces el descenso por la superficie de error (3.10) en t+1 es mayor. Sin embargo, si en t el incremento era positivo y en t+1 es negativo, el paso que se da en t+1 es más pequeño, lo cual es adecuado, ya que eso significa que se ha pasado por un mínimo y que los pasos deben ser menores para poder alcanzarlo. El valor de μ puede mantenerse relativamente alto (sólo en el intervalo [0,1]) aunque este valor también debe reducirse a medida que la red va aprendiendo.

Cabe señalar y recalcar que al hablar de redes backpropagation o redes de propagación hacia atrás hacemos referencia a un algoritmo de aprendizaje más que a una arquitectura (topología) determinada, es decir, el modelo es un perceptrón multicapa con aprendizaje backpropagation o simplemente BP cuya función de activación de la capas ocultas es no lineal y las función de activación de la salida es lineal. El algoritmo consiste en propagar el error hacia atrás[10], es decir, de la capa de salida hacia la capa de entrada, pasando por las capas ocultas intermedias y ajustando los pesos de las conexiones con el fin de reducir dicho error.

3.9.2. El problema de la generalización

Unos de los factores importantes para las RNA's es su capacidad de generalizar a partir de algunos ejemplos. Por *generalización* se entiende la capacidad de dar una respuesta correcta ante observaciones que han sido empleadas en su entrenamiento. Una red entrenada correctamente generalizará, lo que significa que la red ha aprendido adecuadamente no solo los ejemplos presentados, sino que responderá correctamente ante observaciones nunca antes vistas.

En la fase de entrenamiento se deben considerar dos tipos de error importantes, por una parte, un **error de aprendizaje**, que suele calcularse con 3.10. Por otro lado, existe un **error de generalización**, el cual se puede medir con un conjunto representativo de observaciones diferentes a los utilizados en el entrenamiento. De esta manera podemos entrenar una RNA

[10]De allí su nombre.

haciendo uso de un conjunto de aprendizaje, y comprobar su eficiencia real, o error de generalización, mediante un **conjunto de prueba**. Entonces, si representamos a la vez el error en aprendizaje y el error en generalización, obtenemos una gráfica como 3.7: la cual tras una fase inicial, en la que pueden aparecer oscilaciones en el valor del error, el de aprendizaje tiende a disminuir monótonamente, mientras que el error de generalización a partir de cierto punto comienza a incrementarse, lo cual indica una degradación progresiva del aprendizaje.

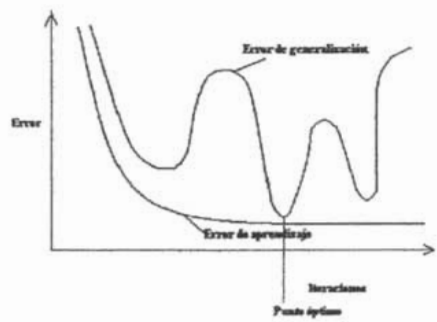


Figura 3.7: Ejemplo de una curva de aprendizaje

La explicación a la figura 3.7 es: Al principio la red se adapta paulatinamente al conjunto de aprendizaje (o entrenamiento), ajustándose al problema y mejorando la generalización. Sin embargo, en un momento dado el sistema se ajusta demasiado a las particularidades de las observaciones empleadas en el entrenamiento (a los cual muchos autores dicen que la red esta memorizando las observaciones), por lo que crece el error que cometerá ante observaciones diferentes a las empleadas en el entrenamiento (error de generalización), entonces podemos ver que en este punto la red no se ajusta bien al problema, sino simplemente está memorizando las observaciones del conjunto de aprendizaje, lo que técnicamente se denomina **sobreaprendizaje o sobreajuste**. Entonces, lo que debería de hacerse ante un problema como este es entrenarse a la red hasta un punto óptimo en el que el error de generalización es mínimo lo cual se denomina **Validación cruzada**[11], aunque en

[11]Este procedimiento consiste en entrenar y validar a la vez para detenerse en el punto

realidad se pueden presentar varios mínimos para el error de generalización, debiéndose detener el aprendizaje en el mínimo error de generalización, y no quedarnos en el primer mínimo que aparezca.

Por último, cuando se entrena una red backpropagation (en general redes supervisadas feedforward) debemos tener en cuenta todo esto, y la técnica de validación cruzada suele ser un buen remedio (varios autores consideran forzosa su utilización), usualmente, de todo el conjunto de entrenamiento se emplea aproximadamente un 80 %(aunque no es una regla) de las observaciones para entrenar, reservándose un 20 % como conjunto de validación para poder realizar la validación cruzada [José R. Hilera, 1995].

3.9.3. Fases de las RNA's

Existen dos fases en toda aplicación de las RNA's: la fase de aprendizaje (o entrenamiento) y la fase de validación.

- **Fase de entrenamiento o aprendizaje.** En la fase de entrenamiento, se usa un conjunto de entrenamiento o aprendizaje. Estos datos son los que definen el comportamiento de la red, y sirven para determinar los pesos (w_i). Los pesos, juegan un papel muy importante en esta etapa, ya que las redes aprenden a través de la actualización (o cambios) de los pesos.

- **Fase de validación.** Una vez entrenado este modelo, se usará en la llamada fase de validación, en la que se procesan datos que no son conocidos por la red (ya no son datos muestra), esto debido a que son datos de prueba y son guardados específicamente para validar la generalización de la red a determinado tiempo de ejecución. Cuando los valores de las neuronas de la última capa han sido calculados, se comparan con la salida deseada para determinar la validez del diseño, analizándose de esta manera el funcionamiento definitivo de la red.

óptimo.

Podado de pesos (pruning)

Es un técnica para reducir el número de parámetros (enlaces), esto al eliminar algunos de sus pesos, el proceso de podado consiste en entrenar a la red hasta cierto nivel, para luego eliminar aquellos pesos que no aportan prácticamente nada a la operación. Existe una modificación a esta técnica conocido como decaimiento de pesos (weight decay), en el cual durante el aprendizaje se deja a los pesos tender poco a poco a cero, para que aquellos que no se actualicen periódicamente, se anulen y desaparezcan.

3.10. Características de la red para predicción

Como ya nos habremos dado cuenta la red que vamos a utilizar es la red backpropagation(red de perceptrones multicapa con algoritmo de aprendizaje backpropagation), entonces es importante puntualizar las características que debe tener para el fin de pronóstico, claro no son reglas, pero dan un panorama general del modelo.

3.10.1. Escoger la arquitectura adecuada

Desde luego no se puede dar una receta para decir que arquitectura es la más adecuada, pero con ayuda del teorema de la aproximación universal, sabemos que con tres capas son suficientes para poder aproximar cualquier función, siempre y cuando las funciones de activación de las capas de entrada y salida sean lineales y las de la capa(s) oculta(s) sean no lineales (ver teorema Funahashi 1989) a menos que el problema sea muy complicado y se necesite agregarle otra capa oculta.

3.10.2. Datos de entrada

Los datos deben de ser normalizados antes de meterlos en la red, esto debido a que cuando las entradas son los datos originales, la red puede converger a un mínimo local ó aprender muy lentamente. Por otro lado, los datos se tienen que procesar, por ejemplo quitarle la tendencia a la serie o quitarle la estacionalidad, aunque las RNA's pueden generalizar aun con estacionalidad, pero no con tendencia, esto debido a que las salidas de la red quedarían por debajo de los valores deseados. Por último, ponemos la forma en como normalizar los datos en un rango de [0,1].

$$X_t' = \frac{X_t - min\{X\}}{max\{X\} - min\{X\}} \tag{3.11}$$

En caso de que se desee otro tipo de intervalo $[z_0, z_1]$ donde $z_0 < z_1$ se tiene

$$X_t' = \frac{X_t - min\{X\}}{max\{X\} - min\{X\}} \cdot (z_0 - z_1) + z_1 \tag{3.12}$$

Siendo min{X} el mínimo de la serie, max{X} el máximo y X_t la observación al tiempo t.

3.10.3. Capa de entrada

Como ya sabemos, la capa de entrada (durante el entrenamiento), está constituida por datos muestra (solo un porcentaje de los datos originales), es decir, son los datos que le van a enseñar a la red, el comportamiento deseado. Entonces, aunque no existe una regla para decir cuantas neuronas de entrada poner, a veces conviene poner el numero de neuronas según el periodo de la serie, aunque como se menciono, no es una regla, solo un punto de partida.

Para cuestiones de predicción de series de tiempo, los datos de entrada son definidos moviendo la red sobre la secuencia original de la serie, tantos lugares como se desee, como se muestra en la figura 3.8, aunque para predecir más de un periodo se necesita retroalimentar la red con las respuestas obtenidas por ella misma.

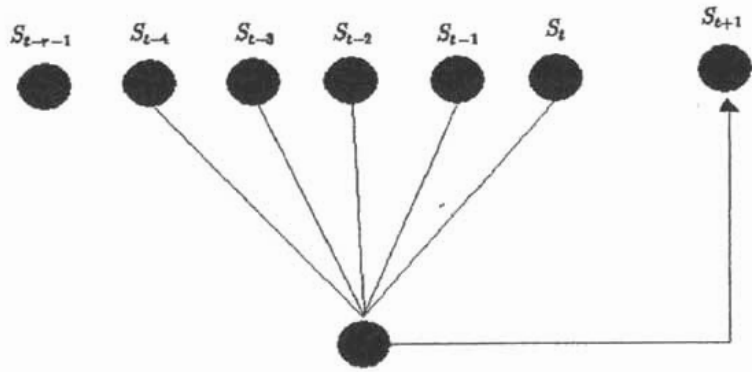


Figura 3.8: El perceptrón utiliza r observaciones de la serie de tiempo en cada neurona de entrada y se mueve uno para predecir el siguiente, donde S es un vector de observaciones.

3.10.4. Pesos y funciones de activación

Como se menciono anteriormente los pesos iniciales deben de ser aleatorios entre 0 y 1. En cuanto a la función de activación se tomará la tanh, esto debido a que la tangente hiperbólica puede exhibir diferentes dinámicas de aprendizaje, durante el entrenamiento puede acelerar el aprendizaje para algunos modelos.

3.10.5. Número de neuronas en la capa intermedia

Como se comento anteriormente, no se puede dar una regla para determinar el número de neuronas, aunque para el caso de aprendizaje supervisado existen varios resultados que son importantes de mencionar. Primero un número muy grande de neuronas en la capa intermedia hace que la red requiera menos iteraciones de entrenamiento para aprender, aunque cada uno de los pasos requiere más tiempo de cómputo ya que, en general, hay más pesos que

ajustar. El número de neuronas en la capa intermedia tampoco debe ser igual al de patrones de entrenamiento, ya que esto favorece a la memorización de los patrones, puesto que la red hace que cada una de las neuronas de la capa intermedia se encargue de reconocer uno de los patrones de entrenamiento, en vez de generalizar a partir de casos individuales.

El número de neuronas debe ser, en general, menor que el número de patrones de entrenamiento para evitar problemas de memorización, un buen punto de partida es la mitad del número de neuronas de entrada, y después ir aumentando o disminuyendo según el caso. El tamaño de la capa de entrada y salida viene dado por la naturaleza del problema.

Es posible eliminar neuronas ocultas si la red converge sin problemas, adoptando la que tiene menos neuronas intermedias. Si la red no converge, quizá sea necesario aumentar este número. Por otro lado, examinando los valores de los pesos de las neuronas ocultas periódicamente en la fase de aprendizaje, se pueden detectar aquellas cuyos pesos cambian muy poco, y reducir por tanto el número de neuronas que apenas participan en el proceso de aprendizaje.

3.10.6. Conjunto de entrenamiento

El conjunto de datos utilizados para el entrenamiento de la red es el que determina qué es lo que la red debería de hacer. Normalmente se reserva un número pequeño de patrones sin utilizar en la etapa de entrenamiento para validar (conjunto de validación o de "test") la capacidad de la red a la hora de solucionar el problema. Un entrenamiento insuficiente de la red hace que está no sea capaz de proporcionar respuestas claras y, por el contrario, un exceso de entrenamiento hace que la red memorice los vectores de entrenamiento. También es necesario que el conjunto de datos de entrenamiento sea representativo del problema, para que la red sea capaz de responder satisfactoriamente a cualquier tipo de entrada, usualmente, se emplea aproximadamente un 80 % de las observaciones para entrenar, reservándose un 20 % como conjunto de validación [Martín del Brío Bonifacio y Alfredo Sanz Molina, 1997 [26]].

3.10.7. Validación

Una vez que la red termino de aprender se le presentan las observaciones apartadas para la validación (Test) del modelo, si la respuesta es la deseada o se ajusta a nuestros (antes de que el aprendizaje comience a degradarse) requerimientos, se puede utilizar esta red para poder pronosticar. En caso contrarío se tiene que modificar el modelo ya sea agregando o quitando neuronas a la capa oculta, iniciando con pesos diferentes, ver si el conjunto de entrenamiento es representativo del problema, entre otros. Cabe aclarar que cuando la red alcanza una solución acceptable, no existe la garantía de que ha alcanzado el mínimo global. Aunque si el error no es grave, no importa si el entrenamiento se ha detenido en alguno de los mínimos locales.

