

24021
54



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
CAMPUS ACATLAN

"LA DISPONIBILIDAD DE TRENES DE LA LÍNEA 1 Y LÍNEA 5
DEL SISTEMA DE TRANSPORTE COLECTIVO METRO DE LA
CIUDAD DE MEXICO PARA CUBRIR LA DEMANDA DE
USUARIOS".

T E S I S

PARA OBTENER EL TÍTULO DE:
**LICENCIADO EN MATEMÁTICAS APLICADAS Y
C O M P U T A C I O N**
P R E S E N T A :

JORGE ALBERTO ZUÑIGA HERRERA



Asesor: Act. María del Carmen González Videgaray

MAYO 2003

A



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

***El honor prohíbe acciones
que la ley tolera.***

***El gobierno más difícil es
el de uno mismo.***

***Lo que has de decir; antes
que a otro, dítelo a ti mismo.***

***La parte de nuestro cuerpo
más sana, es la que más se ejercita.***

SÉNECA

Agradecimientos

**A Dios nuestro señor por darme
vida y la oportunidad de
esforzarme para lograr
mis objetivos hasta este
punto de mi vida.**

**Para ti mamá , no se que
escribir solo que te quiero
y te extraño, donde quiera
que estés, gracias.**

**Gracias papá y
hermana por su apoyo.**

**A todo el personal de la UNAM
que colaboraron con mi formación.**

**Al personal del STCM
de la Ciudad de México
por el apoyo.**

Y muy en especial a ti.

**TESIS CON
FALLA DE ORIGEN**

C

Contenido

Introducción.....	4
-------------------	---

Capítulo I

Sistema de Pilotaje Automático

1.1 Antecedentes.....	7
1.1.2 Equipo fijo.....	8
1.1.3 Equipo de embarcado.....	8
1.1.4 Qué funciones realiza en PA.....	9
1.2 Cuáles son los modos de conducción para los trenes.....	10
1.3 La señalización del Sistema de Transporte Colectivo Metro.....	11
1.4 Antecedentes del número de fallos en material rodante.....	12
1.5 En caso de fallo en PA.....	13
1.6 Crecimiento del Sistema de Transporte Colectivo Metro	13
1.7 Servicio del Sistema de Transporte Colectivo Metro	14
1.8 Afluencia de usuarios en toda la red.....	14
1.9 Mantenimiento del Sistema de Transporte Colectivo Metro.....	15

Capítulo II

Modelo para la proyección: series de tiempo

2.1 Qué es una serie de tiempo.....	17
2.2 Qué tipos de serie históricas existen.....	17
2.3 Objetivo del análisis de series de tiempo.....	18
2.4 Enfoque de Box y Jenkins.....	18
2.5 Modelo autorregresivo.....	25
2.6 Modelo de promedios móviles.....	30
2.7 Modelo autorregresivo y de promedios móviles.....	32

2.8 Modelo autorregresivo e integrado de promedios móviles.....	34
2.9 Series de tiempo estacionales.....	36
2.10 Operadores y modelos estacionales.....	37
2.11 Modelo multiplicativo estacional.....	38
2.12 Identificación del modelo estacionario.....	39
2.13 Identificación del modelo estacional.....	40

Capítulo III

Selección y Diagnostico del modelo para la proyección.

3.1 Metodología Box y Jenkins.....	44
3.2 Identificación del modelo.....	45
3.3 Estabilización de la varianza.....	45
3.4 Estabilización de la tendencia de la serie.....	46
3.5 Análisis de la función de autocorrelación y autocorrelación parcial.....	47
3.6 Estimación por momento: valores parametrales iniciales.....	50
3.7 Verificación: ¿ es adecuado el modelo seleccionado?.....	52
3.8 Análisis de residuales.....	52
3.9 Pronóstico.....	60
3.10 Intervalo de confianza para los pronósticos.....	61

Capitulo IV

Aplicación del modelo y toma de decisiones

4.1 Metodología para obtener los fallos.....	62
4.2 La serie de tiempo de la Línea 1.....	63
4.3 Identificación de la serie histórica para la Línea 1.....	63
4.4 Estimación de parámetros para la Línea 1.....	70
4.5 Análisis de residuales para la Línea 1.....	70
4.6 La Serie de tiempo de la Línea 5.....	76

TESIS CON
 FALLA DE ORIGEN

2

4.7 Identificación de la serie histórica para la Línea 5.....	76
4.8 Estimación de parámetros para la Línea 5.....	83
4.9 Análisis de residuales para la Línea 5.....	83
4.10 La disponibilidad de trenes en las Líneas 1 y 5.....	89
Conclusiones.....	99
Bibliografía.....	102
Anexos I (Línea 1).....	103
Anexos II (Línea 5).....	107

Planteamiento del Problema:

En el Sistema de Transporte Colectivo Metro (STCM) de la Ciudad de México la ocurrencia de un fallo en el Pilotaje Automático (PA), en los trenes de la Línea 1 y Línea 5, afecta en todo su recorrido al prestar el servicio, ocasionando un mayor retardo en la traslación de un tren a una estación, también sus condiciones de seguridad estarán disminuidas. Estos fallos en el PA ocasionara aglomeraciones de usuarios en las horas pico, pérdida de tiempo durante el servicio, desgaste de instalaciones, costos y con la falta de trenes la demanda del servicio estará afectada por la indisponibilidad de trenes en cada una de las dos Líneas.

Justificación:

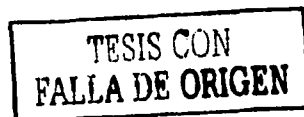
El STCM de la Ciudad de México tiene bajo su responsabilidad la atención de usuarios, su transportación, su seguridad y si presenta algún fallo alguno de estos trenes para éstas dos Líneas estarán afectadas sus responsabilidades durante su recorrido al prestar el servicio, por eso un mantenimiento oportuno, reducirá la cantidad de usuarios afectados al utilizar este medio de transporte.

Objetivo General:

En esta investigación se hace un análisis comparativo entre el número de trenes disponibles de Línea 1 y Línea 5, con respecto a la proyección de la afluencia de usuarios para una mejor toma de decisiones en el mantenimiento correctivo y preventivo de los trenes del sistema.

Objetivo Secundario:

Se calcula la cantidad de gente que será afectada por los fallos en PA durante su recorrido al realizar sus actividades cotidianas, basándose en la proyección de



usuarios realizada por el STCM junto con los fallos pronosticados por la metodología de Box y Jenkins, para un periodo de seis meses

Hipótesis:

Si se consideran estos fallos como una serie de tiempo, entonces existirá la posibilidad una dependencia a través del tiempo, y el "deterioro" de los trenes puede ser pronosticado mediante la metodología de Box y Jenkins, y así se tendrá una herramienta para tomar la decisión de cuando brindar el mantenimiento preventivo o correctivo a los trenes de estas dos Líneas, y con ello reducir la cantidad de usuarios que son afectados

Se selecciono dos series históricas una para la Línea 1 y otra para la Línea 5, para pronosticar el comportamiento de los fallos en el PA, y así ver la disponibilidad de trenes que tienen estas dos Líneas para cubrir la demanda del servicio, respecto a la proyección de afluencia de usuarios realizada por el STCM para el año 2002.

Está investigación el enfoque que se utiliza es el Modelo Autorregresivo Integrado de Media Móvil (ARIMA), que no asume ninguna tendencia particular en los datos históricos de la serie a pronosticar, utiliza un enfoque iterativo de identificación de un modelo útil a partir de modelos de tipo general. Existen dos enfoques básicos para el análisis de series de tiempo: el clásico y el estocástico, la teoría que presenta en este trabajo versa sobre series univariadas con una representación lineal.

TESIS CON
FALLA DE ORIGEN

Estructura de la Investigación:

La organización de esta investigación es la siguiente, en primer capítulo se presenta lo que es el Pilotaje Automático (PA) y sus antecedentes, como los fallos representan un problema de indisponibilidad de trenes para la Línea 1 y Línea 5 del STCM de la Ciudad de México; En el segundo capítulo, se presentan los diferentes modelos según Box y Jenkins; en el tercer capítulo, la selección y diagnóstico del modelo para la proyección de fallos de los trenes, en el cuarto capítulo se explica la aplicación del modelo y la toma de decisiones obtenida por el resultado del modelo y las conclusiones obtenidas.

Se utiliza el paquete estadístico STATGRAPHICS PLUS versión 4.0.

REGIS CON
FALLA DE ORIGEN

Capítulo I

Sistema de Pilotaje Automático.

1.1 Antecedentes

El Metro de la Ciudad de México es un transporte que lleva consigo tecnología de la más avanzada y las ventajas propias de un servicio eléctrico. La red del Metro trabaja con sistema de Pilotaje Automático (PA) implantado a raíz del accidente del 20 de octubre de 1975. Mediante el PA, la distancia entre los trenes, lo mismo su velocidad y frenado; se gobiernan de manera automatizada. Debido a las necesidades de la Ciudad de México, y a la explotación y regulación de la línea, se instala un equipo electrónico de (PA) que se encarga de vigilar la seguridad de los trenes y los usuarios.

El Sistema de PA autoriza la circulación de los trenes del Sistema de Transporte Colectivo Metro (STCM), de la Ciudad de México cuando se reúnen el máximo de condiciones de seguridad, el pilotaje automático permite una uniformación del arranque de los trenes en forma automática por medio de un programa de velocidad impuesto al tren, se obtiene entonces una optimización de la explotación de la línea, respetando los límites físicos de los materiales y normas de seguridad, además asegura la buena marcha de los trenes sin importar la carga y verifica que las condiciones de seguridad sean cumplidas; permite controlar la velocidad en un valor fijo evitando la influencia de la carga que tenga el tren sobre la precisión y el confort al detenerse el tren.

El Pilotaje Automático depende de las condiciones de la señalización y tiene también una relación con el tráfico, ya que el equipo puede variar su traslación del tren dependiendo de qué tan saturado esta la línea.

También este sistema proporciona un modo de conducción automático, confiando a los equipos y dispositivos del tren la ejecución de funciones repetitivas la seguridad de su funcionamiento se basa, en la permanencia del enlace entre la vía y el tren. A falta de algún de estos parámetros debe pararse automáticamente.

El Sistema de Pilotaje Automático llamado equipo de embarcado, que esta colocado a bordo del tren y se encarga de captar y procesar las informaciones del programa para traducirlos en instrucciones para el tren, está constituido por un programa inscrito en la vía (equipo fijo), y por un dispositivo electrónico (equipo embarcado), dentro del tren, el cual mantiene su velocidad generando sobre los equipos de tracción y frenado propios del tren.

El dispositivo del Sistema de Pilotaje Automático comprende dos conjuntos:

- 1) Instalaciones fijas.
- 2) Equipo de embarcado.

1.1.2 Equipo Fijo

Se localiza a lo largo de las vías, está constituido por un conjunto de diferentes módulos electrónicos separados, este equipo recibe la información de P.C.C (Puesto Central de Control), generándose a su vez la alimentación al programa de arranque del tren.

1.1.3 Equipo de embarcado

Al hablar de equipo de embarcado se hace alusión al que se localiza a bordo de los trenes.

1.1.4 Qué funciones realiza en PA

La función del equipo de embarcado es la de recibir e interpretar las informaciones provenientes del equipo fijo, y así poder efectuar automáticamente las funciones que debe realizar (salida, parada del tren, respetar las velocidades establecidas, apertura de puertas, etc.).

El equipo de embarcado interpreta las señales recibidas y las traduce en funciones, para los órganos del tren.

Los equipos PA se encuentran enlazados con los órganos de tracción, frenado, detección y velocidad (puertas, cabina de conducción) del tren. Estos equipos realizan:

- La decodificación y la interpretación de datos.
- La seguridad a bordo del tren.

El Sistema de Pilotaje Automático debe cumplir en forma segura a partir de las informaciones del programa, las siguientes funciones:

- Salida del tren en Estación.
- Circulación en intersecciones respetando las limitaciones de velocidad y señalización establecidas por las condiciones locales de operación.
- Arribo y parada del tren en las estaciones.
- Maniobras de cambio de vías en las Terminales y Servicios Provisionales.
- Autorización de apertura de puertas a vagones.

1.2 Cuáles son los modos de conducción para los trenes

Los modos de conducción que usan los trenes para las Líneas 1 y 5 del STCM de la Ciudad de México, equipados con Pilotaje Automático (PA) son, por orden prioritario:

- 1 Pilotaje Automático (PA).
- 2 Conducción Manual Controlada (CMC).
- 3 Conducción Manual Limitada (CML).
- 4 Conducción Manual Registrada (CMR).
- 5 Conducción Libre a Tracción 2 (CL.-T2).

Los trenes circulan normalmente en Pilotaje Automático, en este modo de conducción, después de haber recibido la autorización de la salida del tren, la acción del conductor se limita a cerrar las puertas y ordenar el arranque del tren.

El tren circulará automáticamente hasta la siguiente estación respetando las limitaciones de velocidad y los de señalización.

La toma de Pilotaje Automático se hace al estar en paro total el tren. es decir:

- Una vez que las condiciones de señalización lo permitan.
- En interestación, enfrente de toda señal que no indique el alto.

Si la toma en PA no ha sido ejecutada, la salida y circulación en PA será imposible de realizar; el conductor puede en todo momento confirmar el cierre de puertas, apoyándose al sonar el timbre, el tren debe estar parado, además las condiciones de señalización deben permitir la salida.