3.10.8. Cuando detener el entrenamiento

- Un máximo de 10000 iteraciones de validación ó más.
- Durante la fase de validación se calcula el error cuadrático medio en el conjunto de validación, el cual debe decrecer durante el entrenamiento. Cuando el error incrementa un 20 % del mínimo error alcanzado, se debe detener el entrenamiento.

Superada esta fase, la arquitectura, el numero de neuronas y conexiones, los pesos sinápticos quedan fijos pudiendo el sistema operar en modo recuerdo. El modo recuerdo es el modo de operación normal del sistema: dada una entrada proporcionará una salida en consonancia con el aprendizaje recibido.

Capítulo 4

Aplicaciones

En este capitulo aplicamos la teoría ya descrita para los métodos estadísticos y para las redes neuronales, analizaremos las series y trataremos de pronosticar un año(corto plazo) dependiendo de la periodicidad de la serie. Cabe mencionar que se excluyo del modelo un año (tanto en los métodos tradicionales como en las RNA's) a efecto de poder comparar los resultados de ambas técnicas al final y, de está manera, poder decir cual arrojó mejores resultados.

Utilizaremos en el caso de los métodos estadísticos el software statgraphics 4.0 y en el caso de redes neuronales se utilizo en programa DataEngine 2.1.

4.1. Primera aplicación: método estadístico ("Serie Ventas")

Supongamos que el cuadro 4.1 contiene las ventas de una compañía la cual desea pronosticar el siguiente año completo, esto es, dado la periodicidad de la serie es trimestral, entonces se deberá pronosticar 4 trimestres adelante. Esta serie la trataremos con un método de alisamiento (atenuación) y luego trabajaremos esta misma serie con redes neuronales, para después usar sólo

la metodología de Box y Jenkins (ARIMA) para las demás series a trabajar durante el resto del capítulo.

t	Ventas	t	Ventas	t	Ventas	t	Ventas
1	1849.9	19	1854.3	37	1707.4	55	1690.3
2	1550.8	20	1851	38	2018.8	56	1642.3
3	1515.3	21	2042.2	39	1898.5	57	1762.3
4	1685.4	22	2272.8	40	1453.6	58	2001.5
5	1733	23	2217.7	41	1706.2	59	1788.6
6	1576	24	1672.2	42	1878.2	60	1724.8
7	1619.5	25	1890.7	43	1752.1	61	1658.2
8	1281.3	26	2242.2	44	1560.4	62	1938.4
9	1401.4	27	2246.9	45	1445.1	63	1845.3
10	1535.3	28	1827.2	46	1683.9	64	1688.9
11	1327.9	29	1669.3	47	1566.6	65	1749.4
12	1493.6	30	1972.8	48	1421.3	66	2087.7
13	1456.9	31	1878.2	49	1455.4	67	1837.1
14	1875.8	32	1560.6	50	1746.1	68	1579.5
15	1646.2	33	1914	51	1571.7	69	1704.1
16	1814.1	34	2076	52	1503.4	70	1870.8
17	1994.6	35	1787.1	53	1483.5	71	1769.7
18	2251.8	36	1782.3	54	1917.9	72	1581.8
						73	1547.3
						74	1961.6

Cuadro 4.1: Serie de tiempo para la serie ventas

Primeramente se gráfica la serie para tener la noción de su comportamiento (véase figura 4.1)

Como podemos observar, no contiene ninguna tendencia, y para reafirmar esto construimos las FAC y FACP correspondientes (véase figura 4.2). Como podemos ver en la figura 4.2 las autocorrelaciones caen rápidamente, esto nos indica la falta de tendencia, pero podemos ver que tiene estacionalidad. Entonces se ataca con un método de alisamiento, recordando que el único método capaz de manejar la estacionalidad, es el método de Winter, por lo cual este método se emplea en este trabajo.

Entonces, una vez que se hicieron los cálculos correspondientes, obtenemos los parámetros óptimos del modelo, los cuales son: $\alpha = 0{,}3948$, $\beta = 0{,}0348$ y $\gamma = 0{,}6351$. Entonces, con estos parámetros obtenemos las salidas del modelo y sus respectivos pronósticos (forecasts). Solo se muestran los últimos 25 observaciones de la serie y sus respectivas salidas, aplicando el modelo Winter,

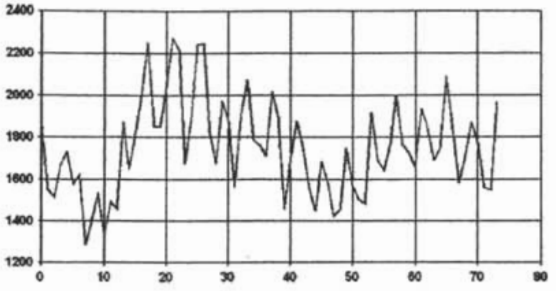


Figura 4.1: Gráfica de serie ventas)

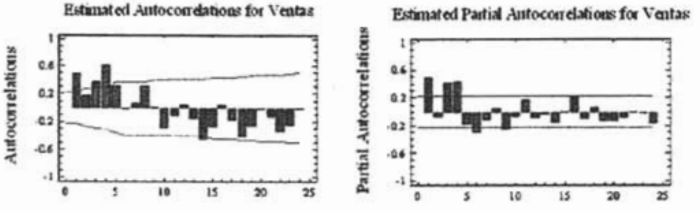


Figura 4.2: FAC y FACP de la serie ventas

así como sus respectivas predicciones (véase el cuadro 4.2), las cuales son las correspondientes a un año (4 trimestres).

Recordando que el método de Winter optimiza los parámetros de tal forma que minimicen los errores cuadráticos medios (ecuación 2.8), y los cuales son los α, β y γ ya mencionados. En el software statgraphics 4.0 se pueden comparar hasta cinco modelos al mismo tiempo y así poder ver cual de ellos minimiza el MSE y satisface varias pruebas para los residuos, como lo vemos en el cuadro 4.3, en el que podemos ver el modelo Winter(a) comparado[1] con un modelo ARIMA(b), una regresión lineal (c), una media móvil de

[1]Como podemos observar el mejor modelo para esta serie es un modelo ARIMA, pero

Forecast Table for Ventas

Model: Winter's exp. smoothing with alpha = 0.3746, beta = 0.0040, gamma = 0.6931

Period	Data	Forecast	Residual
2/34	1748.1	1678.24	69.8593
3/34	1571.7	1604.39	-27.6988
4/34	1503.4	1393.76	109.642
5/34	1462.5	1446.2	17.2962
6/34	1917.9	1733.62	184.276
7/34	1689.3	1663.6	24.5036
8/34	1842.3	1527.18	315.117
9/34	1762.3	1570.69	191.609
10/34	2001.6	2013.65	-12.0628
11/34	1766.6	1788.05	-21.4464
12/34	1724.8	1689.23	35.5178
1/35	1658.2	1718.9	-60.7035
2/35	1930.4	1979.1	-48.704
3/35	1845.3	1748.25	97.0459
4/35	1686.9	1702.78	-15.8769
5/35	1748.4	1678.35	70.0512
6/35	2087.7	2006.78	80.9247
7/35	1837.1	1871.73	-34.6319
8/35	1579.5	1729.77	-150.269
9/35	1704.1	1685.07	19.0258
10/35	1870.6	1987.26	-116.656
11/35	1769.7	1709.0	60.7
12/35	1661.8	1582.52	80.7205
1/36	1547.3	1651.05	-103.752
2/36	1961.5	1835.57	126.029

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
3/36	1748.71	1598.32	2087.24
4/36	1596.44	1632.15	2651.72
5/36	1597.91	854.609	2341.21
6/36	1917.87	719.683	3116.06

Cuadro 4.2: Serie ventas junto con las salidas del modelo Winter, así como sus pronósticos

orden 5(d) y una atenuación exponencial simple(c), después vemos el error cometido por cada uno de los modelos ya mencionados, como lo son: MSE (error cuadrático medio), el MAE (error absoluto de la media), el MAPE (porcentaje de error medio absoluto), el ME (error medio) y el MPE (porcentaje medio de error), que en nuestro modelo el MSE es 27540.4 para el periodo estimado. En la parte inferior a los errores aparece una tabla donde se muestran algunos OK, estos nos indican las pruebas que pasaron los residuos, la primera columna (RMSE) no es más que la raíz cuadrada de MSE, la segunda columna (RUNS) equivale a el quinto supuesto mencionado en el capitulo 2 (no existen observaciones aberrantes), la tercer columna (RUNM) equivale al cuarto supuesto (los residuos tienen distribución normal, para to-

se decidió trabajar con los métodos de alisamiento para esta serie.

da t), la cuarta columna (AUTO) corresponde al tercer supuesto (los residuos no estén autocorrelacionados), la quinta columna (MEAN) corresponde al primer supuesto ($\mu = 0$) y la última columna (Var) corresponde a que se cumpla el segundo supuesto ($Var(\widehat{e} = \sigma^2)$). Por otro lado, las pruebas que son aceptadas son las que tienen un OK, con un 90 % de confianza[2] y las que tienen un * son las que son violadas con cierto grado de confianza (como se ve en la parte inferior del cuadro 4.3).

```
Model Comparison
----------
(A) Winters's exp. smoothing with alpha = 0.3948, beta = 0.0048, gamma = 0.5351
(B) ARIMA(1,0,1)x(0,1,1)4 with constant
(C) Linear trend = 2735.36 + 0.396867 t
(D) Simple moving average of 5 terms
(E) Simple exponential smoothing with alpha = 0.3305

Estimation Period
Model  MSE          MAE          MAPE         ME           MPE
-------------------------------------------------------------------------------
(A)    27569.4      121.962      7.0695       19.1594      0.626545
(B)    19021.9      105.702      5.1008       2.47855      -0.497181
(C)    52125.2      180.304      10.4315      1.41345E-13  -1.54648
(D)    37110.5      151.34       8.62342      3.5771       -0.786181
(E)    40069.9      158.766      9.06078      3.81568      -0.840484

Model  RMSE         RUNS  RUNM  AUTO  MEAN  VAR
-----------------------------------------------
(A)    165.989      OK    OK    OK    OK    ***
(B)    138.849      OK    OK    OK    OK    ***
(C)    228.312      **    OK    ***   OK    **
(D)    192.641      *     OK    ***   OK    **
(E)    200.995      *     OK    ***   OK    OK

(1) the mean squared error (MSE)
(2) the mean absolute error (MAE)
(3) the mean absolute percentage error (MAPE)
(4) the mean error (ME)
(5) the mean percentage error (MPE)

Key:
RMSE = Root Mean Squared Error
RUNS = Test for excessive runs up and down
RUNM = Test for excessive runs above and below median
AUTO = Box-Pierce test for excessive autocorrelation
MEAN = Test for difference in mean 1st half to 2nd half
VAR = Test for difference in variance 1st half to 2nd half
OK = not significant (p >= 0.10)
* = marginally significant (0.05 < p <= 0.10)
** = significant (0.01 < p <= 0.05)
*** = highly significant (p <= 0.01)
```

Cuadro 4.3: Comparación del modelo Winter con otros modelo estadísticos.

SE puede observar que el modelo pasa 4 de las 5 pruebas, lo cual es muy bueno, debido a que esto nos dice que el modelo esta captando la mayor información posible. Ahora bien, en el caso de la varianza podemos ver la gráfica de los residuos contra el tiempo (ver figura 4.3), y podemos observar

[2]Esto gracias al P-value

que al final las oscilaciones se estabilizan, por tal razón, se opto por no hacer ninguna transformación para normalizar la serie y dejar así los resultados arrojados por el modelo.

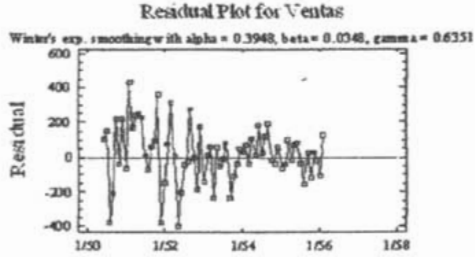


Figura 4.3: Gráfica de los residuos para la serie ventas

Por último se presenta la comparación entre los datos reales y las salidas del modelo Winter(ver la figura 4.4).

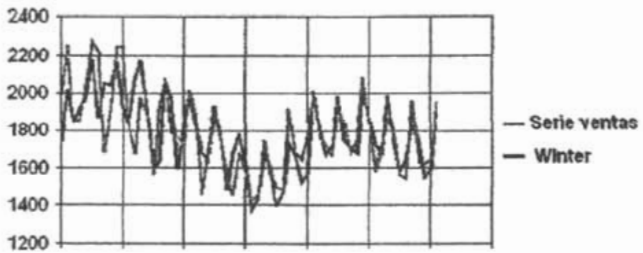


Figura 4.4: Comparación entre la serie original (ventas) y las salidas del método de Winter

Después de tratar esta serie con redes neuronales se comparará con las predicciones arrojadas por el método de Winter para saber que modelo se ajusto mejor a esta serie.

4.2. Primera aplicación; redes neuronales ("Serie Ventas")

Como ya sabemos, el modelo a utilizar es un perceptrón multicapa con mecanismo de aprendizaje backpropagation, entonces se opto por utilizar tres capas gracias al teorema de Funahashi, se trabajo con un número diferente de neuronas tanto de entrada como ocultas para ver cual minimizaba el error de entrenamiento y el de validación, se emplearon inicialmente 4 neuronas de entrada, 2 intermedias (ocultas) y 1 de salida, pero debido a que arrojaba mucho error se opto por utilizar más neuronas de entrada, entonces se procedió a trabajar con 8 neuronas de entrada, 4 intermedias y 1 de salida (8:4:1), pero los resultados no fueron muy buenos, luego se opto por una red (16:8:1) y los resultados mejoraron un poco, pero se termino con una red (12:2:1) esto debido a que se inicio con una red (12:6:1) y se fue incrementando una neurona intermedia cada vez y después se disminuyo para ver cual daba mejores resultados, y la arquitectura que arrojo mejores resultados fue: 12 neuronas en la capa de entrada, 2 neuronas en la capa oculta y 1 neurona de salida (como se muestra en la figura 4.5).

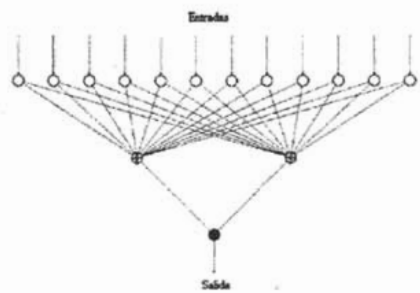


Figura 4.5: Representación de la red utilizada para la serie ventas

Las funciones de activación para las neuronas de entrada fueron lineales al igual que el de la neurona de salida y la función de activación de las neuronas de la capa oculta es tanh. En relación con el método de aprendizaje se utilizo el momento (μ) con valor de .25 y factor de aprendizaje (α) con valor .1,

también se utilizo el podado de pesos para optimizar el modelo y eliminar los enlaces innecesarios, los pesos iniciales fueron dentro del intervalo (-.1,.1), el número de iteraciones fueron hasta 18200 y la ejecución del conjunto de test cada 100 iteraciones, es decir, que después de 100 iteraciones de la fase de entrenamiento, en donde los pesos se ajustan, se introduce el conjunto de prueba para ver la eficacia del modelo al generalizar a partir del conjunto de entrenamiento. En relación con el número de iteraciones, se emplearon 18200 debido a que en este rango se minimizo el error de prueba(véase figura 4.6).

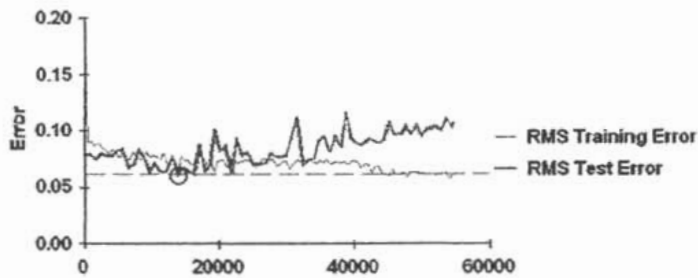


Figura 4.6: Curva de aprendizaje de la serie ventas

Nótese que a partir de un cierto número la línea de prueba empieza a divergir, y esta divergencia comienza después de 18200 por esa razón se detuvo el entrenamiento en la iteración 18200. El error que se obtuvo en la fase de prueba fue cerca del 6% en la fase de entrenamiento fue del 7%. Por último, se presentan, las salidas (véase cuadro 4.4) de la red comparadas con las salidas deseadas, y podemos ver que la red arroja algunos resultados certeros o muy próximos a las salidas deseadas.

Antes de proseguir a comparar los pronósticos arrojados tanto por el método de Winter como el de redes neuronales, hago un recordatorio en la fase del preprocesamiento de datos, la serie no debe de contener tendencia aunque no necesariamente estacionalidad, esto debido a lo expuesto en el capítulo 3. La serie no tiene que tener tendencia (esto se puede hacer sacando una media móvil dependiendo de la periodicidad de la serie y restarsela a la original, aunque después se le tiene que agregar) para poder hacer los cálculos, ya

Serie original	Salida de la red
1746.1	1788.8
1571.7	1736.3
1503.4	1635.4
1483.5	1648.0
1917.9	1872.5
1690.3	1797.3
1642.3	1697.5
1782.3	1683.3
2001.5	1967.0
1786.5	1809.3
1724.8	1850.9
1656.2	1805.2
1938.4	1947.2
1945.3	1945.5
1886.9	1805.0
1749.4	1723.7
2067.7	1984.0
1837.1	1928.9
1579.5	1903.1
1704.1	1729.1
1870.5	1973.7
1769.7	1772.7
1561.8	1517.2
1547.3	1676.1
1951.5	1882.9

Cuadro 4.4: Comparación de las salidas deseadas con las salidas de la red

que si se meten los datos con tendencia las salidas de la red quedarán por debajo de las salidas deseadas, pero no hay problema si existe estacionalidad en la serie, esto se comprueba en el cuadro 4.4 donde se trabajo con la serie con estacionalidad y la red aprendió este fenómeno. Por otro lado, como se pronosticaron 4 valores adelante, se retroalimento la red con los valores obtenidos por ella misma. Después de haber aclarado lo anterior, se analizan los resultados obtenidos por los dos métodos, esto se muestra en el cuadro 4.5,

t	Serie original	Redes neuronales	Winter
75	1633.1	1719.074	1740.71
76	1621.9	1733.185	1554.44
77	1646.3	1762.803	1597.95
78	1951	1899.848	1917.87
		MSE RNA	MSE ARIMA
		17992.608	9783.041

Cuadro 4.5: Tabla con los pronósticos de ambos métodos contra los valores reales (serie ventas)

en la cual, como podemos observar, la primera columna muestra los valores reales de la serie, en la segunda columna se encuentran los pronósticos obtenidos por redes neuronales y en la parte inferior su respectivo error cuadrático medio (MSE); en la tercera columna encontramos los pronósticos arrojados por el método de Winter, así como su respectivo error, entonces se observa que el que minimiza los MSE para estos periodos es el método de Winter y también el que se ajusto mejor a esta serie fue el método de Winter. Por último se muestra la comparación de los dos métodos y la serie original (ver figura 4.7)

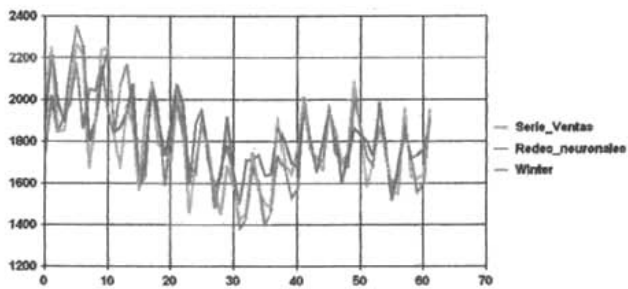


Figura 4.7: Comparación de los dos métodos y la serie original (ventas)

4.3. Segunda aplicación; método estadístico ("Serie food")

Supongamos ahora que tenemos las ventas mensuales (véase cuadro 4.6) de una compañía que venda alimento para perro y queremos pronosticar los siguientes 12 meses, como se había mencionado anteriormente esta vez se utilizará la metodología de Box y Jenkins para tratar esta serie y posteriormente se comparará con los resultados obtenidos por RNA's.

Primeramente, se gráfica (véase figura 4.8) la serie contra el tiempo como se hizo en el ejemplo anterior, esto para poder ver que componentes posee la

t	Ventas food	t	Ventas food	t	Ventas food	t	Ventas food
1	53.5	13	52.1	25	52.3	37	53.3
2	53	14	51.5	26	51.5	38	53.1
3	53.2	15	51.5	27	51.7	39	53.5
4	52.5	16	52.4	28	51.5	40	53.5
5	53.4	17	53.3	29	52.2	41	53.9
6	56.5	18	55.5	30	57.1	42	57.1
7	65.3	19	64.2	31	63.6	43	64.7
8	70.7	20	69.6	32	68.8	44	69.4
9	68.9	21	69.3	33	68.9	45	70.3
10	58.2	22	58.5	34	60.1	46	62.6
11	55.3	23	55.3	35	55.6	47	57.9
12	53.4	24	53.6	36	53.9	48	55.8

Cuadro 4.6: Serie de tiempo de las ventas de alimento (food),las ventas están en miles

serie y poder tomar las medidas pertinentes, esto se comento en el capítulo 2; para aplicar un modelo ARMA la serie tiene que ser estacionaria, por tal razón, debemos ver si no existe tendencia y/o estacionalidad, en caso de existir cualquier componente se aplicará el operador diferencia dependiendo del caso.

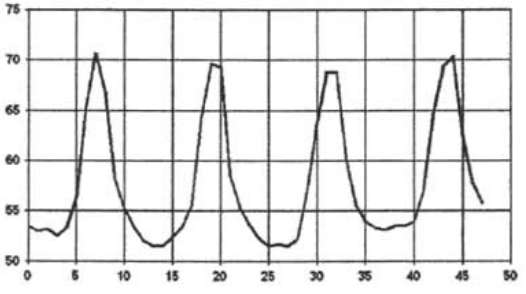


Figura 4.8: Gráfica de la serie food

Al observar la serie podemos darnos cuenta a simple viste de la existencia de estacionalidad y esto también se comprueba con la FAC y la FACP como se muestra en la figura 4.9, en donde podemos ver las oscilaciones estacionales con facilidad las cuales se repiten cada 12 meses.

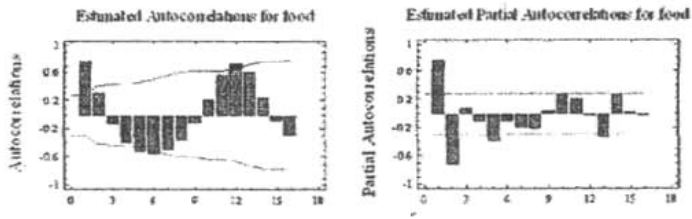


Figura 4.9: Correlogramas de la serie food antes de aplicar una diferenciación estacional

Debido a que la serie mostró estacionalidad se opto por hacer una diferencia estacional con periodo 12 y se obtuvo las nuevas FAC y FACP, como se muestran en la figura 4.10.

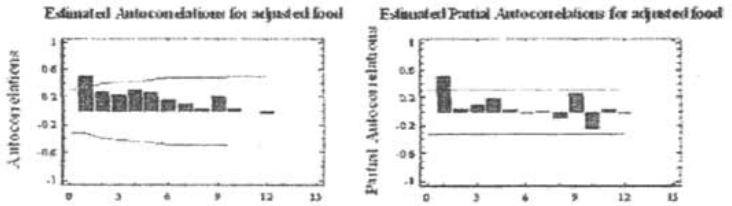


Figura 4.10: Correlogramas de la serie food después de aplicar una diferencia estacional

Una vez hecha la diferenciación podemos observar que solo quedan las primeras autocorrelaciones para ambos correlogramas, esto nos da una idea del modelo a escoger, como puede ser un modelo ARIMA$(1,0,0)x(0,1,0)_{12}$ o un ARIMA$(1,0,1)x(0,1,0)_{12}$, las correlaciones son muy altas y entonces se recomienda emplear un AR en vez de un MA, pero no se descarta la posibilidad de que puedan ser ambos. Esto se soluciona fácilmente gracias a que Statgraphics cuenta con una opción en la cual puedes comparar hasta 5 modelos y ver cual de ellos pasa la mayor cantidad de pruebas para los residuos, además de que modelo tiene el menor MSE. Dicho esto, pasamos a ver las comparaciones de estos dos modelos ARIMA y poder así ver cual es mejor.

Para esto observemos el cuadro 4.7.

Model Comparison

Data variable: food
Number of observations = 49
Start index = 1/50
Sampling interval = 1.0 month(s)
Length of seasonality = 12

Models

(A) ARIMA(1,0,0)x(0,1,0)12 with constant
(B) ARIMA(1,0,1)x(0,1,0)12 with constant
(C) ARIMA(0,0,1)x(0,1,0)12 with constant
(D) Simple moving average of 5 terms
(E) Simple exponential smoothing with alpha = 0.0197

Estimation Period

Model	MSE	MAE	MAPE	ME	MPE
(A)	1.09251	0.817768	1.40271	0.0854837	0.0242936
(B)	1.02331	0.691311	1.18042	0.219935	0.362143
(C)	1.21159	0.917331	1.58106	0.0158627	-0.00874748
(D)	62.8161	6.68791	11.1386	0.822326	-0.280287
(E)	41.1407	5.09893	8.44499	0.797612	0.30273

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	1.04523	OK	OK	OK	*	OK
(B)	1.01149	OK	OK	OK	OK	OK
(C)	1.10072	OK	OK	OK	**	OK
(D)	7.90671	***	***	***	OK	OK
(E)	6.4141	***	***	***	OK	OK

Cuadro 4.7: Comparación de modelos para la serie food.