Los trenes pueden circular igualmente en conducción manual controlada (CMC), esta conducción la única diferencia es que el conductor conduce al tren, y disminuyen las condiciones de seguridad, ocasionando una disminución en la

velocidad del tren, esto afecta a todos a los demás trenes que están circulando en toda la línea, ocasionando una aglomeración de usuarios en las estaciones, pero si sobre pasa la velocidad dictada por el pilotaje del sistema de control de conducción del tren este aplica un frenado de urgencia, En los trenes en los casos, la ausencia de la señal recibida de la instalación fija provoca inmediatamente el paro del tren. Existen zonas que no están equipadas para circular en pilotaje automático ni en la Conducción Manual Controlada (CMC) sólo esta prevista para la circulación en Conducción Manual Libre (CML) .

El modo de conducción C.M.R. es el único posible de obtener en zonas no equipadas con pilotaje automático.

El modo de conducción C.L.-T2 es el único que puede obtener en cualquier punto de la vía equipada o no con pilotaje automático y no es dependiente de PA.

A partir de los datos que capta de la vía el pilotaje automático efectúa automáticamente el arranque del tren en cada estación. El PA interpreta las señales y las traduce en funciones para el tren.

1.3 La señalización del Sistema de Transporte Colectivo Metro

La señalización del Metro para las Líneas 1 y 5, son señales luminosas que indican, de acuerdo con sus colores(amarillo, verde, rojo), reducción de velocidad, marcha normal y alto en el avance del tren esta función se realiza por medio de circuitos de vías, es decir, seccionamientos de la vía en donde por cada seccionamiento hay una señal, la cual debe ser respetada en todos los puntos de recorrido del tren y en todas las circunstancias, tanto para los trenes que circulan en condición en CMC, CML como para los que circulan en pilotaje Automático.

1.4 Antecedentes del número de fallos en material rodante¹

Los fallos en 1999 del material rodante en todo su conjunto registraron, un promedio de 4.6 fallos por cada 10,000 kilómetros recorridos, los cuales se debió a diversas causas, pero principalmente a deficiencias en estos cinco sistemas:

- Pilotaje automático.
- Sistema de puertas.
- Motriz inactiva total.
- Motriz inactiva a la tracción.
- Los motocompresores de los trenes.

Año con año estos cinco sistemas son la causa del mayor índice de afectaciones al servicio originando que haya trenes indisponibles para cubrir la demanda de usuarios, hay varias razones que son la causa de estos fallos, tales como: mala calidad en alguno de ellos, problemas de mantenimiento, de ajuste y de supervisión. El parque del material rodante de la red, registra una promedio de 14 años, aunque algunos de los trenes ya cuentan con 30 años de operación y se requiere continuar su rehabilitación y modernización, por lo que se tiene que mantener programas permanentes de rehabilitación y equipamiento para los sistemas.

En las horas pico los fallos más frecuentes en los trenes se ubican principalmente en dos sistemas:

- Pilotaje Automático.
- Apertura y cierre de puertas.

¹ Plan de empresa 2000-2006; Sistema de Transporte Colectivo Metro de la Ciudad de México

1.5 En caso de fallo en PA

Evento de fallo bajo estudio es: conducción en PA.

Con la falla en este modo de conducción afectara a todos los demás trenes que estén en servicio en las Líneas 1 y 5, ya que reducen su velocidad, también tardan más tiempo de lo debido en cada estación hasta que el tren lo rehabilitan o sale de circulación, esto ocasionara una aglomeración de usuarios, perdida de tiempo, costos, también seria un tren no disponible en cada una las dos Líneas, mientras es llevado al laboratorio de mantenimiento para reparalo.

En está tesis, se utilizara la aplicación de una serie de tiempo que plantea que los fallos en PA pueden estar correlacionados y que con el tiempo puede mostrar ya sea una tendencia creciente, decreciente o constante, se analizaran los datos recopilados de la bitácora de operación del STCM de las Líneas 1 y 5. Estos fallos que ocurren en el tiempo, si se consideran como un proceso de serie de tiempo, el "deterioro" del los trenes estas dos Líneas puede estimarse.

1.6 Crecimiento del Sistema de Transporte Colectivo Metro

El STCM ha registrado un crecimiento en sus 32 años de operación de 27,861,122 de usuarios trasportados en 1969 a 1,433,671,034² de usuarios en diciembre de 2001, entre 1969 y 2002 la expansión de la red ha sido importante hasta alcanzar una extensión de 202 Kilómetros de vías dobles,1750 estaciones. El Metro de la Ciudad de México tiene un lugar sobresaliente entre los 87 trenes metropolitanos del mundo.

² Sistema de Transporte Colectivo Metro, Dirección de Operación , Gerencia de ingeniería y desarrollo

1.7 Servicio del Sistema de Transporte Colectivo Metro

El STCM constituye el medio eficaz del transporte urbano de la ciudad de México, la gran aceptación con que cuenta por parte de los habitantes, ha permitido que en sus 32 años de servicio se hayan brindado atención a casi 34,487,239,999² usuarios. Durante 2001 el STCM atendió el desplazamiento de 1,433,671,034 pasajeros en las nueve líneas de la red. Las metas de transportación de usuarios para el STCM de la línea 1 y 5 son de 271,900,734 y 78,862,801 respectivamente, para todo este año de 2002, según las estimaciones de la gerencia de ingeniería y desarrollo del STCM. La Línea 1 presenta niveles de afluencia de 827,168 de pasajeros, mientras que la Línea 5 su afluencia de usuarios es de 233,141 de pasajeros, en día laborable.

1.8 Afluencia de usuarios en toda la red

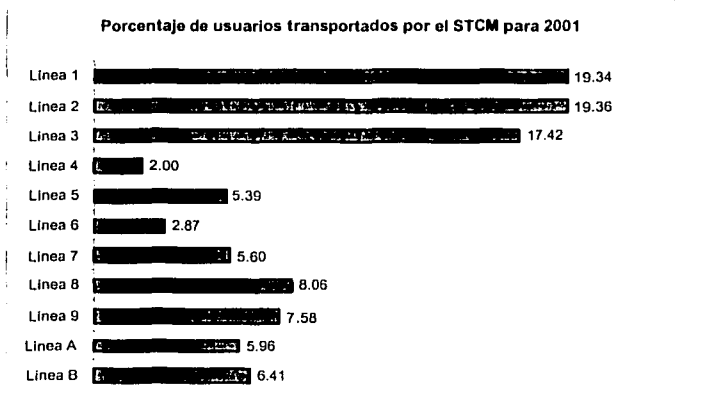
La distribución de usuarios en todo el STCM para el año 2001 es:

Tabla 1
Usuarios transportados por el STCM para 2001.

Línea 1	277,300,804
Línea 2	277,583,973
Línea 3	249,714,499
Línea 4	28,726,317
Línea 5	77,335,281
Línea 6	41,205,501
Línea 7	80,243,293
Línea 8	115,484,156
Línea 9	108,714,223
Línea A	85,503,064
Línea B	91,859,923

**TESIS CON
FALLA DE ORIGEN**

La tabla anterior muestra que la Línea 1 ocupa el segundo lugar de afluencia de usuarios en cambio la Línea 5 esta en noveno lugar en toda la red del STCM para el año de 2001, el porcentaje de participación de estas dos Línea se muestra en la siguiente gráfica:



Para el año de 2001 se tuvo una afluencia de usuarios para todos los siete días de la semana de 1,433,671,034 usuarios con la participación de la Línea 1 del 19%, como lo muestra la anterior gráfica, que cuenta con veinte estaciones con terminal de Pantitlán a Observatorio y la Línea 5 con un 5% del total de la red y esta tiene 13 estaciones con terminal de Pantitlán a Politécnico, como se observa en la anterior gráfica.

1.9 Mantenimiento del Sistema de transporte Colectivo Metro

Para asegurar la disponibilidad de los trenes y su seguridad en el servicio en las Líneas 1 y 5, el organismo realiza permanentemente acciones orientadas a elevar

**TESIS CON
FALLA DE ORIGEN**

la eficiencia y la productividad³ de sus funciones. Entre estas labores la de mantenimiento es primordial para garantizar las condiciones óptimas de operación del material rodante. Por la magnitud de las intervenciones que se deben ejecutar el mantenimiento preventivo se clasifica en menor y mayor de acuerdo con el tiempo de operación o kilometraje recorrido por los trenes. En ambos casos se subdivide en sistemático y cíclico de acuerdo con su aplicación a unidades completas o partes y equipos de los trenes, los proyectos de conservación y mantenimiento del material rodante son de vital importancia ya que en función de su aplicación se determina el tiempo que duren en operación los trenes en las líneas y así puedan prestar un servicio seguro, eficiente y cómodo. Por su carácter contingente, el mantenimiento correctivo no es programable, sino que se realiza cuando un tren presenta un fallo durante el servicio y su objetivo fundamental es restablecer el material rodante a la operación en el menor tiempo posible.

³ la productividad es los usuarios transportados por carro, pasajeros atendidos por estación, pasajeros transportados por empleado, empleados por estación.

Capítulo II

Modelo para la proyección: series de tiempo.

2.1 Qué es una serie tiempo

Una serie de tiempo es una colección de observaciones generadas en forma secuencial a través del tiempo.

2.2 Qué tipos de serie históricas existen

En la práctica se puede encontrar distintos tipos de series de tiempo: estacionarias, no estacionarias, estacionales y no estacionales, que más adelante se presentarán sus características. También existen series multivariadas y univariadas, así como las lineales y no lineales. Las series multivariadas son aquellas cuyas observaciones dependen o son el resultado de combinar los valores tomados por otras series temporales; Las univariadas son las que los valores no dependen de otras series sino sólo del tiempo, en cuanto a las series lineales, puedo decir que son aquellas en cuya representación no se requieren variables dependientes del tiempo elevadas a una potencia mayor a 1.

Cuando las series de tiempo pueden ser registradas continuamente en el tiempo se les llama series continuas y a las que sólo toman valores en intervalos de tiempo específico se les llama series de tiempo discretas.

El uso más común de una serie de tiempo es la predicción de valores, tanto futuros como pasados; entre los cuales los primeros son los mas requeridos ya que las decisiones para la planeación en casi cualquier campo se basan en lo que espera ocurrirá en el futuro. La aplicación de la metodología para el análisis de las

series de tiempo se da en muy diversas áreas, las siguientes son algunos ejemplos:

- a) Planeación de las finanzas y de la economía.
- b) Planeación de la producción.
- c) Control de los inventarios.
- d) Control de la producción.

2.3 Objetivo del análisis de series de tiempo

- I. DESCRIPCIÓN, simplemente enunciar el comportamiento de un proceso, permitiendo una visualización clara.
- II. EXPLICACIÓN, permitiendo un análisis de tipo causa-efecto.
- III. PRONÓSTICO, en el caso que se desee estimar valores futuros con el fin de tomar decisiones..
- IV. CONTROL, si no solo se desea observar sino modificar el comportamiento del proceso para obtener un beneficio.

2.4 Enfoque de Box y Jenkins

El análisis de las series cronológicas según Box y Jenkins (1976) se empieza con la descripción de un proceso estocástico según la siguiente definición:

Un proceso estocástico es una familia de variables aleatorias asociadas a un conjunto índice de números reales, de tal forma que a cada elemento del conjunto le corresponda una y solo una variable aleatoria⁴.

Con base en esta definición se tiene que una serie de tiempo es un proceso estocástico cuyo conjunto índice se toma con la relación al tiempo. En este

⁴ Guerrero, Víctor M., Análisis estadístico de series de tiempo económicas, p.5.

contexto es importante hacer notar que dado una serie de tiempo es el resultado de una realización del proceso estocástico que la genera; esto refleja su carácter probabilístico. Se denotará por Y_t , tanto a la variable aleatoria observada en el tiempo t como al valor observado de la misma, el supuesto principal en materia de series de tiempo es que existe una dependencia entre las observaciones registradas, por lo que deben considerarse los estimadores de los momentos de primer y segundo orden:

- $E(Y_t) = \mu.$
- $VAR(Y_t) = E[(Y_t - \mu_t)^2]$ y
- $COV(Y_t, Y_{t+k}) = E[(Y_t - \mu_t)(Y_{t+k} - \mu_{t+k})].$

La estacionaridad se presenta cuando existen funciones del proceso estocástico que no varían con el tiempo, es decir que llega a un nivel de equilibrio cuyo valor no depende del tiempo. Existen dos tipos de estacionaridad: los momentos de primer y segundo orden que no dependen del tiempo, y la estricta, que implica que la función de densidad para un conjunto arbitrario de variables que están incluidas en el proceso no varía como consecuencia de desplazamientos en el tiempo.

Es decir, $f(Y_t, Y_{t+1}, Y_{t+2}, \dots, Y_{t+m}) = f(Y_{t+k}, Y_{t+k+1}, Y_{t+k+2}, \dots, Y_{t+k+m})$

Así pues como la más común es suponer que la distribución correspondiente a las series de tiempo es la normal, cuando en esta tesis haga referencia a la estacionaridad deberá entenderse que me refiero a la estacionaridad de segundo orden por lo que solo será necesario conocer la media, varianza y la función de autocovarianza de las variables del proceso.

El concepto de estacionaridad es de la mayor importancia ya que la teoría de Box y Jenkins parte del supuesto de que la serie a analizar es estacionaria. Es cierto que en la práctica la mayoría de las series temporales no son estacionarias, pero afortunadamente se han desarrollado métodos para volver estacionarias a las series que en un principio no lo son; esos métodos serán presentados más adelante en esta tesis.

Los operadores que a continuación se describen son un elemento importante en el modelo de Box y Jenkins:

Operador	Notación
De retraso	B
Diferencia	∇

Usando estos operadores se define los llamados polinomios de retraso que tienen la siguiente representación:

$$G(B) = 1 - g_1 B - g_2 B^2 - \dots - g_k B^k = 1 - \sum_{j=1}^k g_j B^j$$

Mediante el uso de estos polinomios es como se presentan los modelos de las series en el enfoque estocástico.

A los modelos de Box y Jenkins para series de tiempo se hacen referencia a las ecuaciones lineales en diferencia para procesos deterministas, que son el equivalente discreto de las ecuaciones diferenciales, es decir, miden la variación entre los valores de la variable respecto al tiempo transcurrido entre las observaciones. Considerando que en el presente caso el tiempo relacionado con tales observaciones es discreto y varía en una unidad en cada registro, entonces

las ecuaciones en diferencia se reducen a medir las variaciones entre los valores de la variable.

Una ecuación en diferencia de orden p se denota así:

$$Y_t - a_1 Y_{t-1} - a_2 Y_{t-2} - a_3 Y_{t-3} - \dots - a_p Y_{t-p} = a_0 \quad (2.4.1)$$

con $t = 0, 1, 2, \dots, p$; $a_i = \text{cte}$, $a_p \neq 0$

puede expresarse, utilizando el operador de retraso, como:

$$(1 - a_1 B - a_2 B^2 - a_3 B^3 - \dots - a_p B^p) Y_t = a_0 \quad (2.4.2)$$

Por el teorema Fundamental del Álgebra este polinomio de retraso puede expresarse:

$$(1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_p B) = G(B) \quad (2.4.3)$$

de este modo:

$$(1 - a_1 B - a_2 B^2 - a_3 B^3 - \dots - a_p B^p) = (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_p B) \quad (2.4.4)$$

las raíces de la ecuación característica $G(x) = 0$ relativa a este polinomio serán

$x = g_1^{-1}, g_2^{-1}, g_3^{-1}, \dots, g_p^{-1}$ y permitirán conocer la solución general de la ecuación

(2.4.1); dicha solución esta dada por

$$Y_t = \left[(1 - g_1 B)(1 - g_2 B) \dots (1 - g_p B) \right]^{\alpha_0} + s_1 g_1^t + s_2 g_2^t + \dots + s_p g_p^t \quad (2.4.5)$$

Cuyos valores $s_i, i = 1, 2, \dots, p$ son constantes que se pueden determinar con base a p condiciones iniciales conocidas que pueden ser los valores Y_1 a Y_p , el nivel medio del proceso o alguna cota superior o inferior. Para confirmar que la

ecuación (2.4.5) es la solución general de (2.4.1) solo es necesario multiplicar (2.4.5) por (2.4.3) y considerar la equivalencia expresada en (2.4.4), esto es, para $i \leq p$ tenemos:

$$\begin{aligned}
 s_i g_i' G(B) &= s_i g_i' (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_i B) \dots (1 - g_p B) \\
 &= s_i g_i' (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_{i-1} B)(1 - g_{i+1} B) \dots (1 - g_p B) \\
 &= ((s_i g_i') - s_i g_i B g_i') (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_{i-1} B)(1 - g_{i+1} B) \dots (1 - g_p B) \\
 &= (s_i g_i' - s_i g_i g_i') (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_{i-1} B)(1 - g_{i+1} B) \dots (1 - g_p B) \\
 &= (s_i g_i' - s_i g_i g_i') (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_{i-1} B)(1 - g_{i+1} B) \dots (1 - g_p B) \\
 &= 0
 \end{aligned}$$

Con lo que el producto de (2.4.5) por (2.4.3) se reduce a:

$$\begin{aligned}
 (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_p B) Y_i &= e_0 / ((1 - g_1 B)(1 - g_2 B) \dots (1 - g_p B)) * (1 - g_1 B)(1 - g_2 B)(1 - g_3 B) \dots (1 - g_p B) \\
 \Rightarrow (1 - e_1 B - e_2 B^2 - e_3 B^3 - \dots - e_p B^p) Y_i &= e_0
 \end{aligned}$$

En este punto se debe señalar que la convergencia del proceso expresado como (2.4.5) a un nivel de equilibrio dependerá del módulo de los valores $g_i, i=1,2,\dots,p$ el proceso será convergente si los p valores cumplen con $|g_i| < 1$ es decir que g_i^{-1} se encuentra fuera del círculo unitario en el plano complejo. Estas condiciones surgen del hecho de que al representar al proceso determinista como:

$$Y_i = (1 - g_1 B)^{-1} (1 - g_2 B)^{-1} \dots (1 - g_p B)^{-1} e_0$$

Este sólo tiene en la practica sí:

$$(1-g_i)^{-1} = \sum_{n=0}^{\infty} (g_i B)^n = \sum_{n=0}^{\infty} g_i^n = \frac{1}{1-g_i}; \quad \text{si } |g_i| < 1, i=1,2,\dots,p \quad (2.4.6)$$

Además esto permite que:

$$\lim_{n \rightarrow \infty} g_i^n = 0 \quad \text{si } |g_i| < 1, i=1,2,\dots,p \quad (2.4.7)$$

las condiciones dadas en (2.4.6) y (2.4.7) no se cumplirán si $|g_i| \geq 1$.

El último tema para el análisis de series de tiempo se refiere a la función de autocorrelación **FAC** y la función de autocorrelación parcial, **FACP**.

La importancia de estas dos funciones es que permiten describir algunas características importantes del proceso generador de la serie de tiempo por medio del cual es posible identificar el modelo correspondiente a ésta.

Función de autocorrelación: permite evaluar la correlación entre dos observaciones distintas de una misma variable en dos puntos temporales diferentes, es decir separados por k periodos:

$$\rho_k = \frac{\text{Cov}(Y_t, Y_{t+k})}{\sqrt{[\text{Var}(Y_t)] * \text{Var}(Y_{t+k})}} \quad (2.4.8)$$

Función de autocorrelación parcial: mide la correlación entre dos observaciones distintas de una misma variable en dos puntos temporales diferentes; mientras las correlaciones de los periodos intermedios son controladas o mantenidas constantes, es decir, la **FACP** es una autocorrelación condicionada en el que el

valor de esta función entre dos observaciones mediatizado por los valores de las correlaciones para las observaciones intermedias.

Como consecuencia, la **FACP** no puede obtenerse por medio del uso de una sola ecuación irán cambiando a medida que los retardos se incrementen, pues serán más observaciones intermedias implicadas.

El cálculo de esta función se realiza mediante la regla de Cramer.

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \quad (2.4.9)$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_2 & \dots & \rho_{k-3} & \rho_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}} \quad (2.4.9)$$

Una ventaja de las funciones mencionadas es que al estar referidas a las correlaciones y ser éstas adimensionales, no se ven influidas por cambios extremos en los valores de la serie. En la siguiente sección se presentan los diferentes modelos bajo la metodología Box y Jenkins.

2.5 Modelo autorregresivo

Estos modelos tienen una representación similar a la de las ecuaciones en diferencias mostradas en (2.4.1), pero se consideran una variable aleatoria a_t , en vez que una constante, es decir, el modelo se convierte en:

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p} = e_t \quad (2.5.1)$$

$t = 0, 1, 2, 3, \dots$; $\phi_1, \phi_p \neq 0$; $e_t =$ variable aleatoria

Se llaman modelos autorregresivos debido a que el valor actual de la variable se estima con base en los valores que ésta misma tomó en tiempos anteriores a t , valores que se ponderaran con los coeficientes autorregresivos $\phi_i, i = 1, 2, \dots, p$. La necesidad de introducir la variable aleatoria se debe a que una serie cronológica no es proceso determinista, si no resultado de un padrón influido por movimientos aleatorios.

Si se cumple que Y_t tenga media constante $E(Y_t) = \mu$ y se hace el cambio de variable $Y_t = Y_t - \mu$ entonces el modelo en (2.5.1) puede representarse tal y como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (2.5.2)$$

donde: $\phi < 1$ condición de estacionaridad

A la ecuación (2.5.2) se le llama modelo autorregresivo de orden p y se simboliza como **AR(p)** otra forma de representar a (2.5.1) es:

$$\phi(B)Y_t = e_t \quad (2.5.3)$$

Considerando que $\{e_t\}$ es un proceso de ruido blanco⁵ con $\mu=0$, σ_e^2 y distribución normal que mide las variaciones o errores aleatorios del proceso.

En el caso de este tipo de proceso, es imposible decir que se alcance un punto de equilibrio debido a que la parte aleatoria que involucra; sin embargo, puede hacerse referencia a su estacionaridad.

Recordando lo revisado anteriormente sobre las ecuaciones en diferencia, se tiene por analogía la estacionaridad del proceso dependerá de los módulos de la ecuación característica $\phi(x)=0$.

Retomando que $\phi(B)$ puede escribirse como:

$(1-g_1 B)(1-g_2 B)(1-g_3 B)\dots(1-g_p B)=\phi(B)$ se llega a que (2.5.2) será un proceso estacionario si $|g_i| < 1$ para $i=1,2,\dots,p$. Esto equivalente a pedir que los valores de $g_1^{-1}, g_2^{-1}, g_3^{-1}, \dots, g_p^{-1}$.

Se encuentren todos fuera del círculo unitario resultado que permite determinar la estacionaridad de la serie, que hace posible deducir que al proceso es estacionario si las siguientes p determinantes son todos positivos.

$$D_1 = \begin{vmatrix} -1 & \phi_p \\ \phi_p & -1 \end{vmatrix} \quad (2.5.4)$$

⁵Un proceso de ruido blanco es una sucesión de choques aleatorios de una variable aleatoria con media constante (generalmente se le considera igual a cero) y varianza σ_e^2 .

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

$$D_p = \begin{vmatrix} -1 & 0 & \dots & 0 & \phi_p & \phi_{p-1} & \dots & \phi_1 \\ \phi_1 & -1 & \dots & 0 & 0 & \phi_p & \dots & \phi_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \phi_{p-1} & \phi_{p-2} & \dots & -1 & 0 & 0 & \dots & \phi_p \\ \phi_p & 0 & \dots & 0 & -1 & \phi_1 & \dots & \phi_{p-1} \\ \phi_{p-1} & \phi_p & \dots & 0 & 0 & -1 & \dots & \phi_{p-2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \phi_1 & \phi_2 & \dots & \phi_p & 0 & 0 & \dots & -1 \end{vmatrix} \quad (2.5.4)$$

Si $\phi(B)Y_t = e_t$, resulta ser estacionario entonces se puede expresar como una suma infinita en términos de los errores aleatorios

$$Y_t = \phi(B)^{-1} e_t = \Psi(B)e_t = e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \Psi_3 e_{t-3} + \dots \text{con } \sum_{i=1}^{\infty} \Psi_i < \infty$$

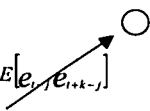
A esta propiedad se le conoce con el nombre de invertibilidad. Adicionalmente puedo mencionar que es mejor representar el modelo con número finito de parámetros que con una infinidad de los mismos; a esta medida de elegir el modelo con el menor número de parámetros se le llama principio de parsimonia.

Considerando que el proceso $\mathbf{AR}(p)$ es estacionario, la media no depende del tiempo, ya que $E(\mathbf{Y}_t) = 0$ y los valores de la varianza y autocovarianza, basados en la representación invertida de Z_t , serán:

$$\gamma_0 = \text{Var}(\mathbf{Y}_t) = E[(\mathbf{Y}_t - E(\mathbf{Y}_t))^2] = E(\mathbf{Y}_t^2)$$

$$\gamma_0 = E[(\mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} + \Psi_3 \mathbf{e}_{t-3} + \dots)^2]$$

$$\begin{aligned} \gamma_0 &= E[(\mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} + \Psi_3 \mathbf{e}_{t-3} + \dots) \mathbf{e}_t] \\ &+ E[(\mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} + \Psi_3 \mathbf{e}_{t-3} + \dots) \Psi_1 \mathbf{e}_{t-1}] \\ &+ E[(\mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} + \Psi_3 \mathbf{e}_{t-3} + \dots) \Psi_2 \mathbf{e}_{t-2}] \\ &+ E[(\mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} + \Psi_3 \mathbf{e}_{t-3} + \dots) \Psi_3 \mathbf{e}_{t-3}] + \dots \end{aligned}$$

$$\gamma_0 = E(\mathbf{e}_t^2) + \Psi_1^2 E(\mathbf{e}_{t-1}^2) + \Psi_2^2 E(\mathbf{e}_{t-2}^2) + \Psi_3^2 E(\mathbf{e}_{t-3}^2) + \dots + \sum_{j=1}^{\infty} \Psi_j \Psi_j E[\mathbf{e}_t \mathbf{e}_{t+k-j}]$$


$$\gamma_0 = \sigma_e^2 \sum_{j=1}^{\infty} \Psi_j^2 \quad \text{con} \quad \Psi_0 = 1 \quad (2.5.5)$$