Como podemos ver en este cuadro, se comparan 5 modelos para la serie en estudio, los primeros 3 corresponden a los posibles modelos ARIMA, el cuarto a una media móvil de orden 5 y el quinto a una atenuación exponencial simple. Con lo anterior podemos ver que los últimos dos modelos dejan mucho que desear al fallar en 3 pruebas sin contar que su MSE es mucho mayor con respecto a los tres primeros. Ahora, haciendo un análisis de los primeros 3 modelos se observa que el mínimo MSE es el (b), aunque el modelo (a) también es aceptable, pero al analizar las pruebas de los residuos nos encontramos con que el modelo (b) pasa todas las pruebas mientras el modelo (a) no pasa una al igual que el modelo (c). Recordemos que entre más pruebas pase el modelo, mayor será la captación de información del comportamiento de la serie, por tal motivo se opto por escoger el modelo (b) correspondiente a un modelo ARIMA$(1,0,1)x(0,1,0)_{12}$. Por último, mostramos las salidas del modelo con respecto a la serie y sus respectivos pronósticos (véase cuadro 4.8) en donde podemos ver los 12 periodos de pronóstico, al igual que en el ejemplo anterior, primero trabajaremos la serie food con redes neuronales y luego compararemos resultados para ver cual modelo se ajusto mejor a la

serie en estudio.

Forecast Table for food
Model: ARIMA(1,0,1)x(0,1,0)12 with constant

Period	Data	Forecast	Residual
12/81	33.8	37.5199	0.0001837
1/82	42.3	32.2842	0.048787
2/82	34.3	31.8704	-0.134404
3/82	41.7	41.8369	0.0620491
4/82	51.3	52.3662	-1.06623
5/82	38.8	57.155	-0.963021
6/82	57.1	58.0948	2.00097
7/82	60.6	64.2941	-0.784147
8/82	68.8	69.5724	-0.772442
9/82	68.6	69.0579	-0.157994
10/82	60.1	59.024	1.07604
11/82	58.6	58.3582	0.0347718
12/82	33.6	33.0087	0.000137858
1/83	39.3	38.6102	0.688839
2/83	43.1	42.0222	1.07778
3/83	39.3	38.387	0.930004
4/83	53.3	52.6341	0.668827
5/83	62.3	53.3946	0.300383
6/83	57.1	58.5902	-1.4902
7/83	64.1	64.0801	0.0199236
8/83	68.4	68.8762	-0.476218
9/83	70.3	59.9464	5.6555
10/83	62.8	61.187	1.613
11/83	67.9	67.1078	0.790223
12/83	53.8	56.6408	0.104619

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
1/84	56.0666	53.0369	57.1543
2/84	64.8003	62.7582	67.0451
3/84	33.3031	33.0777	37.3284
4/84	52.0098	53.3068	57.838
5/84	57.7245	53.3329	58.0384
6/84	58.8192	58.4646	61.2777
7/84	66.8238	63.9989	68.0484
8/84	73.2203	68.5327	73.8213
9/84	72.1331	69.4767	74.7813
10/84	64.6279	61.7138	67.3387
11/84	58.7424	56.8547	61.528
12/84	67.6467	64.8024	68.4861

Cuadro 4.8: Salidas del modelo ARIMA$(1,0,1)x(0,1,0)_{12}$, así como sus pronósticos (serie food)

Po último mostramos la comparación del método ARIMA y la serie original (ver figura 4.11)

4.4. Segunda aplicación: redes neuronales ("Serie food")

Una vez trabajada la serie food con el modelo ARIMA, pasamos a tratar esta misma serie, pero esta vez con redes neuronales, para pronosticar 12 periodos y poder así compararlos con los pronósticos arrojado por el método

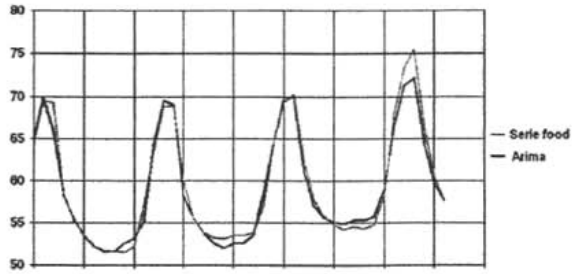


Figura 4.11: Comparación de la serie original (food) y las salidas del modelo ARIMA

estadístico.

Recordemos que nuestra modelo siempre va a ser un perceptrón multicapa con aprendizaje backpropagation, lo interesante aquí es encontrar la arquitectura adecuada que minimice el error de generalización. Se opto por utilizar una red de tres capas, se inicio con 12 neuronas de entrada y se fue incrementando de una a una las neuronas ocultas y una neurona de salida, pero esta vez no funciono tener 12 neuronas de entrada, esto porque no siempre se puede trabajar todas las series con una misma red, aún cuando tengan la misma periodicidad. Luego se opto por utilizar 9 neuronas de entrada y se repitió el proceso con las neuronas de entrada y una neurona de salida, pero tampoco dio buenos resultados, esto debido a que la curva de aprendizaje divergía rápidamente y el error de test era muy grande, luego se escogió 8 neuronas de entrada y se repitió lo anterior tampoco sin éxito, hasta que se trabajo con una red con 6 neuronas de entrada y con 3 neuronas en la capa oculta y una de salida, arrojaba mejores resultados, pero estos fueron mejorados con una arquitectura(6:5:1) la cual minimizo el error de test aun más que el modelo anterior, luego se trato con una red de 4 neuronas de entrada y se fue incrementando de una en una las neuronas ocultas, pero volvía a incrementarse el error de test, por tanto se opto por trabajar con la red con 6 neuronas de entrada, 5 ocultas y una de salida, como se ve en la figura 4.12.

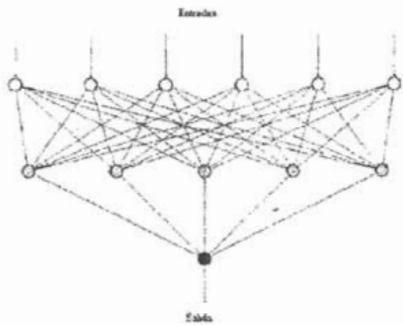


Figura 4.12: Representación de la red utilizada para la serie food

Los datos se normalizaron en el intervalo [.1,.9] de acuerdo a la ecuación 3.12. Las funciones de activación para las neuronas de entrada fueron lineales, para las neuronas ocultas se tomo la tanh y para la neurona de salida una función de activación lineal. Para el método de aprendizaje el valor del momento fue $\mu = .25$, para la tasa de aprendizaje se tomo $\alpha = .05$, también se utilizo el podado de pesos para eliminar los enlaces innecesarios. Los pesos iniciales están dentro del intervalo(-.1,.1), por último el número de iteraciones fue 300, esto debido a que fue en el punto donde se minimizo el error de prueba(véase figura 4.13) el cual fue de 5.7% y el error de entrenamiento fue del 6.3%. Al ver la figura 4.13 no damos cuenta que a partir de una cierta iteración la línea de prueba empieza a divergir, entonces lo que se hizo fue detener el aprendizaje en la iteración donde se encontró el mínimo error de prueba la cual fue la iteración número 300.

Una vez visto lo anterior, se ponen los pronósticos de ambos métodos para compararlos con los valores reales y poder decir cual de ellos produjo mejores resultados, esta comparación la vemos en el cuadro 4.9.

Podemos ver que en este caso se ajusta mejor el método ARIMA que las redes neuronales, sin embargo analizando los datos obtenidos por redes neuronales podemos ver que conforme pasa el tiempo los pronósticos empiezan a deteriorarse drásticamente, una posible causa a esto puede ser a la retroal-

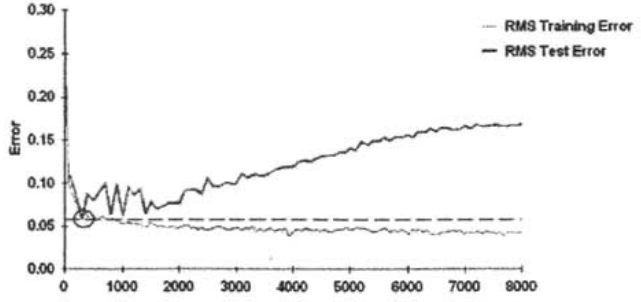


Figura 4.13: Curva de aprendizaje para la serie food

imentación de la red, esto debido a que la red se va retroalimentando con sus propias salidas y por supuesto esa salida tiene error, entonces es lógico pensar que entre más periodos se quieran pronosticar mayor será el error cometido, otra posible razón es una tendencia en el último año (año a pronosticar) que no pudo captar la red, entonces aunque el modelo ARIMA arrojo mejores resultados para esta serie, no se puede descartar la idea de utilizar las redes neuronales para esta serie en un periodo corto de tiempo.

Por último se presenta la comparación entre los datos reales y las salidas del modelo ARIMA(ver la figura 4.14).

4.5. Tercera aplicación: método estadístico ("Serie Pasajeros")

A continuación mostramos la serie (véase cuadro 4.10), la cual corresponde al numero de pasajeros de una compañía de transporte que contratan su servicio de autobuses, la serie es mensual y se desea pronosticar el numero de pasajeros que se espera que usen su servicio el próximo año (los siguientes 12 meses).

t	originales	RNA	ARIMA(1,0,1)x(0,1,0)12
49	54.8	54.83	55.10
50	54.2	53.78	54.90
51	54.6	52.75	55.31
52	54.3	53.20	55.31
53	54.8	55.56	55.71
54	58.1	59.46	58.92
55	66.1	63.46	66.52
56	73.3	65.35	71.23
57	75.5	63.83	72.13
58	68.4	58.79	64.44
59	60.5	56.05	59.74
60	57.7	54.94	57.85
		MSE RNA	MSE ARIMA
		394.31	26.14

Cuadro 4.9: Comparación de los pronósticos arrojados por ambos métodos y la serie original (food)

Nuevamente utilizaremos la metodología de Box y Jenkins para ajustar un modelo ARIMA que describa el mejor modelo para esta serie, como hemos venido haciendo, primero graficaremos la serie (véase figura 4.15) para tratar de visualizar sus diferentes componentes y así empezar a ajustar el mejor modelo.

Al ver la gráfica se nota que contiene tanto estacionalidad, como tendencia, esto nos dice desde luego que no es estacionaria, para comprobar nuestras hipótesis sobre las componentes que contiene la serie, veamos los correlogramas (véase 4.16) para ver verificar este hecho.

Al analizar la FAC y la FACP de la serie podemos ver que tiene una tendencia y estacionalidad de orden 12, para atacar esto primero aplicamos una diferenciación estacional para eliminar la estacionalidad, lo cual nos queda como se muestra en la figura 4.17:

La figura 4.17 muestra que el decaimiento de las correlaciones no es tan rápida como en el ejemplo anterior, pero tampoco caen lentamente, digamos que la caída se da a una velocidad intermedia y esto nos conduce a tener dos posibilidades: la primera sería que la serie no necesitara una diferenciación simple para quitar la tendencia, y la otra sería que si se necesitara la aplicación de una diferenciación para quitar la tendencia, pero este problema se soluciona viendo las series teóricas del cuadro 2.1, que al observalas se nota que se asemeja a un modelo AR(1), entonces, esto nos dice que no se necesi-

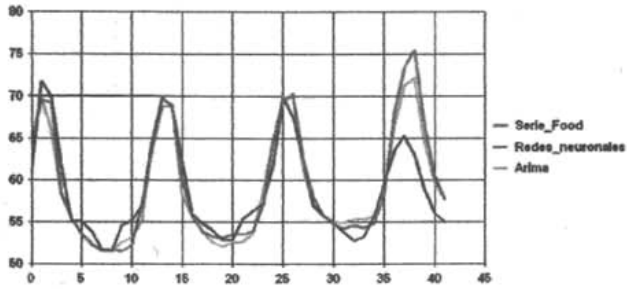


Figura 4.14: Comparación entre los dos métodos y la serie original (food)

ta una diferencia simple para quitar la tendencia, que solo con la diferencia estacional basta. Para afrontar este hecho veamos el cuadro 4.11.

Como podemos observar en el cuadro 4.11, los modelos a comparar según lo anterior son: (a) ARIMA$(1,0,0)x(0,1,0)_{12}$ que nos representa al modelo solo con la diferenciación estacional, (b) ARIMA(1,1,0)x$(0,1,0)_{12}$ que corresponde al modelo donde se aplico tanto la diferenciación estacional como la diferenciación simple para quitar la tendencia, (c) ARIMA$(0,0,1)x(0,1,0)_{12}$ que nos representa lo mismo que (a) pero esta vez se utilizo un MA en vez de un AR para el modelo, (d) una media móvil de grado 5 y (e) una atenuación exponencial simple. Gracias a esto podemos ver que tanto el modelo (a) como el (b) se ajustan bien al problema, i,e, pasan todas las pruebas de los residuos, pero el modelo que tiene un MSE menor, además por el principio de parsimonia[3], el modelo (a) es el modelo óptimo para esta serie, esto solo viene a confirmar lo que se había dicho sobre que el mejor modelo era un ARIMA$(1,0,0)x(0,1,0)_{12}$. Entonces, una vez seleccionado el modelo adecuado que nos describa la mayor información de la serie, se muestran las salidas (solo las últimas 25) del modelo, así como sus pronósticos respectivos, los cuales se muestran en el cuadro 4.12.