Puede advertirse que tampoco $\gamma_0 = \text{Var}(\mathbf{Y}_t)$ depende del tiempo sino de la convergencia de la suma infinita y, finalmente:

$$\gamma_k = \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) = E[(\mathbf{Y}_t - E(\mathbf{Y}_t))(\mathbf{Y}_{t+k} - E(\mathbf{Y}_{t+k}))] = E(\mathbf{Y}_t \mathbf{Y}_{t+k})$$

$$\gamma_k = E \left[\begin{aligned} &(\mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} + \dots) \\ &(\mathbf{e}_{t+k} + \Psi_1 \mathbf{e}_{t+k-1} + \Psi_2 \mathbf{e}_{t+k-2} + \dots + \Psi_{k-1} \mathbf{e}_{t+k-(k-1)} + \Psi_k \mathbf{e}_{t+k-k} + \Psi_{k+1} \mathbf{e}_{t+k-(k+1)} + \dots) \end{aligned} \right]$$

$$\begin{aligned} \gamma_k &= E[(e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \Psi_3 e_{t-3} + \dots) e_{t-k}] \\ &+ E[(e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \Psi_3 e_{t-3} + \dots) \Psi_1 e_{t-k-1}] \\ &+ E[(e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \Psi_3 e_{t-3} + \dots) \Psi_2 e_{t-k-2}] + \dots \\ &+ E[(e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \Psi_3 e_{t-3} + \dots) \Psi_{k-1} e_{t-k-(k-1)}] \\ &+ E[(e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \Psi_3 e_{t-3} + \dots) \Psi_k e_{t-k}] \\ &+ E[(e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \Psi_3 e_{t-3} + \dots) \Psi_{k+1} e_{t-k-(k+1)}] + \dots \end{aligned}$$

$$\gamma_0 = E(\Psi_k e_t^2) + E(\Psi_1 \Psi_{k+1} e_{t-1}^2) + E(\Psi_2 \Psi_{k+2} e_{t-2}^2) + \dots + \sum_{j=k-1}^{\infty} \Psi_j \Psi_j E[e_{t-j} e_{t+k-j}]$$

$$\gamma_0 = \sigma_e^2 \left[\Psi_k + \sum_{j=1}^{\infty} \Psi_j \Psi_{k+j} \right] \quad \text{con} \quad \Psi_0 = 1 \quad (2.5.6)$$

Así, en (2.5.6), $\gamma_k = \text{Cov}(Y_t, Y_{t+k})$ tampoco depende del tiempo, sino solo de la separación de las variables y de la convergencia de la suma infinita. Las primeras p autocorrelaciones que forman la **FAC** pueden obtenerse en términos de los parámetros ϕ y las ecuaciones de Yule-Walker siendo estas:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_2 \rho_{p-2} \\ &\dots \dots \dots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \rho_p \end{aligned} \quad (2.5.7)$$

Las demás autocorrelaciones se obtienen de la relación

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}; \quad \text{para } k \geq p+1 \quad (2.5.8)$$

La **FAC** converge a cero exponencialmente dependiendo de las raíces de $\phi(B)=0$. Todo proceso **AR(p)** tiene asociada una única **FAC** que se

encuentran determinada por las condiciones (2.5.7) y (2.5.8), cuyas correlaciones obviamente no dependen del tiempo.

En cuanto a la **FACP**, se tiene que sólo las primeras p autocorrelaciones parciales son distintas de cero y que el modelo.

Depende de p autocorrelaciones parciales son distintas de cero debido a que el modelo depende de p retrasos en Y_t (la autocorrelación se ve definida sólo por p variables consecutivas). Para ver esto se tiene que para la autocorrelación parcial ϕ_{kk} para $k > p$, señalada en (2.4.9), la última columna en la matriz del numerador puede ser escrita como una combinación lineal de las columnas previas de la matriz por lo que para $K > p$ la **FACP** desaparece.

2.6 Modelo de promedios móviles

Estos modelos conciben a la serie de tiempo $\{Y_t\}$, formada por observaciones dependientes, como si se hubiera generado por una combinación lineal finita de choques aleatorios independientes $\{e_t\}$. Si se toman las desviaciones de $\{Y_t\}$ Si se toman las desviaciones de $\{Y_t\}$, respecto de su media μ es factible generar la serie $\{Y_t\}$, donde $Y_t = Y_t - \mu$, y formular el modelo de promedios móviles de orden q como a continuación se describe:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2.6.1)$$

donde: $\theta < 1$ condición de invertibilidad

Equivalentemente a:

$$Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t = \theta(B) e_t \quad (2.6.2)$$

Y a:

$$Y_t = \theta(B) e_t \quad (2.6.3)$$

Donde: $\theta_i, i=1,2,\dots,q$ son las ponderaciones asociadas a los choques en el periodo t-1,t-2,...,t-q. Por las representaciones anteriores se nota que al ser $\{Y_t\}$ una combinación lineal finita de variables aleatorias estacionarias -- *recuérdese que se supone que $\{e_t\}$ es un proceso de ruido blanco*-- entonces todo proceso de promedios móviles es estacionario.

El proceso en (2.6.1) se simboliza como **MA(q)** y se llama proceso de promedios móviles de orden **q**. Probablemente la denominación de promedios móviles proviene de la idea de promediar o ponderar los choques aleatorios, aunque en este caso no se trata necesariamente de un promedio pues no es indispensable que la suma de las ponderaciones sea uno, ni que estas sean positivas. Las restricciones para invertibilidad de este modelo se obtienen por dualidad de las condiciones de estacionaridad del modelo autorregresivo en (2.5.4), es decir, (2.6.3) será un proceso invertible si las **q** determinantes que se presentan a continuación son todos positivos:

$$D_1 = \begin{vmatrix} -1 & \theta_p \\ \theta_p & -1 \end{vmatrix} \quad D_2 = \begin{vmatrix} -1 & 0 & \theta_q & \theta_{q-1} \\ \theta_1 & -1 & 0 & \theta_q \\ \theta_q & 0 & -1 & \theta_1 \\ \theta_{q-1} & \theta_q & 0 & -1 \end{vmatrix} \quad (2.6.4)$$

$$D_q = \begin{vmatrix} -1 & 0 & \dots & 0 & \theta_q & \theta_{q-1} & \dots & \theta_1 \\ \theta_1 & -1 & \dots & 0 & 0 & \theta_q & \dots & \theta_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{q-1} & \theta_{q-2} & \dots & -1 & 0 & 0 & \dots & \theta_q \\ \theta_q & 0 & \dots & 0 & -1 & \theta_1 & \dots & \theta_{q-1} \\ \theta_{q-1} & \theta_q & & 0 & 0 & -1 & \dots & \theta_{q-2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_1 & \theta_2 & \dots & \theta_q & 0 & 0 & \dots & -1 \end{vmatrix} \quad (2.6.4)$$

2.7 Modelos autorregresivos y de promedios móviles

Los modelos denotados por **ARMA**(p, q) surgen al combinar el modelo **AR**(p) con el **MA**(q). El modelo **ARMA**(p, q) se presenta así:

$$\phi(B)Z_t = \theta(B)e_t \quad (2.7.1)$$

Equivalente a:

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p} = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2.7.2)$$

Con $\phi(B)$ y $\theta(B)$ polinomios de retraso de órdenes p y q respectivamente,

$\{e_t\}$ un proceso de ruido blanco.

TESIS CON
 FALLA DE ORIGEN

Y $\{Y_t\}$ es la serie de desviaciones $\{Y_t\}$ respecto de su media μ , Para que el proceso sea estacionario se requiere que las raíces de $\phi(x) = 0$ estén fuera del círculo unitario y para que sea invertible es necesario que las de $\theta(x) = 0$ lo estén también, por lo que deberán cumplirse que todos los determinantes mostrados en (2.5.4) y (2.6.4) sean positivos.

Como consecuencia de la estacionaridad y de invertibilidad que presentan los procesos **AR**(p) y **MA**(q) que se combinan, se tiene que la esperanza, varianza, covarianza y autocorrelación, del **ARMA**(p, q) no depende del tiempo.

Para derivar la **FAC** del proceso basta con despejar a $\{Y_t\}$ de (2.7.2), multiplicar el resultado por $\{Y_{t+k}\}$ y tomar la esperanza de cada uno de los términos con los que se obtiene:

$$\begin{aligned} \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \phi_3 \gamma_{k-3} + \dots + \phi_p \gamma_{k-p} & k \geq q+1 & \Rightarrow \\ \rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_3 \rho_{k-3} + \dots + \phi_p \rho_{k-p} & k \geq q+1 & \quad (2.7.3) \end{aligned}$$

Por lo que sigue la misma ecuación en diferencia que la **FAC** del proceso **AR**(p) expresada en (2.5.8), sólo que en (2.7.3) esta función inicia su convergencia a cero después del retraso $k=q$, en tanto que las primeras q autocorrelaciones dependen tanto de los parámetros autorregresivos como de promedios móviles, lo que implica que tengan un comportamiento irregular. Por su parte, la **FACP** será una sucesión infinita convergente a cero.

2.8 Modelos autorregresivos e integrados de promedios móviles

Este tipo de modelo es una extensión de los modelos **ARMA** en la que se considera una serie temporal que no es estacionaria en la media. La idea del modelo **ARIMA** es la siguiente: si la serie original $\{Y_t\}$ tiene una tendencia polinomial no determinística (u homogénea, como también se le llama) que provoca que la serie no sea estacionaria entonces es factible la construcción de segundo proceso $\{W_t\}$ que si sea estacionario.

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 estacionaridad mostrado por algunas serie de tiempo podría representarse mediante la simple toma sucesiva de diferencias de la serie original, esto permite gran flexibilidad de representación a los modelos **ARMA**, puesto los que en realidad se hace es aplicar el operador diferencia ∇^d

Es eliminar una posible tendencia polinomial de orden d , presente en la serie que se analice. Visto desde otro ángulo, si el proceso original $\{Y_t\}$ adolece de no estacionaridad causada por una tendencia polinomial no-determinista (a la cual se le denomina no estacionaridad homogénea) es posible construir el proceso estacionario $\{W_t\}$, donde:

$$W_t = \nabla^d Y_t; \text{ para toda } t \quad (2.8.1)$$

para esta nueva serie podría ser ya posible obtener un modelo **ARMA**

$\phi(B)W_t = \theta(B)a_t$, la cual serie equivalente a considerar el modelo **ARIMA**

$$\phi(B)\nabla^d Y_t = \theta(B)a_t, d \geq 1 \quad (2.8.2)$$

Para $\{Y_t\}$ en donde $\{a_t\}$ es un proceso de ruido blanco. El termino integrado se refiere a que Z_t , se obtiene de la relación de ecuación (2.8.1) por inversión del operador ∇^d , dando como resultado una suma infinita de términos W_t , y así, si $W_t = \nabla Y_t$; se tendría:

$$Y_t = \nabla^{-1} W_t = W_t + W_{t-1} + W_{t-2} + \dots \quad (2.8.3)$$

El orden del polinomio de retraso $\phi(B)$ del exponente en el operador diferencia y el orden del polinomio de retraso $\theta(B)$, se acostumbra mencionarlos en secuencias en esa secuencia, de manera que el 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 de la ecuación (2.8.2) puede escribirse como:

$$W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} - \dots - \phi_p W_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \dots \quad (2.8.4)$$

donde:

$$W_t = \nabla^d Y_t$$

Ya que el interés primordial se ubica en modelos que sean 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:

$$\varphi(B) = \phi(B)\nabla^d \quad (2.8.5)$$

la condición es que de las raíces de $\varphi(x) = 0$ sean unitarias, mientras que las restantes deben estar fuera del círculo unitario. Haciendo uso de la ecuación (2.8.5) la expresión (2.8.4) se convierte en:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_{p,d} Y_{t-p,d} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \dots \quad (2.8.6)$$

Y_t , la representación en términos de suma ponderada de choques aleatorios viene dada por:

$$Y_t = \varphi^{-1}(B)\theta(B)a_t = \psi(B)a_t \quad (2.8.7)$$

2.9 Series de tiempo estacionales

Un tipo especial de no estacionaridad presente en muchas series, el la estacionalidad que, es un fenómeno que consiste en la presencia en la serie de cualquier comportamiento periódico con duración menor o igual a un año que se registre regularmente a lo largo del tiempo.

No es posible representar este tipo de series como un simple modelo **ARIMA**. Debido a esa característica estacional y es por eso que a continuación se presenta la metodología aplicable en estos casos.

Por una serie estacional se entenderá una serie de tiempo que aparte de contener una tendencia (y /o ciclos) de larga duración, muestre fluctuaciones que se repiten anualmente, quizá con cambios graduales a través de los años:

En las series de tiempo observadas a intervalos mensuales o trimestrales, las cimas los valles tienen tendencia a ocurrir alrededor del mismo trimestres o meses sucesivos de años sucesivos, en el caso de algunos fenómenos como el turismo, enfermedades, ventas, producción, fallos, etc. Parece lógico entonces, tratar de explotar la correlación entre los mismos meses o trimestres de meses sucesivos.

2.10 Operadores y modelos estacionales

Para representar las series estacionales se introduce el operador de diferencia estacional ∇_i^D definido como:

$$\nabla_i^D = (1 - B^S)^D Y_i$$

y el *polinomio de retraso estacional* de orden **D** con coeficientes constantes g_1, g_2, \dots, g_k que está dado por:

$$G(B^S) = 1 - g_1 B^S - g_2 B^{2S} - \dots - g_k B^{kS} = \sum_{r=0}^k g_r B^{rS}$$

Al aplicarle el operador ∇_i^D se pierde **DxS** observaciones.

Utilizando estas notaciones se puede escribir un modelo **ARIMA(P,D,Q)**, puramente estacional como:

$$\Phi(B^S) \nabla_i^D Y_i = \Theta(B^S) e_i$$

Donde $\Phi(B^s)$ es el polinomio autorregresivo estacional de orden P , $\Theta(B^s)$ es un polinomio de promedios móviles estacional de orden Q y e_t es ruido blanco. El comportamiento de la **FAC** de un proceso **ARIMA**(P,D,Q)_s es análogo al de la **FAC** de un modelo **ARIMA**(p,d,q) no estacional por lo que la metodología de construcción revisa anteriormente se aplica sin mayor modificación a los procesos puramente estacionales.

2.11 Modelo Multiplicativo Estacional

Muchas series temporales con estructuras **ARIMA** estacional también tiene estructura **ARIMA** regular, estas estructuras pueden incorporarse aditivamente, pero para el modelo resulte más poderoso, las estructuras deben combinarse multiplicativamente. Box y Jenkins propusieron el modelo general de la forma:

$$\Phi(B^s)\nabla^d Y_t = \Theta(B^s)a_t$$

Donde: a_t no es ruido blanco, si no un proceso que sigue un modelo **ARIMA**(p,d,q)

$$\phi(B)\nabla^d a_t = \theta(B)e_t$$

Con estas dos expresiones se obtiene el modelo **ARIMA** multiplicativo estacional representado por **ARIMA**(p,d,q)X(P,D,Q)_s

$$\phi(B)\Phi(B^s)\nabla^d \nabla^D Y_t = \theta(B)\Theta(B^s)e_t$$

Donde los términos P , D y Q tienen una interpretación similar a la p , d y q , s indica la longitud del periodo estacional. Como puede notarse, a través de este modelo se consideran relaciones entre observaciones contiguas dentro

del mismo año y entre observaciones presentadas en los mismos periodos de diferentes años.

Finalmente, cabe destacar que el modelo $ARIMA(p,d,q) \times (P,D,Q)_s$ deberá tener los parámetros restringidos a las bandas de estacionaridad e invertibilidad para los procesos **ARIMA** previamente diferenciados d veces de forma ordinaria y D veces de forma estacional, es decir, se deberá cumplir que las raíces de $\phi(x)=0$, $\Phi(x^s)=0$, $\theta(x)=0$ y $\Theta(x^s)=0$ se encuentren fuera del círculo unitario. Obviamente, debido a la mayor complejidad del modelo, la estructura de las autocorrelaciones es más complicada.

2.12 Identificación del modelo estacionario

Podemos dividir la identificación de modelos en 2 fases:

El primer paso, en la identificación de modelos será la gráfica de la serie de tiempo para determinar si es necesario aplicar alguna transformación en la varianza o en la tendencia de los datos o la existencia de un comportamiento estacional. En ocasiones no será fácil determinar a simple vista en la gráfica de la serie temporal, si se necesitan hacer transformaciones, entonces nos apoyamos en las gráficas de **FAC** y **FACP**.

El segundo paso, una vez obtenida una serie de tiempo estacionario, las gráficas de la **FAC** y **FACP** se verán de manera diferente, de acuerdo al tipo de modelo que se trate.

- Si la gráfica de la **FAC** disminuye rápidamente y la grafica de la **FACP** se corta entonces es un **AR(p)**.

- Para un modelo **MA(q)** la gráfica de la **FAC** es la que indicara el orden del modelo de acuerdo al número donde se corta y la **FACP** disminuirá de una forma acelerada.
- Para un **ARMA(p,q)** la gráfica de la **FAC** se corta después de $q-p$ intervalos, similarmente, la grafica de la **FACP** se corta después de $p-q$ intervalos.

Lo anterior expuesto se resume de esta manera:

Modelo	FAC	FACP
AR(p)	Decrece	Se trunca después de p
MA(q)	Se trunca después de q	Decrece
ARMA(p,q)	Decrece	Decrece

2.13 Identificación del modelo estacional

Las series temporales con estructura **ARIMA** estacional también puede ser detectadas a través de la **FAC** y de la **FACP** correspondientes.

La identificación se vuelve ahora más difícil debido a que se incrementan la cantidad de modelos que se pueden utilizar. Para una estacionalidad de longitud s estas funciones presentan elevamientos y decaimientos significativos en los retardos $s, 2s, 3s, \dots$, pero por lo demás tienen un mismo comportamiento que un proceso no estacional

**TESIS CON
FALLA DE ORIGEN**

Cabe señalar que en la forma explícita del modelo se utiliza a $W_t = \nabla^d \nabla^D T(Y_t)$ y que los valores d y D se determinan de tal forma que W_t sea aproximadamente estacionaria. Para tomar esta decisión puede ser útil la gráfica la **FAC** muestral de varias diferencias por ejemplo, $\{T(Y)\}$, $\{\nabla T(Y)\}$, $\{\nabla^2 T(Y)\}$, $\{\nabla \nabla T(Y)\}$, $\{\nabla^2 \nabla T(Y)\}$ y por inspección visual el grado de diferenciación adecuado.

Para facilitar la especificación de los modelos estacionales, autores como Anau, McCleary, Hamilton y Watts recomiendan que se tengan en cuenta las consideraciones que se siguen:

- a) La ausencia de estacionaridad estacional se detecta inspeccionando la **FAC**. Si la serie es estacional, la **FAC** presentará valores grandes y casi iguales en los retardos estacionales, es decir **FAC(s) ≈ FAC(2s) ≈ FAC(3s).....**
- b) Se puede esperar que los procesos **ARIMA(P,0,0)_s** presentan un patrón de decaimiento en la **FAC** donde la tasa de decaimiento estará en función de los valores $\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_m$ y, por el contrario, que su **FACP** presente **P** picos en los primeros **P** retardos, mientras que los restantes valgan cero.
- c) Es posible que los procesos **ARIMA(0,0,Q)_s** presentan **Q** picos en los términos **Q** retardos estacionales de su **FAC**, mientras que los restantes valgan cero y, por el contrario, un patrón de decaimiento progresivo en la **FACP** cuya tasa de decaimiento estará en función de los valores $\Theta_1, \Theta_2, \Theta_3, \dots, \Theta_n$.

d) Generalmente, las componentes regular y estacional del modelo **ARIMA**(p,d,q) $X(P,D,Q)_s$ serán de la misma clase, es decir, raramente ocurrirá que la parte regular quede adecuadamente descrita por un proceso **AR** y la estacional por un proceso **MA** o a la inversa.

La identificación de un modelo **ARIMA**(p,d,q) $X(P,D,Q)_s$ implica los siguientes pasos: Determinar mediante la **FAC** si la serie es o no estacionaria (sino es estacionaria conviene diferenciarla); Después se utilizan la **FAC** y la **FACP** para identificar el modelo correspondiente a la parte regular, posteriormente se inspecciona la **FAC** y la **FACP** para determinar los valores **P** y **Q** de la parte estacional respectivamente.

Para los modelos multiplicativos su identificación como se menciono anteriormente es más complicada pues en muchas ocasiones las gráficas de **FAC** y **FACP** no serán lo suficientemente claras para interpretar el comportamiento del proceso, en este caso es necesario probar con varios modelos para comprobar el ajuste al proceso de cada modelo; a continuación se muestra el cuadro de comportamiento teórico de las **FAC** y las **FACP**.

Modelo	FAC	FACP
AR(p)*SAR(P)	Decrece	Se corta después de $p+sp$ intervalos
MA(q)*SMA(Q)	Se corta después de $q+sQ$ intervalos	Decrece
Mezclados	Decrece	Decrece

**TESIS CON
FALLA DE ORIGEN**

En la identificación de modelos ARIMA deben procurarse siempre los modelos parsimoniosos, es decir, que el mejor modelo es aquel que siendo el más sencillo se ajusta correctamente a los datos de la serie

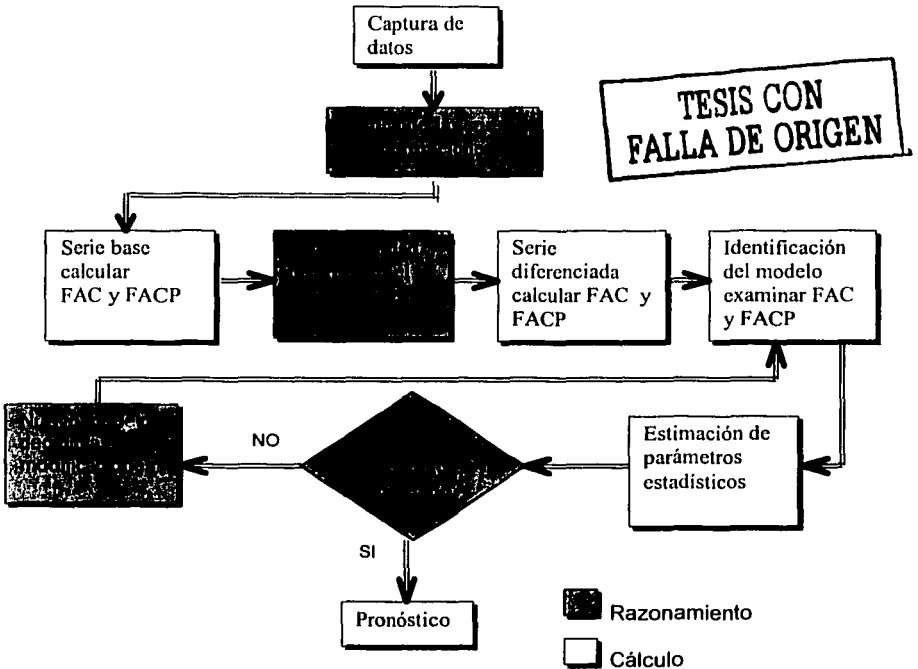
Capítulo III

Selección y diagnóstico del modelo para la proyección.

3.1 Metodología Box y Jenkins

La serie de tiempo se descompone en varios componentes, algunas veces llamados "filtros", precisamente porque la filosofía del método consiste en detectar las distintas componentes usando los filtros correspondientes, hasta obtener residuales no predecibles cuyo comportamiento tiene poca influencia en el resultado final.

Diagrama funcional del método Box y Jenkins



3.2 Identificación del modelo

En esta etapa se determina la transformación de la serie temporal necesaria para estabilizar la varianza y el número d de veces que deberá diferenciarse la serie transformada para obtener una serie sin tendencia. Posteriormente con base en los comportamientos de la **FAC** y **FACP**, se establecen los órdenes de los polinomios autorregresivos y de promedios móviles (p,q) para el proceso **ARMA** relacionada con la serie estacionaria.

3.3 Estabilización de la varianza

Para empezar a identificar el modelo lo primero que se debe hacer es determinar si es posible considerar que el proceso tiene varianza constante. Para lograr este objetivo, se forman grupos homogéneos de observaciones contiguas, es decir, si se tiene N observaciones se dividen éstas en H grupos con R observaciones ($R=[N/H]$) y si sobran $n=N-R*H$ observaciones éstas deberán eliminarse del principio o del final de la serie para realizar estas pruebas *--en la práctica se recomienda formar grupos de entre 8 y 15 observaciones--*

A continuación se determinan las medias y desviaciones estándares de cada grupo H : Sea $Y_{h,r}$ la observación r -ésima del grupo h -ésima

$$Y_h = \sum_{r=1}^R Y_{h,r} / R \quad h=1,2,\dots,\dots,\dots, H \quad (3.3.1)$$

$$S_h^2 = \sum_{r=1}^R (Y_{h,r} - Y_h)^2 / (R-1) \quad h=1,2,\dots,\dots,\dots, H \quad (3.3.2)$$

Si resulta que la varianza por grupos no es "estable" se determina el valor de λ (parámetro de transformación) para el cual la serie transformada cumpla con

esa condición. El valor de λ seleccionado será el que minimice el coeficiente de variación de λ ($cv(\lambda)$) con:

$$CV(\lambda) = S(\lambda)/M(\lambda) \quad (3.3.3)$$

$$M(\lambda) = \sum_{h=1}^n (S_h / Z_h^{\lambda})^H \quad (3.3.4)$$

$$S^2(\lambda) = \sum_{h=1}^n [S_h / Y_h^{1-\lambda} - M(\lambda)]^2 (H-1) \quad (3.3.5)$$

Para estacionarizar la varianza se utiliza la propuesta por Box-Cox (1964) que sugiere una transformación del tipo:

$$T(Y_t) = \begin{cases} (Y_t^\lambda - 1)/\lambda & \text{si } \lambda \neq 0 \\ \log(Y_t) & \text{si } \lambda = 0 \end{cases} \quad (3.3.6)$$

3.4 Estabilización de la tendencia de la serie

Para tendencia de la media habrá que diferenciar la serie un número d de veces adecuando (generalmente de 1 a 3 veces) aplicando el operador ∇ , la nueva serie obtenida será estacionaria y con d observaciones menos que la original. Esto supone que existe un polinomio de grado d que se ajusta aproximadamente al comportamiento de la serie. Para decir el número d de veces que se debe aplicar el operador diferencia es aconsejable seguir las técnicas que a continuación se describen.

Una de las formas para determinar d es a través de las graficas de la **FAC** muestral de algunas diferencias de la serie original que va a ser analizada

$\{T(Y_t)\}, \{\nabla T(Y_t)\}, \{\nabla^2 T(Y_t)\}, \dots, \{\nabla^n T(Y_t)\}$ ya que el decaimiento rápido de la **FAC** hacia cero indica que un proceso estacionario. En la práctica no es frecuente el uso de más de tres diferencias sucesivas y no es recomendable

la sobre diferenciación puesto que amplía la varianza y se pierden datos de la serie.

3.5 Análisis de la función de autocorrelación y de autocorrelación parcial

Obviamente, no se conoce los valores reales de la **FAC** ni de la **FACP** por lo que si se cuenta con **N** observaciones, se procede a analizar los valores de la **FAC** muestral definida por:

$$r_k = \frac{\sum_{t=1}^{N-k} (Y_t - M(Y_t))(Y_{t+k} - M(Y_{t+k}))}{\sum_{t=1}^N (Y_t - M(Y_t))} \quad (3.5.1)$$

Donde: **M(Y_t)** es la media de las observaciones y **M(Y_t) = M(Y_{t+k})**. Asimismo, Se revisan los valores de la **FACP** muestral que se obtiene al sustituir en (2.4.9) a ρ_i por su estimación.