Estos resultados serán comparados con las predicciones arrojadas por redes neuronales para determinar que modelo se ajusta mejor a esta serie.

[3]Escoger el modelo con el menor número de parámetros que describa el modelo.

t	pasajeros	t	pasajeros	t	pasajeros	t	pasajeros	t	pasajeros	t	pasajeros	t	pasajeros	t	pasajeros
1	112	18	140	35	146	52	235	69	259	86	277	103	455	120	337
2	118	19	170	36	166	53	229	70	229	87	317	104	467	121	360
3	132	20	170	37	171	54	243	71	203	88	313	105	404	122	342
4	129	21	158	38	180	55	264	72	229	89	318	106	347	123	406
5	121	22	133	39	193	56	272	73	242	90	374	107	305	124	396
6	135	23	114	40	181	57	237	74	233	91	413	108	336	125	420
7	148	24	140	41	183	58	211	75	267	92	405	109	340	126	472
8	148	25	145	42	218	59	180	76	269	93	355	110	360	127	548
9	136	26	150	43	230	60	201	77	270	94	306	111	362	128	559
10	119	27	178	44	242	61	204	78	315	95	271	112	348	129	463
11	104	28	163	45	209	62	188	79	364	96	306	113	363	130	407
12	118	29	172	46	191	63	235	80	347	97	315	114	435	131	362
13	115	30	178	47	172	64	227	81	312	98	301	115	491	132	405
14	126	31	199	48	194	65	234	82	274	99	356	116	505		
15	141	32	199	49	196	66	264	83	237	100	348	117	404		
16	135	33	184	50	196	67	302	84	278	101	355	118	359		
17	125	34	162	51	236	68	293	85	284	102	422	119	310		

Cuadro 4.10: Serie de tiempo de la serie pasajeros

Por último se presenta la comparación entre los datos reales y las salidas del modelo ARIMA(ver la figura 4.18).

4.6. Tercera aplicación: redes neuronales ("Serie Pasajeros")

De igual manera como hemos estado haciendo, pasamos a tratar esta serie con redes neuronales y pronosticar 12 periodos a futuro, y con esto poder comparar con los resultados obtenidos con el método estadístico.

La red fue la misma que en los ejemplos pasados y el mecanismo de aprendizaje backpropagation. Se utilizaron varias arquitecturas(con diferentes numeros de neuronas en la capa de entrada e intermedias), pero la que mejores resultados arrojo fue una arquitectura con 12 neuronas de entrada, 8 neuronas ocultas y 1 de salida (ver figura 4.19).

Las funciones de activación son las mismas, lineales para las neuronas de entrada y de salida y tanh para las neuronas de la capa oculta. Para el método de aprendizaje se tomo para el momento $\mu = .25$, para el caso de la tasa de aprendizaje se tomo $\alpha = .08$, nuevamente se utilizo el podado para

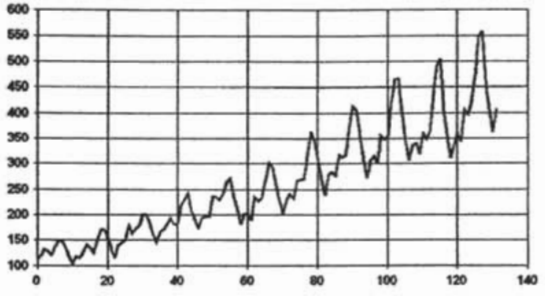


Figura 4.15: gráfica de la serie pasajeros

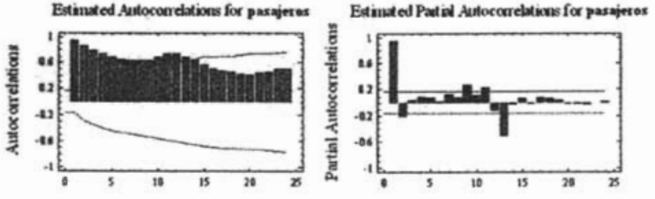


Figura 4.16: correlogramas de la serie pasajeros

eliminar los enlaces innecesarios, los pesos están dentro del intervalo (-1,1) y los datos se normalizaron en el intervalo [0,1] de acuerdo al a ecuación 3.11. Por último el número de iteraciones fue de 10900, debido a que fue el punto donde se minimizo el error de prueba(ver figura 4.20).

El error de prueba fue cercano al 4% y el error de entrenamiento fue cercano al 3%, que es considerablemente bajo a pesar de que la serie cuenta con tendencia y estacionalidad, pero hay que recordar que cuando una serie tiene tendencia se le tiene que quitar para poder trabajar con una serie estacionaria, como en el caso de los modelos ARIMA que exijen que la serie sea estacionaria, ahora regresando al caso de redes, cuando se trabaja la serie se esta trabajando con una serie estacionaria debido que primeramente tuvo que haber un preprocesamiento de los datos para quitar estas componentes

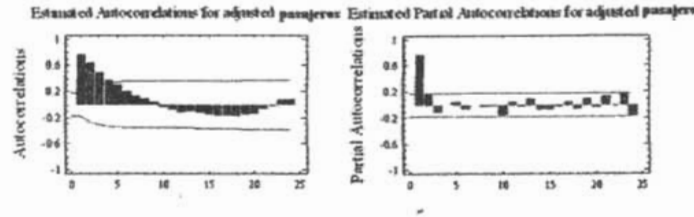


Figura 4.17: correlogramas de la serie pasajeros una vez que se aplico una diferenciación estacional

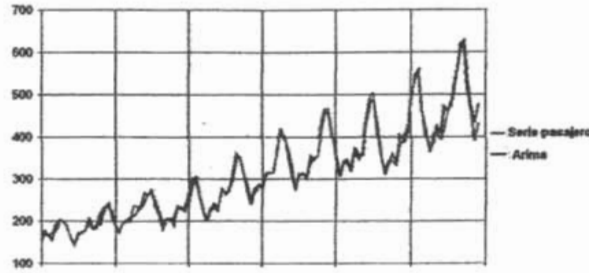


Figura 4.18: Comparación entre el modelo ARIMA y la serie original (pasajeros)

y después de que haya terminado la red en aprender y después de que haya arrojado su pronostico se le suma la tendencia para que este completa la serie.

Una vez expuesto lo anterior se muestran los pronósticos arrojados por ambos métodos (ver cuadro 4.13) y se comparan con los valores reales respectivos a sus periodos, y así poder concluir cual modelo arrojo mejores resultados.

Al observar el cuadro podemos notar que las redes neuronales son las que mejor se ajustan a esta serie en particular, al tener un error cuadrático medio menor al obtenido por el ARIMA, por tanto podemos decir que se ajusto mejor el método de redes neuronales que el método estadístico.

```
Model Comparison
----------------
Data variable: pasajeros
Number of observations = 132
Start index = 1/50
Sampling interval = 1.0 month(s)
Length of seasonality = 12

Models
------
(A) ARIMA(1,0,0)x(0,1,0)12 with constant
(B) ARIMA(1,1,0)x(0,1,0)12 with constant
(C) ARIMA(0,0,1)x(0,1,0)12 with constant
(D) Simple moving average of 5 terms
(E) Simple exponential smoothing with alpha = 0.9999

Estimation Period
Model  MSE          MAE          MAPE         ME           MPE
-------------------------------------------------------------------
(A)    106.228      8.18453      3.20567      0.189501     -0.273505
(B)    110.312      8.32657      3.2179       -0.00665951  -0.135474
(C)    174.948      10.566       4.24945      0.0898553    -0.851923
(D)    2627.63      38.4724      13.7621      6.7937       0.969893
(E)    901.759      23.9022      8.91169      2.21988      0.413665

Model  RMSE         MSES  MAES  AUTO  MEAN  VAR
-----------------------------------------------
(A)    10.3067      OK    OK    OK    OK    OK
(B)    10.5029      OK    OK    OK    OK    OK
(C)    13.2268      OK    OK    ***   **    **
(D)    50.2755      ***   ***   ***   OK    ***
(E)    31.333       ***   OK    ***   OK    ***
```

Cuadro 4.11: comparación de diferentes modelos para la serie pasajeros

Por último se presenta la comparación gráfica entre los datos reales y las salidas de los dos modelos (ver la figura 4.21).

4.7. Cuarta aplicación: método estadístico ("Serie Revistas")

Esta vez analizaremos una serie semanal, es decir, corresponde a una serie de ventas semanales de una revista y se desea pronosticar el siguiente mes, es decir, las siguientes cuatro semanas[4], la serie revistas se muestra en el cuadro

[4]Se opto por pronosticar solo un mes y no un año, esto debido a que su periodicidad es semanal, y por tanto sus pronósticos serían bastantes y como consecuencia se iría degradando la exactitud del pronóstico

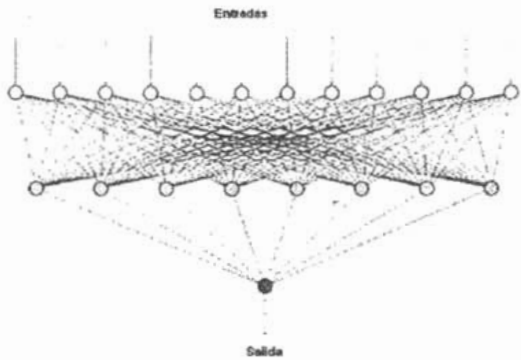


Figura 4.19: Representación de la red utilizada para la serie pasajeros

4.14.

Considerando la gráfica de la serie para poder identificar ciertos componentes que contiene la serie, la cual se muestra en la figura 4.22.

Al observar la gráfica podemos ver que no existe tendencia, pero no es claro la existencia de estacionalidad, por esta razón, tenemos que analizar los correlogramas para poder ver si existe o no estacionalidad (4.23).

Al analizar las FAC y las FACP podemos observar que efectivamente no existe tendencia y, en relación con la existencia de estacionalidad, podemos ver que existen correlaciones que se repiten, pero no más allá de dos veces, esto nos dice que no existe estacionalidad, por tanto no se necesitará aplicar una diferenciación estacional, pero al haber un comportamiento casi estacional, nos habla de la necesidad de integrar uno o más parámetros estacionales. Entonces con base en lo anterior y las características de las FAC y FACP, los modelos a considerar son: un ARIMA(1,0,0)x$(0,0,1)_7$[5], un ARIMA(1,0,1)x$(0,0,1)_7$, un ARIMA(0,0,1)x$(0,0,1)_7$, un ARIMA(1,0,0)x$(1,0,0)_7$

[5]El 7 es porque la serie viene diariamente y su periodo estacional cada 7 dias que corresponde a una semana.

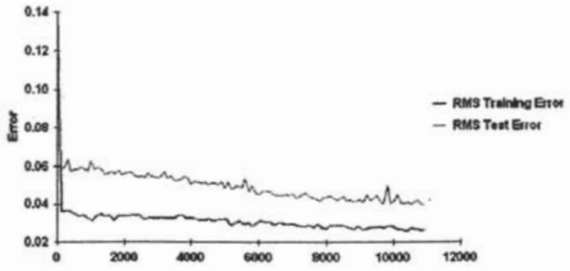


Figura 4.20: Curva de aprendizaje para la serie pasajeros

y un ARIMA(1,0,0)x$(1,0,1)_7$. Estos modelos son comparados (véase cuadro 4.15) para observar cual es el modelo que mejor se ajusta a nuestros datos.

En el cuadro 4.15 podemos ver que el modelo que pasa todas las pruebas, así como el que tiene el mínimo MSE es el modelo (a); que corresponde a un ARIMA(1,0,0)x$(0,0,1)_7$. Entonces una vez que se a elegido el modelo, se muestran las salidas del modelo (solo las últimas 25), así como los pronósticos arrojados por el mismo (véase cuadro4.16), para poder compararlos con los pronósticos arrojados por redes neuronales.

Debe notarse la importancia de saber la periodicidad en la cual la serie viene, por que sí uno escoge una periodicidad diferente, el modelo nunca se va ajustar y provocara muchos errores en los pronósticos.

Por último se presenta la comparación entre los datos reales y las salidas del modelo ARIMA(ver la figura 4.24).

4.8. Cuarta aplicación: redes neuronales ("Serie Revistas")

Ahora pasamos a trabajar esta serie(4.14) con redes neuronales para poder pronosticar 4 periodos a futuro y compararlos con las predicciones arrojadas

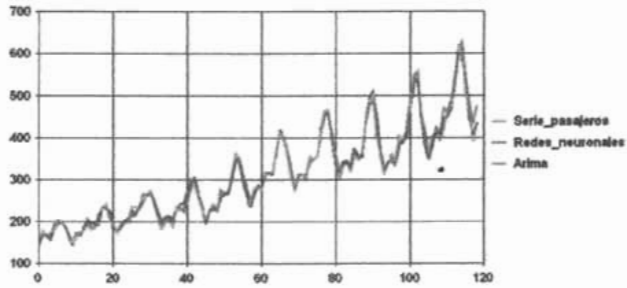


Figura 4.21: Comparación entre los dos métodos y la serie original (serie pasajeros)

por el modelo ARIMA.

Primeramente se normalizaron los datos en un rango [0.1, 0.9] utilizándose la formula 3.12, se trataron varias arquitecturas, pero la que mejores resultados dio fue una red con 12 neuronas de entrada, 5 neuronas ocultas y una de salida(como se muestra en la figura 4.25).

Las funciones de activación fueron lineales tanto para las neuronas de entrada como para la de salida, las función de activación para las neuronas ocultas fue la tanh. Para la fase de aprendizaje se utilizo μ =.25(momento) y una tasa de aprendizaje α =.07, nuevamente se utilizo el podado de pesos para optimizar el modelo y eliminar los enlaces innecesarios, los pesos iniciales fueron dentro del intervalo (-.1,.1). Esta vez se detuvo el entrenamiento en la iteración número 28800, esto debido a que fue el punto en donde se minimizo el error de prueba, para apreciar esto mejor veamos la figura 4.26.

En esta gráfica podemos apreciar que donde se encuentra el mínimo de la curva de prueba es cerca de los 30000 y para ser exactos es en el punto 28800, después de este punto empieza a divergir el error de prueba y como consecuencia a degradarse nuestros pronósticos. El error de prueba fue del 3% y el error de entrenamiento fue del 4.2%, esto quiere decir que se ajusto muy bien la red a la serie. Una vez terminado la fase de entrenamiento se

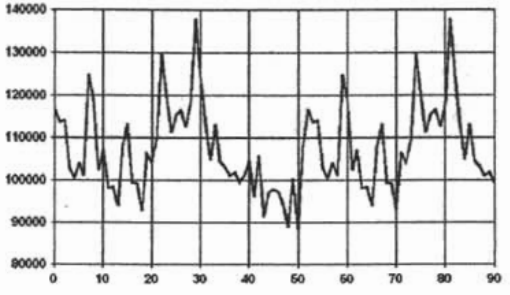


Figura 4.22: Gráfica de la serie revistas

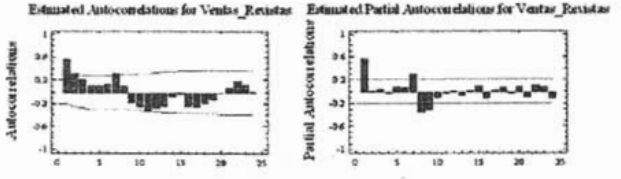


Figura 4.23: Correlogramas de la serie revistas

pasa a mostrar sus ultimas 25 salidas de la red comparadas con las salidas deseadas lo cual se muestra en el cuadro 4.17

Una vez que se mostraron las salidas de la red pasamos a comparar los pronósticos arrojados tanto con el método estadístico como con redes neuronales y así poder ver cual método se ajusto mejor a esta serie. Para poder observar esto veamos el cuadro 4.18

Finalmente podemos observar que el método de redes neuronales arrojo mejores pronósticos, al tener un MSE mucho menor que al arrojado por el método estadístico, esto nos habla de que la red aprendió bastante bien el comportamiento de la serie. Con esto vemos la importancia que tiene el tener un buen entrenamiento y un bajo nivel de error de prueba, para un buen pronóstico.

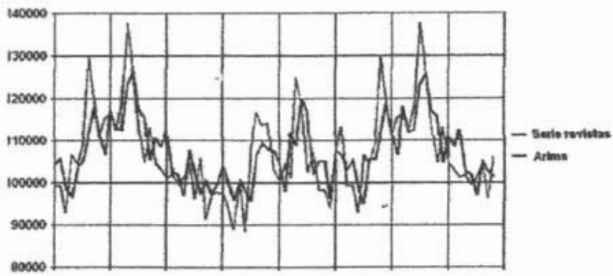


Figura 4.24: Comparación entre los dos métodos y la serie original (revistas)

Por último se presenta la comparación entre los datos reales y las salidas de los dos modelos (ver la figura 4.27).

4.9. Quinta aplicación: método estadístico ("Serie Tiie")

En esta última aplicación se trataran de pronosticar 12 periodos, de la serie de datos históricos mensuales de la Tiie (tasa de interés interbancario) desde Enero de 1980 hasta Abril del 2002. Nuestro propósito es pronosticar los próximos 12 meses. Primeramente se muestra la serie de la Tiie en el cuadro 4.19.

Luego de presentar la serie, la graficamos para hacer un previó análisis sobre sus características antes de pasar a analizar las FAC y FACP de la serie, entonces la gráfica se muestra en la figura 4.28.

Al analizar esta serie parece que no existe estacionalidad, aunque esto lo confirmaremos con ayuda de los correlogramas. En relación con la tendencia parece que si existe una tendencia negativa, aunque esta vez no se ve tan claro como en los ejemplos anteriores, pero hay ocasiones en la que el solo hecho de estudiar la gráfica no basta, por las razones recién descubiertas (no

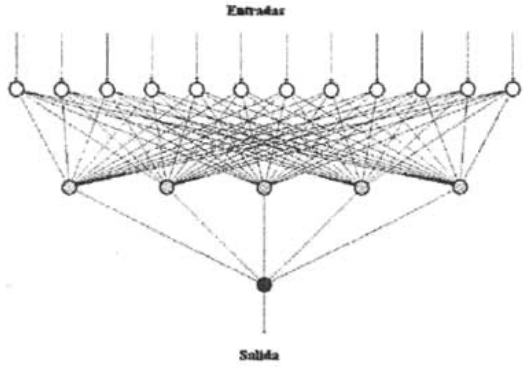


Figura 4.25: Representación de la red utilizada para la serie revistas

siempre se ve a simple vista), entonces para esto pasamos a estudiar la FAC y la FACP de la serie para poder determinar que componentes contiene la serie en estudio. La figura 4.29 muestra la FAC y la FACP respectivamente.

Al analizar los correlogramas observamos que efectivamente existe una tendencia, esto debido a que las correlaciones caen lentamente a cero. Entonces, se aplican operadores en diferencia a la serie para quitar la tendencia y así verificar si realmente no existe estacionalidad, la figura 4.30 muestra la FAC y la FACP ya diferenciadas.

Al analizar los correlogramas de la figura 4.30 observamos que efectivamente no existe estacionalidad, y que ya no se necesita otra diferenciación simple para hacer estacionaria la serie. Entonces, observando las características de los correlogramas se tienen dos posibles modelos a considerar: Un ARIMA(2,1,0) o un ARIMA(2,1,1), esto debido a que el modelo se parece a un modelo teórico (ver el cuadro 2.1) el cual corresponde a un AR(2), pero también cabe la posibilidad de que pueda ser un ARIMA(2,1,1). Comparemos los dos modelos para ver cual se ajusta mejor a la serie, lo cual se muestra en el cuadro 4.20.

Al ver este cuadro nos damos cuenta que el modelo que pasa todas las pruebas

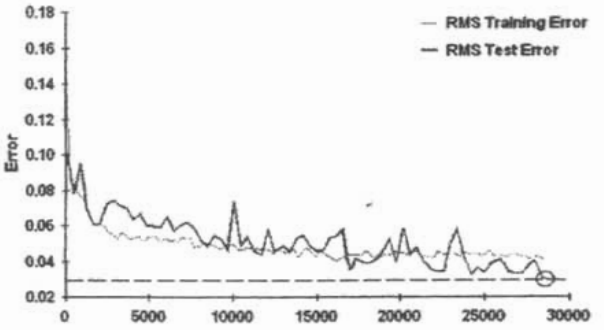


Figura 4.26: Curva de aprendizaje para la serie revistas

es el modelo (a) el cual corresponde a un ARIMA(2,1,1), pero el que tiene el menor MSE es el modelo (b) que corresponde a un ARIMA(2,1,0), entonces entramos en un dilema, pero se opto por escoger el modelo (a), debido a que la diferencia del MSE entre el modelo (a) y (b) es mínima y a que el modelo (a) pasa todas las pruebas para los residuos, esto no quiere decir que no se pueda ocupar el modelo (b), pero a causa de que el modelo (a) pasa todas las pruebas y su error cuadrático medio no dista mucho del modelo (b), se puede suponer que se tendrán mejores pronósticos con el modelo elegido.

Entonces, una vez elegido el modelo, se presentan las salidas (las últimas 25 salidas) del mismo, así como sus respectivos pronósticos, en el cuadro 4.21 se muestra lo anterior.

Cabe mencionar que debido a que la Tiie es una variable en la cual puede haber cambios muy drásticos, se recomendaría que en un caso practico solo se tomara en cuenta el primer pronóstico, esto debido a que si uno desea trabajar con una variable como está, se puede encontrar con muchas sorpresas, como se verá al comparar los resultados obtenidos con los reales. A fin de ejemplificar esto se opto por pronosticar todo el año.

Por último se presenta la comparación entre los datos reales y las salidas del modelo ARIMA(ver la figura 4.31).

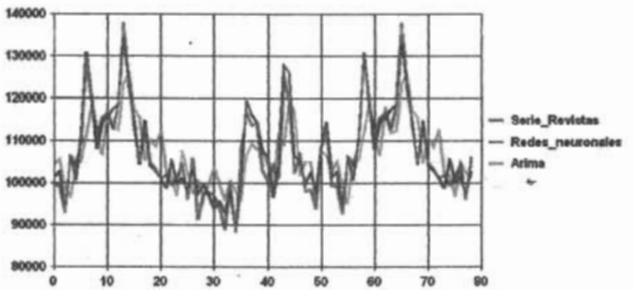


Figura 4.27: Comparación entre los dos métodos y la serie original (revistas)

4.10. Quinta aplicación: redes neuronales ("Serie Tiie")

Ahora pasamos a tratar esta serie con redes neuronales, y así poder pronosticar 12 periodos a futuro para después comparar los resultados con los arrojados por el modelo ARIMA.

Primeramente se normalizaron los datos en un rango de [0.1, 0.9] utilizándose la formula 3.12, luego después de varias arquitecturas se utilizo una red con 12 neuronas de entrada, 6 neuronas en la capa oculta y 1 neurona de salida (ver la figura 4.32), la cual fue la que arrojo el menor error de prueba.

Las funciones de activación fueron lineales tanto para las neuronas de entrada como para la de salida, las función de activación para las neuronas ocultas fue la tanh. Para la fase de aprendizaje se utilizo μ =.2 y una tasa de aprendizaje α =.08, nuevamente se utilizo el podado de pesos para optimizar el modelo y eliminar los enlaces innecesarios, los pesos iniciales fueron dentro del intervalo (-.1,.1). Esta vez se detuvo el entrenamiento en la iteración número 800, esto debido a que fue el punto en donde se minimizo el error de prueba, para observar esto veamos la figura 4.33.

El error de prueba fue del 1% y el error de entrenamiento fue del 2%, lo

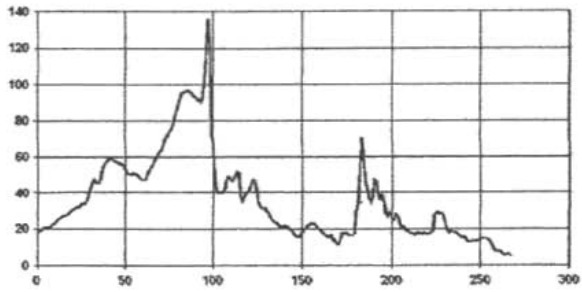


Figura 4.28: Gráfica de la serie Tiie

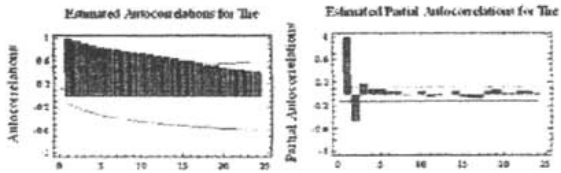


Figura 4.29: Correlogramas de la serie Tiie

cual podemos decir que es bastante bueno, aunque hay que aclarar que esta conclusión es subjetiva debido a que para otros el error óptimo bien podría ser menor (aunque no es este el caso debido a que los errores son muy bajos).

Ahora pasamos a mostrar los pronósticos arrojados tanto con el método estadístico como con redes neuronales y así concluir cual modelo se ajusto mejor a la serie. Para lograr esto observemos el cuadro 4.22

Finalmente podemos observar que el método de redes neuronales arrojo mejores pronósticos, al tener un MSE menor que al arrojado por el método estadístico, aunque en este caso la diferencia entre los modelos no es mucha.

Para finalizar se presenta la comparación gráfica entre los datos reales y las salidas de los dos modelos (ver la figura 4.34).

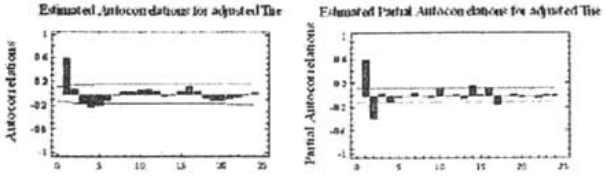


Figura 4.30: Correlogramas de la serie Tiie después de que se aplico una diferenciación simple

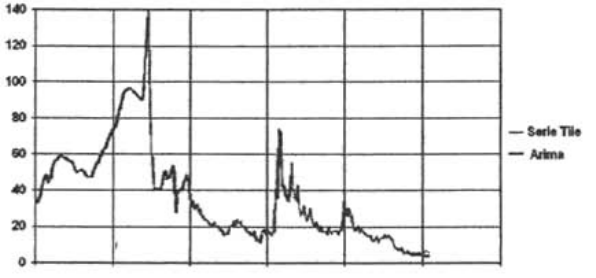


Figura 4.31: Comparación entre la serie original (Tiie) y las salidas del modelo ARIMA

Aquí podemos ver que los dos modelos se ajustaron bien a la serie.