En este punto se debe asociar la **FAC** a un proceso **ARIMA**. Las aproximaciones básicas para la varianzas y autocovarianzas de las autocorrelaciones proporcionadas por Bartlett son:

$$Var(r_k) \approx \sum_{j=-\infty}^{\infty} (\rho_j^2 + \rho_{j+k} \rho_{j-k} - 4 \rho_k \rho_j \rho_{j-k} + 2 \rho_k^2 \rho_j^2) / N$$

a) Si el proceso es generado por ruido blanco con distribución normal, la aproximación para la varianza será:

$$Var(r_k) \approx 1/N$$

Para decidir si realmente todas las correlaciones son cero se realiza un comparativo entre los valores de r_k y sus desviaciones estándar; así r_k será considerada diferente de cero si

$$|r_k| \geq 2[1/N]^{1/2} \quad \forall k$$

b) Si el proceso es originado por un **MA(q)** las autocorrelaciones para atrasos mayores a **q** son cero y entonces la aproximación sería:

$$\text{Var}(r_k) \approx \left(1 + 2 \sum_{j=1}^q \rho_j^2\right) / N \quad \text{para } k > q$$

para decidir si las correlaciones son cero a partir de cierto retraso **q** se realiza un comparativo entre los valores r_k y sus desviaciones estándar, bajo este criterio r_k es diferente de cero si:

$$|r_k| > 2 \left[\left(1 + 2 \sum_{j=1}^q \rho_j^2\right) / N \right]^{1/2} \quad \text{para } k > q$$

Esta medida se basa en el hecho de que para una muestra de más de cien observaciones la distribución de r_k se aproxima a una distribución normal con media cero. Como no se cuenta los valores de ρ_i utilizo su estimador r_i

Por otro lado, el orden de un proceso **AR(p)** no es fácilmente identificable a partir de la **FAC** por lo que se hace uso de la **FACP** estimada; como esta función está influida por errores muestrales, para decidir si el valor ϕ_{ii} es cero se realiza una prueba a su estimador $\hat{\phi}_{ii}$ para cuantificar la variación muestral de éste. Según Quenouille (1949) si el proceso es **AR(p)** las autocorrelaciones parciales muestrales se distribuyen de manera independiente con media $E(\hat{\phi}_{ii}) = \phi_{ii}$ y con

$$\text{Var}(\hat{\phi}_{ii}) \approx 1/N \quad \text{para } i > p$$

Con base en esto se determina que ϕ_{ii} es distinto de cero si ϕ_{ii} se encuentra fuera del intervalo definido por:

$$\pm 1.96[\text{Var}(\phi_{ii})]^{1/2} \approx \pm 2[\text{Var}(\phi_{ii})]^{1/2} \approx \pm 2/(N)$$

Entonces, si se cumple la condición anterior para $i < p$ se tendrá asociado un proceso **AR(p)**

Finalmente, para identificar un modelo **ARMA(p,d)** se cuenta con diferentes referencias. La primera técnica de identificación es revisar que el proceso tenga asociada una **FACP** que no desaparecerá después de un número finito de retrasos, en tanto que la **FAC** tendrá un comportamiento irregular de las primeras q autocorrelaciones y después convergerá a cero.

La segunda técnica consiste en revisar que si $q > p$, la **FAC** tendrá un comportamiento irregular de las primeras $q-p$ autocorrelaciones y después convergerá a cero.

En tanto que la **FACP** no desaparecerá después de un número finito de retrasos; o bien, si $q < p$, la **FACP** tendrá un comportamiento irregular de las primeras $p-q$ autocorrelaciones y después convergerá a cero, en tanto que la **FAC** no desaparecerá después de un número finito de retrasos.

En resumen, si el proceso presenta su **FAC** y su **FACP** iguales a cero para todos sus retrasos entonces se tratará de un ruido blanco, en tanto que un proceso **AR(p)** tiene solo las primeras p autocorrelaciones parciales distintas de cero mientras que su **FAC** converge a cero y, por otra parte, **MA(q)** tendrá todas sus primeras p autocorrelaciones distintas de cero aunque su **FACP**

muestre convergencia a cero y, finalmente, un proceso **ARMA(p,q)** tendrá comportamientos irregulares en las dos funciones, **FACP** y **FAC**.

3.6 Estimación por momentos: Valores parametrales iniciales

Cuando el modelo identificado es **AR(p)** el valor inicial de los p parámetros puede obtenerse a través de las ecuaciones de Yule-Walker estimadas

$$r_1 = \phi_1 + \phi_2 r_1 + \dots + \phi_p r_{p-1}$$

$$r_2 = \phi_1 r_1 + \phi_2 + \dots + \phi_p r_{p-2}$$

.....

$$r_p = \phi_1 r_{p-1} + \phi_2 r_{p-2} + \dots + \phi_p$$

es decir

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{pmatrix} = \begin{pmatrix} 1 & r_1 & r_2 & \dots & r_{p-2} & r_{p-1} \\ r_1 & 1 & r_1 & \dots & r_{p-3} & r_{p-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & r_{p-3} & \dots & r_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ r_2 \\ \dots \\ r_p \end{pmatrix}$$

Cuando el modelo sea **MA(q)**, las relaciones entre los parámetros θ y las correlaciones ρ_1, ρ_2, \dots no son lineales por lo que se requiere resolver el sistema de ecuaciones no lineal.

TESIS CON
 FALLA DE ORIGEN

$$r_1 = (-\theta_1 + \theta_1 \theta_2 + \dots + \theta_{q-1} \theta_q) / (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$$

$$r_2 = (-\theta_1 + \theta_1 \theta_2 + \dots + \theta_{q-2} \theta_q) / (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$$

.....

$$r_q = -\theta_q / (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$$

Este sistema tiene múltiples soluciones pero sólo una de ellas satisfará las condiciones de invertibilidad por lo que se deberá ser elegida.

Para modelos **ARMA**(p, q) se presentan ecuaciones lineales y no lineales que involucran a los parámetros: Para resolver estas ecuaciones se considera el hecho de que $\phi(B)\rho_k = 0$ para $k > p$ por lo que el sistema:

$$r_{q+1} = \phi_1 r_q + \phi_2 r_{q-1} + \dots + \phi_p r_{q-p+1}$$

$$r_{q+2} = \phi_1 r_{q+1} + \phi_2 r_q + \dots + \phi_p r_{q-p+2}$$

.....

$$r_{q+p} = \phi_1 r_{q+p-1} + \phi_2 r_{q-1-2} + \dots + \phi_p r_p$$

Proporciona las estimaciones iniciales para $\phi_1, \phi_2, \dots, \phi_p$ que debe sustituirse en las primeras q autocorrelaciones muestrales para poder determinar los valores iniciales $\theta_1, \theta_2, \theta_3, \dots, \theta_q$:

En lo tocante al término θ_0 la estimación de este parámetro por el método de momentos se obtiene a partir de la relación $E[\phi(B)W_t] = \theta_0$ de forma que

$$\theta_0 = (1 - \phi_1 - \phi_2 - \dots - \phi_p) \underline{W}$$

con $\underline{W} = \sum_{i=d-p+1}^N W_i / (N - d - p + 1)$. El hecho de que se presente una tendencia determinística implica que ésta actuará aun fuera del periodo observado;

como este supuesto es muy fuerte generalmente se elimina el término determinística y sólo en la etapa de verificación se precisa si dicho término debe incluirse.

3.7 Verificación: ¿ Es adecuado el modelo seleccionado?

Una vez que el modelo ha sido identificado y estimado cabe preguntarse si es realmente adecuado. Si es posible encontrar evidencia de que el modelo es seriamente inapropiado.

En la sección siguiente se presenta los métodos para diagnosticar qué supuestos cumple el modelo construido. Un método para confirmar si el modelo es adecuado es la sobreparametrización, es decir, calcular los parámetros de un modelo más general que aquél que se había seleccionado lo que nos permitirá saber si el modelo es inadecuado. Adicionalmente es necesario analizar los residuales, en este tipo de análisis los propios datos sugerirán las modificaciones adecuadas.

3.8 Análisis de residuales

Una vez determinadas las estimaciones paramétrales θ y ϕ del modelo

$$\phi(B)W_t = \theta(B)e_t$$

Los residuales se estiman a través de:

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

considerando $\phi(B)\theta(B)^{-1} = \pi(B)$ el modelo puede escribirse como

$$W_t = e_t + \pi_1 W_{t-1} + \pi_2 W_{t-2} + \dots$$

Por otro lado, la estimación de W_t con base en el polinomio estimado $\pi(B)$ y las observaciones hasta el momento $t-1$ es:

$$\dot{W}_t = \pi_1 W_{t-1} + \pi_2 W_{t-2} + \dots$$

A partir de estas relaciones se sigue que:

$$e_t = W_t - \dot{W}_t$$

Es decir los residuales miden la discrepancia entre los valores observados y los valores estimados por el modelo.

Si un modelo se acerca bastante al proceso **ARIMA** que rige la serie, los errores se comportaran como ruido blanco: tendrán media cero, varianza constante y estarán incorrelacionados.

Una de las formas más claras y simples para detectar violaciones a los supuestos de los modelos es a través del análisis de residuales, en donde, como residual se considera aquella parte de las observaciones que no explica el modelo.

Además cuando el tamaño de muestra es grande, los errores aleatorios y los residuales (que también son variables aleatorias) son esencialmente iguales por esta razón al analizar los residuales observados $\{e_t\}$ se analiza básicamente lo que debería ser una realización del proceso de ruido blanco $\{e_t\}$. Los supuestos acerca del proceso $\{e_t\}$ puede verificarse y posiblemente corregirse, de la siguiente manera:

Supuesto 1. $\{e_t\}$ Tiene media cero:

$$m(e_t) = \sum_{i=1}^N \frac{e_t}{(N-d-p)}$$

$$\sigma_e = \left(\sum_{i=d+p+1}^N (e_t - M(e_t))^2 / (N-d-p-q) \right)^{1/2}$$

Posteriormente se evalúa

$$(N-d-p)^{1/2} M(e_t) / \sigma_e$$

Si el valor absoluto de este cociente es menor que dos en valor absoluto se dice que no hay evidencia de que la media del proceso sea distinta de cero y por lo tanto, no se rechaza el supuesto; sin embargo, si el valor absoluto es mayor o igual a dos, se dice que la media es significativamente distinta de cero (esto querrá decir que todavía existe una tendencia determinista o semideterminística en la serie) y se ha violado el supuesto.

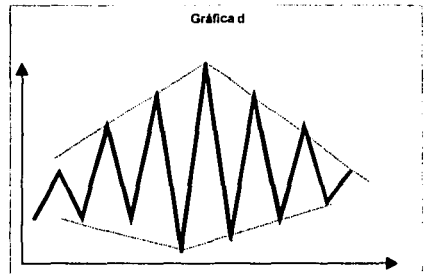
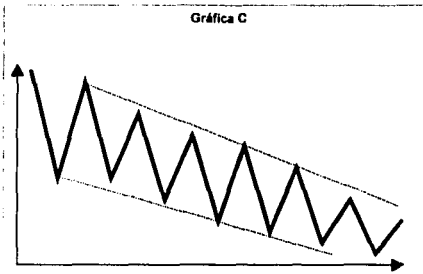
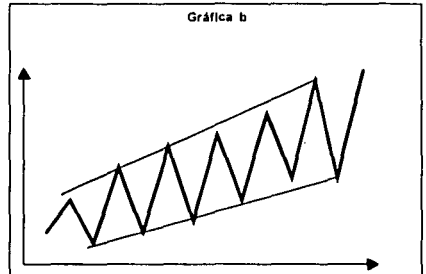
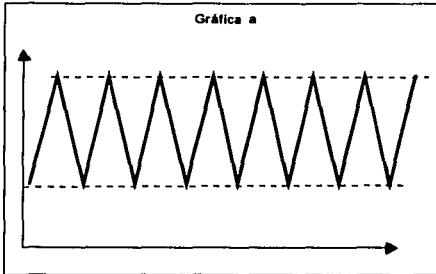
Supuesto 2. $\{e_t\}$ Tienen la varianza constante.

Verificación: Construir una gráfica de los residuales contra el tiempo para observar visualmente, si la varianza parece o no ser constante, pero la idea es que solamente las violaciones muy notorias a este supuesto son las que realmente llegan a causar problemas.

Figura 1

Algún comportamiento típico de la varianza residual:

- a) Varianza constante
- B) Varianza creciente
- c) Varianza decreciente
- d) Varianza no-monótona



Supuesto 3. Las variables aleatorias $\{e_t\}$ son mutuamente independientes.

Para verificar si esto es cierto implica analizar las autocorrelaciones residuales

TESIS CON
FALLA DE ORIGEN

mediante la estimación de la **FAC** muestral de los residuales que, bajo el supuesto de media cero, estarán formada por términos

$$r_k(\mathbf{e}) = \frac{\sum_{t=1}^{N-k} e_t e_{t+k}}{\sum_{t=1}^N e_t^2}, \quad k=1,2,\dots$$

con $l'=d+p+1$

1) Estimarse la desviación estándar de $r_k(\mathbf{e})$ como

$$\sqrt{\text{Var}[r_k(\hat{\mathbf{e}})]} = 1/\sqrt{N-d-p}$$

Para determinar la significación estadística individual de las autocorrelaciones de los residuales se realiza. A continuación considerando que:

$r_k(\mathbf{e}) \rightarrow N(0, [N-d-p]^{-1})$: Si $|r_k(\hat{\mathbf{e}})| \geq 2/\sqrt{N-d-p}$. Se considera que $r_k(\mathbf{e})$ es significativamente distinta de cero.