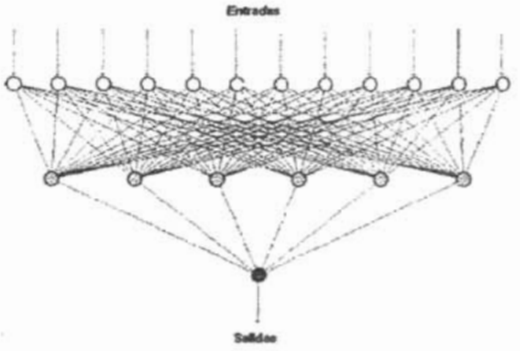


Figura 4.32: Representación de la red utilizada para la serie Tiie

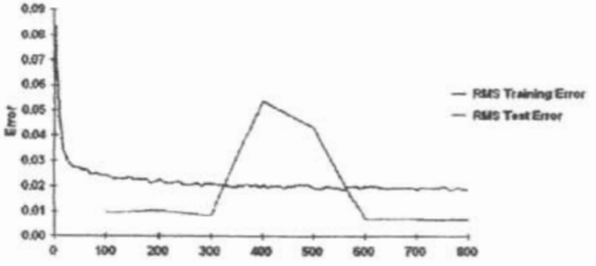


Figura 4.33: Curva de aprendizaje de la serie Tiie

Model: ARIMA(1,0,0)x(0,1,0)12 with constant

Period	Data	Forecast	Residual
12/58	336.0	339.505	-3.50479
1/59	340.0	345.247	-5.24692
2/59	318.0	327.175	-9.17456
3/59	362.0	375.659	-13.6590
4/59	348.0	358.7	-10.6997
5/59	363.0	360.813	2.18712
6/59	435.0	434.329	0.671075
7/59	491.0	481.401	9.59904
8/59	505.0	493.989	11.011
9/59	404.0	440.763	-36.7627
10/59	359.0	352.813	6.18712
11/59	310.0	320.586	-10.5865
12/59	337.0	345.665	-8.66522
1/60	360.0	346.627	13.3726
2/60	342.0	340.102	1.89776
3/60	406.0	387.36	18.6399
4/60	396.0	389.649	6.35064
5/60	420.0	407.907	12.0927
6/60	472.0	487.238	-15.2375
7/60	548.0	526.948	21.0518
8/60	559.0	557.239	1.76245
9/60	463.0	453.794	9.20586
10/60	407.0	412.865	-5.86548
11/60	362.0	354.907	7.09266
12/60	405.0	385.163	19.8349

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
1/61	421.197	400.738	441.656
2/61	397.656	371.269	424.042
3/61	457.143	427.47	486.815
4/61	443.467	411.909	475.025
5/61	464.473	431.554	497.392
6/61	514.935	480.909	547.761
7/61	588.049	553.798	622.3
8/61	597.491	562.837	632.026
9/61	500.114	465.294	534.934
10/61	443.041	408.071	478.011
11/61	397.167	362.099	432.235
12/61	433.455	404.382	474.369

Cuadro 4.12: Salidas del modelo ARIMA$(1,0,0)x(0,1,0)_{12}$, así como sus pronósticos para el siguiente año (serie pasajeros)

t	Valores originales	RNA's	ARIMA
133	417	415.83	424.721
134	391	409.64	408.18
135	419	446.32	472.474
136	461	448.56	463.055
137	472	467.16	487.565
138	535	520.68	540.092
139	622	595.56	616.616
140	606	597.04	628.14
141	508	522.28	532.864
142	461	446.02	477.188
143	390	406.35	432.712
144	432	434.66	476.236
		MSE RNA	MSE ARIMA
		1479.88051	4328.65278

Cuadro 4.13: Comparación de los pronósticos arrojados por ambos métodos y la serie original (serie pasajeros)

t	ventas_revistas	t	ventas_revistas	t	ventas_revistas	t	ventas_revistas
1	116,774	24	119,750	47	97,197	70	99,156
2	113,686	25	111,133	48	93,969	71	92,717
3	114,108	26	115,380	49	88,670	72	106,600
4	103,038	27	116,531	50	100,372	73	104,216
5	100,494	28	112,479	51	88,174	74	109,540
6	104,245	29	118,313	52	108,059	75	129,674
7	101,051	30	137,987	53	116,774	76	119,750
8	125,109	31	124,543	54	113,686	77	111,133
9	118,770	32	113,556	55	114,108	78	115,380
10	102,307	33	104,663	56	103,038	79	116,531
11	107,238	34	113,203	57	100,494	80	112,479
12	98,150	35	104,631	58	104,245	81	118,313
13	98,360	36	103,140	59	101,051	82	137,987
14	93,698	37	101,019	60	125,109	83	124,543
15	107,777	38	101,808	61	118,770	84	113,556
16	113,248	39	99,453	62	102,307	85	104,663
17	99,413	40	101,176	63	107,238	86	113,203
18	99,156	41	104,541	64	98,150	87	104,631
19	92,717	42	95,909	65	98,360	88	103,140
20	106,600	43	105,974	66	93,698	89	101,019
21	104,216	44	91,197	67	107,777	90	101,808
22	109,540	45	96,992	68	113,248	91	99,453
23	129,674	46	97,861	69	99,413		

Cuadro 4.14: Serie de tiempo para la serie revistas

```
Model Comparison
----------------
Data variable: Ventas
Number of observations = 91
Start index = 01/01/50
Sampling interval = 1.0 day(s)
Length of seasonality = 7

Models
------
(A) ARIMA(1,0,0)x(0,0,1)7 with constant
(B) ARIMA(1,0,1)x(0,0,1)7 with constant
(C) ARIMA(0,0,1)x(0,0,1)7 with constant
(D) ARIMA(0,0,1)x(1,0,0)7 with constant
(E) ARIMA(1,0,0)x(1,0,1)7 with constant

Estimation Period
Model  MSE          MAE          MAPE         ME           MPE
-------------------------------------------------------------------
(A)    5.64383E7    5849.54      5.41518      -1.31342     -0.447261
(B)    5.6829E7     5845.2       5.40661      -5.37344     -0.446053
(C)    6.3071E7     6566.83      6.10843      9.04763      -0.526234
(D)    6.75807E7    6663.12      6.19557      -29.4321     -0.609177
(E)    5.653E7      5821.85      5.38681      -1.56054     -0.422385

Model  RMSE         RUNS  RUNM  AUTO  MEAN  VAR
-----------------------------------------------
(A)    7512.54      OK    OK    OK    OK    OK
(B)    7538.5       OK    OK    *     OK    OK
(C)    7941.72      OK    **    ***   OK    OK
(D)    8220.75      ***   **    ***   OK    OK
(E)    7518.64      OK    OK    *     OK    OK
```

Cuadro 4.15: Comparación de diferentes modelos para la serie revistas

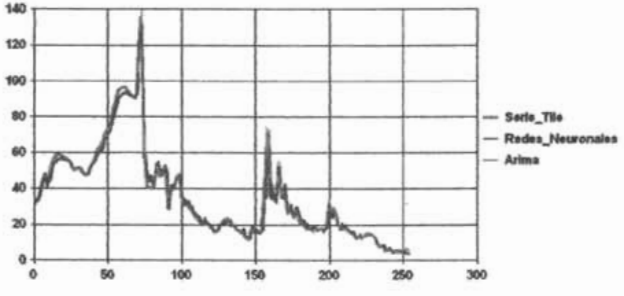


Figura 4.34: Comparación gráfica entre la serie original (Serie Tiie) y los dos modelos

Model: ARIMA(1,0,0)x(0,0,1)7 with constant

Period	Data	Forecast	Residual
08/02/96	107977.0	107824.0	152.962
09/02/96	113246.0	106766.0	6481.63
10/02/96	99413.0	102926.0	-3512.92
11/02/96	97156.0	102292.0	-5135.65
12/02/96	87717.0	95715.8	-7998.82
13/02/96	106600.0	94910.8	11689.2
14/02/96	104218.0	105961.0	-1743.18
15/02/96	109540.0	105410.0	4129.64
16/02/96	129674.0	111909.0	17765.9
17/02/96	117730.0	116671.0	1059.32
18/02/96	111193.0	111389.0	-196.317
19/02/96	115388.0	106315.0	9064.72
20/02/96	116821.0	118195.0	-1374.19
21/02/96	112479.0	112031.0	447.642
22/02/96	116913.0	112430.0	4483.04
23/02/96	137887.0	123964.0	13923.0
24/02/96	124543.0	126052.0	-1509.43
25/02/96	119146.0	117224.0	1921.97
26/02/96	104662.0	115661.0	-10997.0
27/02/96	113203.0	104772.0	8431.36
28/02/96	104621.0	110932.0	-6310.62
29/02/96	103140.0	108463.0	-5322.77
30/02/96	103019.0	112513.0	-11493.1
31/02/96	102886.0	102830.0	421.847
01/04/96	99452.0	101911.0	-2457.99

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
02/04/96	96719.2	81714.6	111724.0
03/04/96	102274.0	87797.9	116732.0
04/04/96	102658.0	84381.2	120936.0
05/04/96	101304.0	81949.8	120658.0

Cuadro 4.16: Salidas del modelo ARIMA$(1,0,0)x(0,0,1)_7$, así como sus pronósticos para el siguiente mes (serie revistas)

Serie original	Salida de la red
107777	108329.1
113248	114459.5
99413	101507.7
99166	102636.9
92717	93961.4
109800	106253.8
104216	100522.2
109540	110470.8
129674	130862.9
119750	119553.4
111133	107949.6
115380	114503.1
116531	115611.5
112479	117518.2
118313	116517.9
132987	134775.0
124643	122729.8
113566	114777.5
104663	104244.1
113203	114768.3
104631	104009.8
103140	102907.9
101019	101191.5
101008	98560.0
99453	105724.0

Cuadro 4.17: Comparación de las salidas deseadas con las salidas de la red de la serie revistas

t	Serie original	RNA's	ARIMA
92	101,176	99926	96719.2
93	104,541	101502	105274
94	95,909	97730	102668
95	105,974	102665	101504
		MSE RNA	MSE ARIMA
		12531769.6	42966128.1

Cuadro 4.18: Comparación de los pronósticos arrojados por ambos métodos red (serie revistas)

t	Tiie	t	Tiie	t	Tiie	t	Tiie	t	Tiie	t	Tiie	t	Tiie	t	Tiie	t	Tiie	t	Tiie
1	17.9	29	36.26	57	50.6	85	95.89	113	49.15	141	21.72	169	13.22	197	29.38	225	27.54	253	14.88
2	19.39	30	39.59	58	49.34	86	96.2	114	51.97	142	21.29	170	11.96	198	27.05	226	29.28	254	14.83
3	19.2	31	43.23	59	48.31	87	96.25	115	51.5	143	20.52	171	11.53	199	28.18	227	27.76	255	14.08
4	19.83	32	46.42	60	47.54	88	95.79	116	38.12	144	19.85	172	14.16	200	27.52	228	28.56	256	13.32
5	20.39	33	47.88	61	47.17	89	94.79	117	35.24	145	18.98	173	17.03	201	24.92	229	29.31	257	11.79
6	20.47	34	45.89	62	47.23	90	93.78	118	37.4	146	18.22	174	17.16	202	25.04	230	28.9	258	9.47
7	20.53	35	45.51	63	49.36	91	92.91	119	39.51	147	16.6	175	17.82	203	28.03	231	22.84	259	8.24
8	20.62	36	48.12	64	51.93	92	92.15	120	40.11	148	15.75	176	17.16	204	26.97	232	19.16	260	7.45
9	21.91	37	50.29	65	53.76	93	91.02	121	42.08	149	15.56	177	18.73	205	24.09	233	17.82	261	7.52
10	22.42	38	54.24	66	54.92	94	90.3	122	44.87	150	16.01	178	15.86	206	21.06	234	16.82	262	7.89
11	22.77	39	56.16	67	57	95	92.37	123	47.15	151	18.07	179	16.34	207	21.1	235	18.08	263	8.56
12	24.25	40	57.21	68	59.06	96	104.3	124	47.2	152	19.54	180	16.86	208	21.67	236	18.17	264	5.81
13	25.46	41	58.14	69	60.98	97	122.5	125	42.62	153	20.16	181	29.87	209	18.73	237	17.94	265	5.52
14	25.98	42	58.63	70	62.29	98	135.9	126	35.16	154	21.86	182	35.98	210	18.78	238	17.25	266	5.01
15	26.59	43	58.73	71	63.39	99	117.2	127	33.05	155	21.79	183	56.82	211	18.05	239	16.26	267	5.82
16	26.91	44	58.23	72	65.66	100	81.03	128	31.27	156	22.76	184	70.26	212	17.34	240	15.42	268	4.86
17	27.22	45	57.78	73	68.55	101	60.59	129	31.11	157	22.79	185	57.86	213	17.16	241	15.29		
18	27.66	46	57.14	74	70.3	102	46.76	130	31.52	158	22.72	186	46.39	214	16.56	242	15.18		
19	28.42	47	56.82	75	71.79	103	40.72	131	29.56	159	21.31	187	41.42	215	17.74	243	13.67		
20	29.5	48	56.44	76	73.48	104	39.9	132	29.23	160	20.16	188	37.1	216	17.79	244	12.48		
21	30.45	49	55.95	77	75.02	105	39.9	133	27.14	161	19.75	189	34.61	217	16.98	245	12.51		
22	31.22	50	55.16	78	76.97	106	40.03	134	25.71	162	18.88	190	37.09	218	17.03	246	13.53		
23	31.77	51	53.11	79	81.36	107	41.65	135	24.29	163	17.36	191	47.54	219	17.37	247	12.98		
24	31.81	52	51.1	80	84.4	108	45.48	136	23.62	164	18.91	192	46.54	220	17.66	248	13.05		
25	32.24	53	50.12	81	87.72	109	49.37	137	23.1	165	18.18	193	40.18	221	16.85	249	13.29		
26	33.43	54	50.38	82	91.48	110	48.7	138	21.79	166	15.57	194	35.91	222	17.24	250	13.51		
27	33.67	55	50.89	83	94.19	111	47.3	139	20.99	167	19.62	195	39.12	223	17.75	251	14.44		
28	34.39	56	50.83	84	95.33	112	46.91	140	20.56	168	14.88	196	35.21	224	19.05	252	14.39		