Durbin demostró que aunque los valores obtenidos con los datos estimados pueden ser una valiosa ayuda para reconocer lo adecuado del modelo, puede ser peligroso utilizar estos valores ya que es posible que se presenten alejamientos importantes respecto de los valores reales. Por su parte Box y Pierce notaron que esta prueba no es completamente válida para las autocorrelaciones correspondientes a retrasos pequeños ($k \leq 3$) pero si lo es para retrasos moderados y grandes: Por lo anterior, en estos casos se procede:

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

$$Q(k) = n(n+2) \sum_{i=1}^k r_i^2(\hat{\epsilon}) / n - k$$

Si el modelo es adecuado, es decir si los errores son ruido blanco, Q se distribuye aproximadamente como una ji cuadrada con $k-p-q$ grados de libertad. Se probara la hipótesis:

$$H_0 = \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k = 0$$

$$H_a = \text{al menos una distinta de cero}$$

Que los errores sean ruido blanco se rechaza cuando los valores de Q son grandes en relación con el valor de tablas de ji cuadrada.

Supuesto 4. $\{e_t\}$ tiene una distribución normal, para toda t

Verificación: 1) Se sabe que una distribución normal aproximadamente del 95% de las observaciones deben localizarse dentro un intervalo de dos desviaciones estándar por abajo y por arriba de la media entonces se cumple que la media de los residuales sea cero, se esperaría que a lo más un total de $\binom{N-d-p}{20}$ observaciones se localizarán fuera del intervalo

$$(-2\hat{\sigma}_e, 2\hat{\sigma}_e).$$

2) podría utilizarse también el un histograma de los residuales mediante el cual se pudiera visualizar la forma de distribución, con esto se podría detectar en particular si la distribución contiene asimetría.

La teoría se basa en el supuesto de que se distribuya de una forma normal, pero es obvio que no necesariamente sigue una distribución normal por lo que pueden esperarse ligeras desviaciones de este supuesto que no causen mayor problema; sin embargo, si el incumplimiento es muy notorio puede aplicarse una transformación normalizante.

Supuesto 5. Implícitamente se ha supuesto que no existan observaciones aberrantes (posiblemente ajenas a la serie de estudio).

Verificación: De nuevo la grafica de los residuales contra el tiempo permitirá visualizar si existe este tipo de observaciones anómalas. De esta manera, como una regla empírica de trabajo, podría considerarse como "sospechosa" a las observaciones cuyos residuales estén fuera del intervalo $(-3\sigma, 3\sigma)$, implicaría que un evento ocurra con una probabilidad del 0.2% aproximadamente (lo que sería muy extraño), o el residual registrado corresponde a una observación que no fue producida por el mismo proceso de la serie en estudio.

Supuesto 6. El modelo considerado es *parsimonioso*

Verificación: En si lo que la parsimonia implica es que no se puede reducir el numero de parámetros involucrados, ya que todos son necesarios para explicar el comportamiento del fenómeno y no pueden ser considerados como

iguales a cero. Para determinar si todos los parámetros incluidos en el modelo son necesarios y pueden ser considerados como iguales a cero. Lo que se hace en este caso es construir intervalos de aproximadamente 95% de confianza del tipo

$$\left(\beta - 2\sqrt{\text{Var}(\hat{\beta})}, \beta + 2\sqrt{\text{Var}(\hat{\beta})} \right) \quad \text{donde } \beta \in \beta = (\phi, \theta) \quad (3.8.1)$$

Para cada uno de los parámetros y observar si el valor cero se encuentra dentro del intervalo, y probar la hipótesis:

$$H_0: \beta_j = 0 \quad \beta_j = (\theta_j, \phi_j)$$

Esto es que el coeficiente del estimador del parámetro es insignificante y que podemos omitirlo del modelo, contra la hipótesis alternativa:

$$H_1: \beta_j \neq 0 \quad \beta_j = (\theta_j, \phi_j)$$

la cual indica que el parámetro es necesario para el modelo.

Si el valor es cero razonable para el parámetro debería de eliminarse este y volver a estimar el modelo sin incluirlo. De cualquier modo, si se tiene conocimiento de que dicho parámetro debe aparecer en el modelo, es aconsejable mantenerlo aunque la prueba no rechace la posibilidad de que el valor del parámetro sea cero.

Supuesto 7. El modelo sobreespecificado se dice que un modelo esta sobreespecificado cuando contiene coeficientes de más esto quiere decir que el modelo es más complicado de los que debiera y entonces hay que quitarle parámetros que están de más, se les llama redundantes o excesivos.

Tenemos entonces la hipótesis nula:

$$H_0: \beta_j = 0 \quad \beta_j = (\theta, \phi_j)$$

Esto es que el coeficiente del estimador del parámetro es insignificante y que podemos omitirlo del modelo, contra la hipótesis alternativa:

$$H_1: \beta_j \neq 0 \quad \beta_j = (\theta, \phi_j)$$

la cual indica que el parámetro es necesario para el modelo.

Se resuelve utilizando el estadístico t de los coeficientes.

$$T = \frac{\beta_j}{\sigma_{\beta_j}} \quad \beta_j = (\theta, \phi_j)$$

Que se comporta aproximadamente como una t-student con $(n-p-q)$ grados de libertad.

Por lo tanto aceptamos la hipótesis nula H_0 cuando: $|T| < 2$

Esto quiere decir que el parámetro está de más en el modelo y lo debemos eliminar para tener un modelo que cumpla con el principio de parsimonia.

Al contrario, rechazando H_0 si $|T| \geq 2$

El parámetro es significativo y debe formar parte del modelo.

3.9 Pronóstico

Hasta este momento, el modelo elegido ha pasado por las pruebas necesarias para asegurar que sea parsimonioso y que arroje resultados confiables, entonces, hemos llegado a la etapa donde se aplica el modelo seleccionado para obtener el pronóstico de la serie de fallos del STCM de la Ciudad de México.

Se pronostica para un horizonte de tiempo H , esto es pronosticando h periodos de tiempo a partir del tiempo t (al cual se llama origen), así a partir de la serie Y_1, Y_2, \dots, Y_t , obtendremos $Y_{t+1}, Y_{t+2}, \dots, Y_{t+h}$. Claro que estos pronósticos tienen un margen de error, estos se definen como $e_{t,(h)} = Y_{t+h} - Y_{t,(h)}$: Esto es, el error para el periodo $t+h$ es la diferencia entre el valor real y valor pronosticado.

3.10 Intervalos de confianza para los pronósticos

En lo referente a los límites de confianza se tiene que dado que se considera que $e_t \rightarrow N(0, \sigma_e^2)$, por lo tanto, los límites del $100(1 - \alpha)\%$ de confianza para $T(Y_{t+h})$, condicionados con el conocimiento de las observaciones Y_t, Y_{t-1}, \dots , viene a ser

$$T(Y_t) \pm z_{\alpha/2} \left(\sum_{j=0}^{h-1} \psi_j^2 \right)^{1/2}$$

donde: $z_{\alpha/2}$ es el punto porcentual tal que $P(Z > z_{\alpha/2}) = \alpha/2$; si $Z(0,1)$ el intervalo de confianza definido por la ecuación anterior se estima al sustituir a $\psi_1, \dots, \psi_{h-1}$ y σ_e por sus respectivas estimaciones.

Capítulo IV

Aplicación del modelo y toma de decisiones.

4.1 Metodología para obtener los fallos.

La metodología para la recopilación de fallos para las dos Líneas se resume en la siguiente figura:

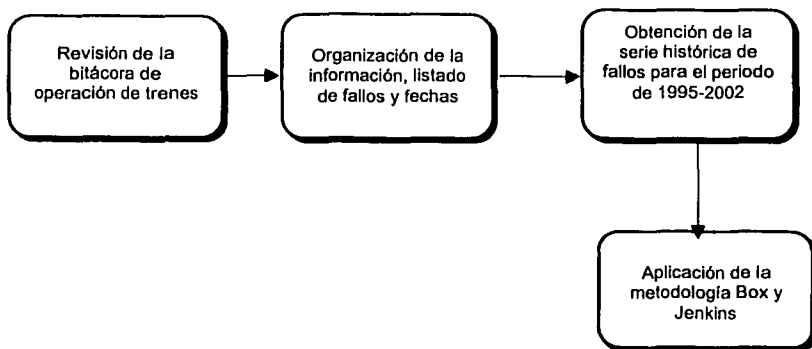


Figura 1: Metodología para la obtención de fallos.

TESIS CON
FALLA DE ORIGEN

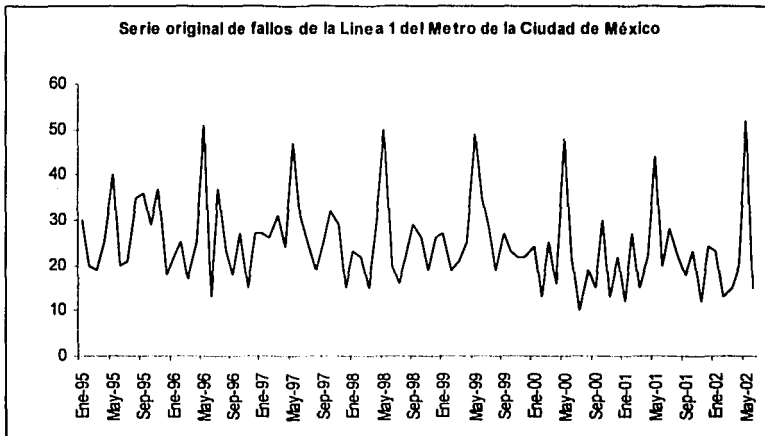
4.2 La serie de tiempo de la Línea 1

Esta serie tiempo representa los fallos de trenes de la Línea 1 del STCM de la Ciudad de México, para ajustar estos datos a un modelo ARIMA se agruparon en una forma mensual, registrados desde enero de 1995 hasta junio 2002.

Es decir, se cuenta con noventa observaciones a las cuales se les aplicara la metodología de Box y Jenkins. La gráfica 1 muestra los valores de los fallos a través del tiempo. Los datos de esta serie están en el anexo I .

Gráfica 1

Fallos del Sistema de Transporte Colectivo Metro del periodo de enero 1995 a junio 2002.



4.3 Identificación de la serie histórica para la Línea 1

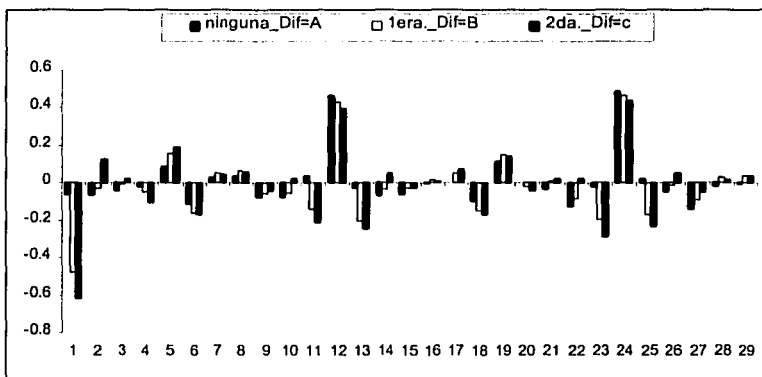
Como puede observarse en la gráfica 1, esta presenta un comportamiento con fluctuaciones que no presentan variabilidad creciente o decreciente a través del tiempo, lo cual es lógico suponer ya que los trenes tienen mantenimiento, se

observa en la gráfica anterior que no se requiere de una transformación estabilizadora de la varianza para la serie, por lo que no se utilizará ninguna transformación, ni la propuesta por Box-Cox en la sección (3.3)

Adicionalmente la serie del STCM para la Línea 1 se observa algún comportamiento estacional por lo que pudiera describirla un modelo **SARIMA** estacional. A continuación se estabiliza el nivel de la serie mediante la aplicación de diferencias ordinarias a la serie original sin transformar, como se menciona en la sección (3.4) para detectar cuál será la serie estacionaria gráfico los primeros 29 valores de **FAC**, los datos se encuentran en el anexo I .

Gráfica 2

Autocorrelaciones de la serie original y la 1era y 2da diferencia ordinaria.



Las **FACs** graficadas corresponden a las siguientes diferencias sobre la serie sin transformar:

La serie A son las **FACs** de serie original sin transformada

La serie B son las **FACs** de la primera diferencia ordinaria

**TESIS CON
FALLA DE ORIGEN**

La serie C son las **FACs** de la segunda diferencia ordinaria

De la anterior gráfica las autocorrelaciones más pequeñas corresponden a la serie A entonces se elige esta serie, por lo que ésta pasa a ser la serie estacionaria se contara con las 90 observaciones, para aplicar la metodología de Box y Jenkins.

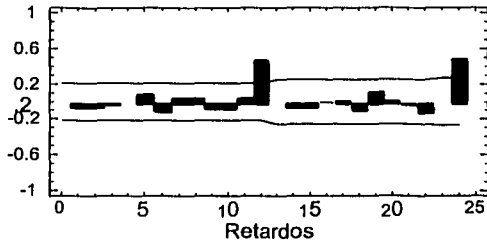
Por otra parte utilizando otra herramienta para precisar si realmente la serie escogida es estacionaria es a través de sus desviaciones estándar de las **FACs** de la serie original como también de la de primeras y segundas diferencias, estos datos se encuentran en el anexo I .

Serie	A	B	C
Desviación. Estándar.	0.141	0.156	0.167

Se obtiene que a través de este análisis la serie de menor desviación estándar es la serie A, con esta forma se ratifica que la serie A es estacionaria. Ya que con las otras series con la aplicación de diferencias aumento su variabilidad, a esta serie estacionaria se tendrá que ajustar un modelo **ARIMA** o **SARIMA**

Para identificar que proceso sigue esta serie de la Línea 1 del STCM, se analizan a continuación su **FAC** y **FACP** muestrales, en el anexo I se encuentran los primeros 29 valores de estas funciones, se calcularon con base a la formula mostrada en la ecuación (3.5.1) y la sustitución de estos valores en la expresión (2.4.9).