Cuadro 4.19: Serie de tiempo para la Tiie

```
Model Comparison
----------------
Data variable: Tiie
Number of observations = 268
Start index = 1/50
Sampling interval = 1.0 month(s)
Length of seasonality = 12

Models
------
(A) ARIMA(2,1,1) with constant
(B) ARIMA(2,1,0) with constant
(C) Linear trend = 57.9011 + -0.159109 t
(D) Simple moving average of 5 terms
(E) Simple exponential smoothing with alpha = 0.9999

Estimation Period
Model  MSE          MAE          MAPE         ME            MPE
-------------------------------------------------------------------------
(A)    10.7287      1.55291      4.71431      -0.000189043 -0.243483
(B)    10.699       1.57742      4.72753      -0.00039688  -0.300826
(C)    442.691      14.5519      49.4568      -6.28353E-15 -29.2127
(D)    78.9347      4.75902      13.5829      -0.160525    -3.84823
(E)    19.4571      2.11514      5.82526      -0.0406614   -0.910731

Model  RMSE         RUNS  RUNM  AUTO  MEAN  VAR
-----------------------------------------------
(A)    3.27547      OK    OK    OK    OK    OK
(B)    3.27093      OK    *     OK    OK    OK
(C)    21.0402      ***   ***   ***   ***   ***
(D)    8.88452      ***   ***   ***   OK    ***
(E)    4.41102      ***   ***   ***   OK    ***
```

Cuadro 4.20: Comparación de diferentes modelos para la serie Tiie

Model: ARIMA(2,1,1) with constant

Period	Data	Forecast	Residual
4/70	12.48	12.3978	0.0822132
5/70	12.11	12.1383	0.271716
6/70	13.13	12.9205	0.20947
7/70	12.98	14.275	-1.29502
8/70	13.05	12.0158	1.03423
9/70	13.29	13.4478	-0.157793
10/70	13.31	13.3134	0.156612
11/70	14.44	13.5821	0.85754
12/70	14.39	15.1274	-0.737335
1/71	14.66	13.8945	0.765534
2/71	14.63	14.9826	-0.352389
3/71	14.08	14.4221	-0.342112
4/71	13.32	12.6281	-0.308095
5/71	11.79	12.891	-1.10097
6/71	9.49	10.7541	-1.28408
7/71	8.24	8.11001	0.129987
8/71	7.45	8.1165	-0.666495
9/71	7.52	7.11864	0.401364
10/71	7.49	7.44747	0.0425345
11/71	6.58	8.08269	-1.50269
12/71	5.61	5.17692	0.53908
1/72	5.52	5.7861	-0.260898
2/72	5.61	5.48362	0.126272
3/72	5.62	6.49272	-0.872713
4/72	6.66	5.27162	-0.313625

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
5/72	4.13361	-2.2939	10.5611
6/72	3.94984	-9.67602	17.5657
7/72	3.89142	-15.0299	23.2928
8/72	4.0017	-18.3456	27.409
9/72	4.02998	-22.4226	30.4827
10/72	3.9649	-25.6664	32.9962
11/72	3.88969	-27.506	35.2054
12/72	3.8486	-29.8096	37.491
1/73	3.78631	-31.9892	39.5659
2/73	3.73662	-34.031	41.5242
3/73	3.68862	-36.0076	43.385
4/73	3.63768	-37.8748	45.1302

Cuadro 4.21: Salidas del modelo ARIMA(2,1,1), así como sus pronósticos para el siguiente año (serie Tiie)

t	valores originales	redes	arima
269	5.02	4.63	4.15361
270	5.37	4.94	3.94484
271	5.56	5.32	3.98143
272	5.05	5.60	4.0317
273	5.17	5.68	4.02008
274	5.48	5.50	3.9649
275	5.37	5.06	3.89969
276	5.13	4.61	3.8406
277	5.37	4.32	3.78831
278	6.26	4.11	3.73862
279	6.38	3.91	3.68862
280	5.89	3.78	3.63768
		MSE RNA	MSE ARIMA
		8.8008095	17.4641554

Cuadro 4.22: Comparación de los pronósticos arrojados por los dos modelos y la serie original (Serie Tiie)

Conclusiones

Una vez expuesta la teoría y algunas aplicaciones de los métodos de redes neuronales y de los modelos autorregresivos y de promedios móviles integrados (ARIMA), se puede concluir que efectivamente las redes neuronales se pueden utilizar con el fin de pronóstico de series de tiempo, cuya precisión depende de un buen entrenamiento, en el cual entre menor sea el error de prueba, mejores serían los pronósticos. Con esto se demuestra la hipótesis planteada en este trabajo en el sentido de que las RNA's pueden obtener resultados comparables a los métodos estadísticos tradicionales, obteniendo en algunos casos incluso ajustes superiores. En este trabajo se han hecho algunos experimentos que sustentan estas afirmaciones, sin pretender dar resultados de carácter general relacionados a una forma de modelar que resulte más efectiva que otra. Para ello, considero que se abre un tema que se encuentra fuera del alcance de este trabajo y que consistiría en hacer un análisis estadístico exhaustivo, ajustando miles de series con ambas metodologías y comparar la bondad estadística de cada una de ellas.

Las ventajas que ofrece el trabajar con RNA's se refieren al hecho de no suponer que siguen alguna distribución en particular, no sólo para el conjunto de entrenamiento, sino para los residuales (por ejemplo normalidad o ruido blanco), como sucede con la mayoría de los métodos estadísticos paramétricos tradicionales. Otra ventaja que ofrecen las RNA's se deriva del hecho de no preocuparse por el problema de estacionalidad, esto debido a que las RNA's pueden aprender bien este comportamiento (aunque también se puede hacer un reprocesamiento de los datos para tratarla). En cambio, al aplicar modelos ARIMA se tiene que hacer una diferenciación estacional (o introducir parámetros estacionales al modelo) para poder hacer estacionaria la serie.

En relación con la tendencia, ambas técnicas exigen la eliminación de ésta ya que, de lo contrario, las salidas de los modelos quedan muy por debajo de las salidas deseadas y, por lo tanto, los pronósticos tendrán este mismo problema.

Un problema que plantean las RNA's se refiere a la arquitectura adecuada que se debe de utilizar, ya que aunque existen aproximaciones heurísticas para definir el número de neuronas en la capa oculta, no existe un método que establezca de manera categórica el número óptimo de éstas para cada problema. Es muy importante en este sentido, tomar en cuenta la práctica y la experiencia como auxiliares en la formulación de una arquitectura adecuada.

Por tanto, es importante darnos cuenta del enorme terreno que están ganando algunas técnicas relativamente recientes, como las RNA's, en problemas tan relevantes como el pronóstico de series de tiempo, entre otros, con resultados muy satisfactorios.

Bibliografía

[1] Aguirre Jaime Armando ***Introducción al tratamiento de series temporales***

Díaz de Santos, Madrid, 1994

[2] Bañuelos Rodríguez Gladys ***Estudio de la complejidad de las series de tiempo, de los mercados financieros, por medio de redes neuronales***

Tesis, UNAM , México2002

[3] Brockwell J. Peter y Richard A. Davis ***Introduction to Time Series and Forescasting***

Springer, EU, 1996, primera edición

[4] Bowerman Connell ***Forescasting and Time Series an applied approach***

Duxbury Press, EU, 1993

[5] Calero Vinelo Arístides ***Estadística***

Ediciones Politécnico, México, 1998

[6] Castillo Enrique, Angel Cobo, José Manuel Gutiérrez y Rosa Eva Pruneda ***Introducción a las Redes Funcionales con aplicaciones***

Paraninfo, España, 1999

[7] CORCHADO JUAN MANUEL, FERNANDO DÍAZ, LOURDES BOURAJO Y FLORENTINO FERNÁNDEZ ***Redes Neuronales Artificiales un enfoque practico***

Norgráfica, España, 2000

[8] DIEBOLD X. FRANCIS ***Elementos de pronósticos***

International Thomson Editores, EU, 1999

[9] DOUGLAS HAMILTON JAMES ***Time Series Analysis***

Princeton, EU, 1994

[10] E.P. BOX GEORGE, GWILYM M. JENKINS Y GREGORY C. REINSEL ***Time Series Analysis***

Prentice Hall, tercera edición, EU, 1994

[11] GUERRERO GUZMÁN VÍCTOR MANUEL ***Análisis Estadístico de series de tiempo económicas***

Thomson, segunda edición, Mexico, 2003

[12] HANKE E. JOHN ARTHUR G. REITSCH ***Pronósticos en los negocios***

Prentice Hall, Eastern Washington University, 1996

[13] HASSER NORMAN B., JOSEPH P. LA SALLE Y JOSEPH A. SULLIVAN ***Análisis Matemático***

vol 1, Trillas, Mexico, 1999

[14] HERNÁNDEZ GODINEZ GABRIELA ***Apuntes para un curso sobre series de tiempo a nivel profesional***

tesis, UNAM, Mexico, 2003

[15] HILERA JOSÉ R. Y VÍCTOR J. MARTÍNEZ ***Redes Neuronales Artificiales fundamentos, modelos y aplicaciones***

Alfaomega Ra-Ma, España, 1995

[16] JUGGE G. JUGGE ***Introduction to the theory and practice of econometrics***

Wiley, Canada, 1988

[17] LAWRENCE JEANNETTE ***Introduction to Neural Networks design, theory and applications***

California Scientific Software, EU, 1994

[18] LAWRENCE STEPHEN ARNOLD Y H. FRIEDBERG INSEL SPENCE ***Linear algebra***

Prentice Hall, EU, 1989

[19] LAU CLIFFORD ***Neural Networks theorical foundations and analysis***

IEEE press, EU, 1991

[20] LYENGAR SITHARAMA S., E.C CHO Y VIR V. PHOHA ***Foundations of Wavelet Networks and Applications***

Chapman y Hall/CRC, EU,2002

[21] M. BISHOP CHRISTOPHER ***Neural Networks for pattern recognition***

Clarenden press Oxford, EU, 1995

[22] M. SKAPURA DAVID ***Building Neural Networks***

Adison Wesley, EU, 1996

[23] MADDALA G.S. ***Introducción a la Econometría***

Prentice Hall, EU, 1996

[24] MAKRIDAKIS SPYROS Y STEVEN C. WHEELWRIGHT ***Forescasting Methods for Management***

Wiley, quinta edición, EU, 1989

[25] MANO M. MORRIS ***Logica Digital y Diseño de Computadoras***

Prentice Hall, EU, 1982

[26] MARTÍN DEL BRÍO BONIFACIO Y ALFREDO SANZ MOLINA ***Redes Neuronales y Sistemas difusos***

Alfaomega Ra-Ma, segunda edición, España, 1997

[27] MENDEN HALL WILLIAM ***Probabilidad y Estadística***
Prentice Hall, cuarta edición, EU, 1995

[28] MORILLA RODRÍGUEZ CARMEN ***Análisis de series temporales***
La muralla, España, 2000

[29] OLMEDA IGNACIO Y SERGIO BARTA ROMERO ***Redes Neuronales Artificiales fundamentos y aplicaćiones***
CICAI, España, 1993

[30] PENA TRAPERO J.BERNARDO Y JULIO A. ESTAVILLO DORADO ***Cien ejercicios de Econometría***
Pirámide, España, 1999

[31] RZEMPOLUCK EDWARD J. ***Neural Networks data analysis using simulnet***
Springer, Canada, 1997

[32] WIE A. WILLIAM ***Análisis de series temporales***
La muralla, España, 2000

[33] YAFFE ROBERT Y MONNIE MCGEE ***Introduction to Time Series Analysis and Forescasting with applications of SAS and SPSS***
Academic press, EU, 2000