Función de Autocorrelación para la Línea 1

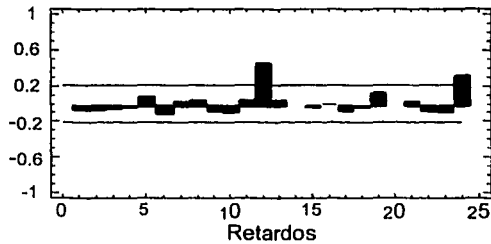


De la **FAC** las autocorrelaciones para la etapa estacionaria se encuentran dentro de la banda de significancia, no requiere de la aplicación de modelos de la parte estacionaria, se analiza la parte estacional de la serie de fallos de trenes.

Tiene valores significativos en $K=12$, $K=24$, entonces; hay que aplicar una diferencia de orden estacional se perderán algunos datos, según el orden de la estacionalidad de la serie.

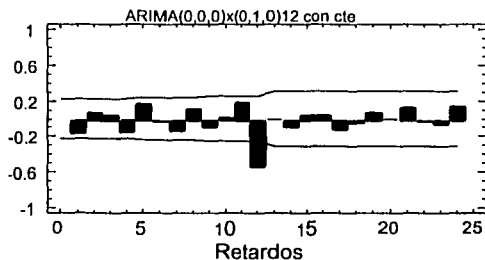
**TESIS CON
FALLA DE ORIGEN**

Función de Autocorrelación Parcial para la Línea 1



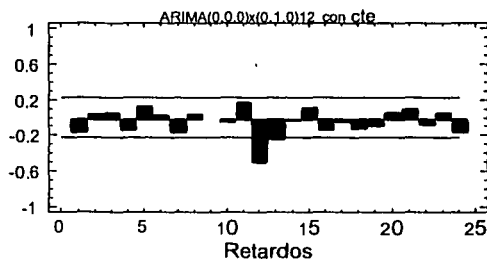
Por el comportamiento de la **FAC** y de la **FACP** muestrales no existen valores significativos, en el periodo estacionario ya que la **FAC** y la **FACP** tiene valores significativos en los retrasos $K=12$, $k=24$, lo que sugiere un comportamiento estacional, hay que aplicar diferencias de orden estacional y se obtiene la gráfica:

Función de Autocorrelación de los residuales para la Línea 1



La **FAC** de presenta valores significativos en los retrasos $K=12$, sugiere un comportamiento de un modelo de orden estacional **SMA(1)₁₂**, según la propuesta de la sección (2.12) y (2.13).

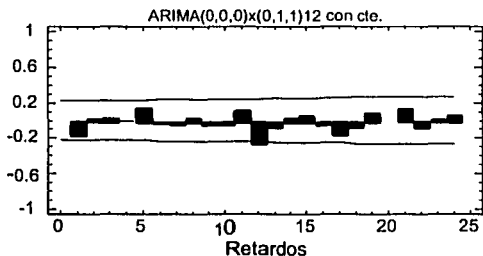
Función de Autocorrelación Parcial de los residuales para la Línea 1



TESIS CON
FALLA DE ORIGEN

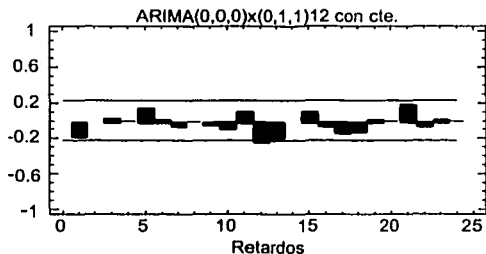
Ahora la **FACP** presenta valor significativo en el retraso $K=12$, también se puede observar un comportamiento decreciente lo cual sugiere suponer que tiene un comportamiento de un modelo de orden estacional **SMA(1)₁₂**.

Función de Autocorrelación de los residuales para la Línea 1



Después de la proposición de modelos de orden estacional la **FAC** de los residuales, el retraso $K=12$ toca la banda de significancia, todavía los residuales no son cero, párese que se comporta como **SAR(1)₁₂** como lo sugiere la sección (2.12) y (2.13), por el comportamiento decreciente de las autocorrelaciones

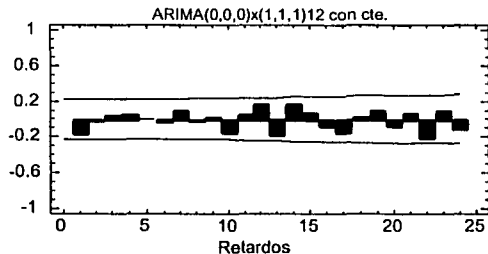
Función de Autocorrelación Parcial de los residuales para la Línea 1



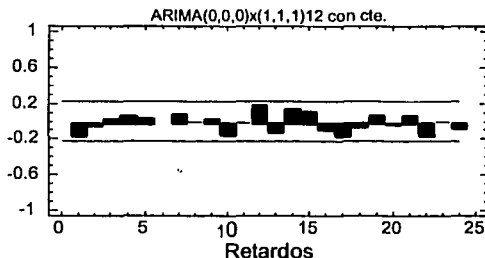
TESIS CON
FALLA DE ORIGEN

Observemos la anterior gráfica y la **FACP** de los residuales todavía el retardo $K=12$ toca la banda lo que sugiere la proposición de un **SAR(1)₁₂**.

Función de Autocorrelación de los residuales para la Línea 1



Función de Autocorrelación Parcial de los residuales para la Línea 1



Tanto la **FAC** y **FACP** de los residuales están todos dentro de la banda párese ser un buen modelo tentativo que explica los fallos de la Línea 1 del STCM de la Ciudad de México para el periodo de enero de 1995 a junio del 2002, esto se

TESIS CON
FALLA DE ORIGEN

verificará más adelante. Ahora se estima sus parámetros del modelo de la serie con el paquete estadístico.

4.4 Estimación de parámetros para la Línea 1

Los resultados de los parámetros que arroja el paquete Statgraphics plus son:

Modelo SARIMA (0,0,0)x(1,1,1)¹² con cte.

Parámetro	Estimación	Error Stnd.	t	P-value
SAR(1)	-0.61381	0.0759119	-8.08582	0.000000
SMA(1)	0.823074	0.0477709	17.2296	0.000000
Media	-0.893322	0.202913	-4.40248	0.000035
Constante	-1.44165			

Backforecasting: yes

Estimación varianza del ruido blanco =31.7126 con 75 grados de libertad

Estimación desviación estándar = 5.63139

Número de iteraciones: 7

4.5 Análisis de residuales para la Línea 1

Se inicia con la revisión de los supuestos para el análisis de residuales que se proponen en la sección (3.8) para aceptar que este es el modelo que explique los fallos de los trenes de la Línea 1 del STCM.

Supuesto 1: e_t tiene media cero. Los valores de la media y desviación estándar

son: $M(e_t) = -0.057$ $S_e = 5.303$

A continuación calculo el cociente:

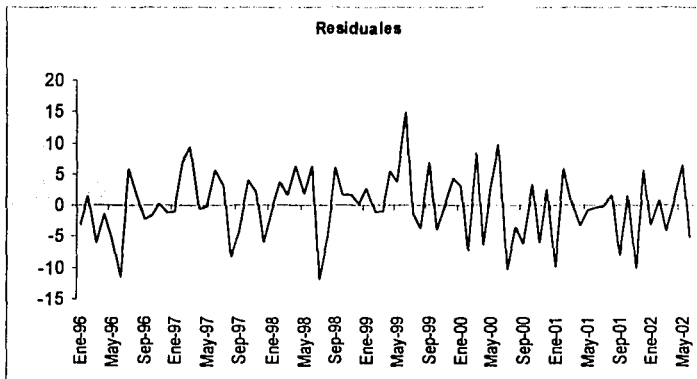
$$\left(\frac{N-d-p}{N} \right)^{1/2} M(e_t) / S_e = (90-0-0)^{1/2} * (-0.057) / (5.303) = -0.5$$

Como si cumple que $-0.5 < 2$ si se aprueba el supuesto.

Supuesto 2: e_t tiene varianza constante. Se presenta a continuación la gráfica de la serie de los errores aleatorios calculados previamente, ya que a través de su revisión sabré si se cumple con el supuesto.

Gráfica 3

Serie de residuales para la Línea 1 del Metro de la Ciudad de México

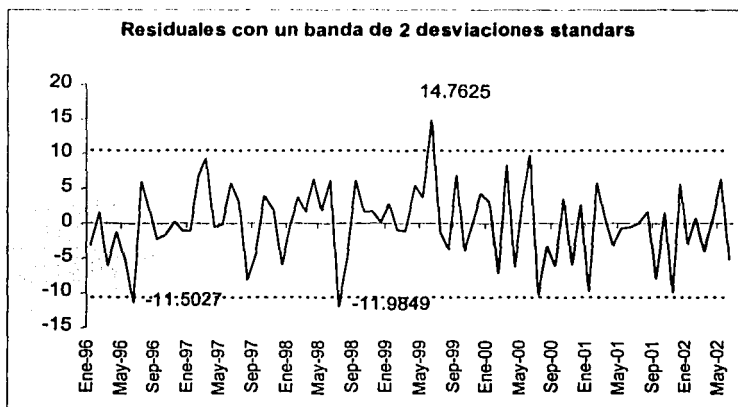


Como puede verse, la gráfica 3 no presenta una tendencia creciente ni decreciente si no un comportamiento estable, a través del tiempo por lo tanto la varianza si es constante y en consecuencia si se cumple el supuesto.

Supuesto 3: Las variables e_t son mutuamente independientes: Calculando las primeras 24 estimaciones de $r_i(e)$ obtengo el siguiente valor de Q que es de: 21.8085. Ahora si el valor de tablas es de $\chi^2_{(0.025, 24)} = 39.3641$, dado que $\chi^2 > Q$ no se rechaza el supuesto de independencia de las variables.

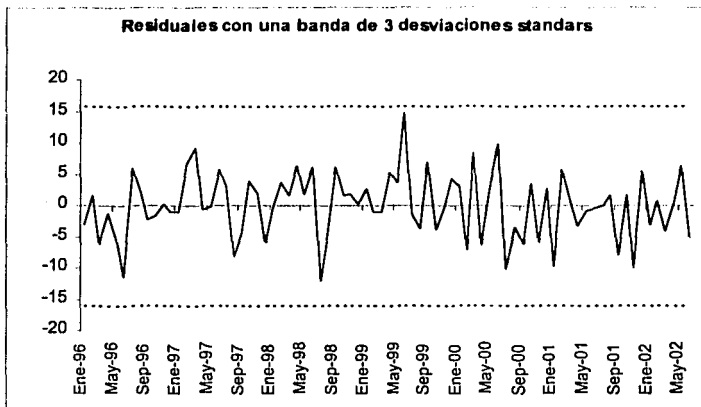
**TESIS CON
FALLA DE ORIGEN**

Supuesto 4: Los e_t se distribuye normal. Cuántos valores de la serie de choques aleatorios se encuentran fuera del intervalo por $(-2S, . 2S) = (-2*5.303, 2*5.303)=(-10.605, 10.605)$ Se observa que los puntos fuera del intervalo, con un 95% de confianza, solo son 3 observaciones y representa él 3.8% del total de las observaciones, como se muestra en la siguiente gráfica, por lo que sí se cumple con el supuesto, es decir que la serie de residuales sigue un comportamiento de una distribución normal.



Supuesto 5: No existen observaciones aberrantes todos los residuales están dentro del intervalo: $(-3S, . 3S)=(-3*5.303 3*503)=(-15.908, 5.908)$, con un 99% de confianza, como lo muestra la siguiente gráfica, entonces se acepta el supuesto.

**TESIS CON
FALLA DE ORIGEN**



Supuesto 6: El modelo es parsimonioso: Para verificar este supuesto se construyen los intervalos de confianza señalados en la sección (3.8.1).

Parámetro β	$[\text{Var}(\beta)]^{1/2}$	$(\beta - 2[\text{Var}(\beta)]^{1/2}, \beta + 2[\text{Var}(\beta)]^{1/2})$
-0.61381	(0.0759119)	(-0.7656338 -0.4619862)
0.823074	(0.0477709)	(0.7275322 0.9186158)
-0.89332	(0.202913)	(-1.299146 -0.487494)

Ya que el intervalo de confianza para el parámetro del **SAR(1)₁₂**, **SMA(1)₁₂** y la **media** no cubren el valor de cero se concluye que la hipótesis:

$$H_0: \beta_j = 0 \quad \beta_j = (\theta, \phi)$$

No se rechaza al nivel del 5%, es decir se requiere la inclusión de los parámetros para explicar el comportamiento de la serie de la Línea 1 del STCM, y el modelo en si resulta ser parsimonioso.

**TESIS CON
FALLA DE ORIGEN**

Supuesto 7. Se analiza el resultado de los parámetros de la metodología Box y Jenkins, arrojados por el paquete Statgraphics Plus que se encuentran en la sección (4.4).

➤ Análisis del estadístico “t” para el parámetro Φ :

$$|-8.08582| > 2$$

Entonces implica que Φ : *si* es significativo y por lo tanto no será eliminado.

➤ Para la *Media* se concluye:

$$|-4.40248| > 2$$

Entonces μ si es significativa por lo tanto no se elimina la constante δ

➤ Análisis del estadístico de “t” para “ Θ ”:

$$|17.2296| > 2$$

Entonces implica que Θ *si* es significativo y por lo tanto no será eliminado

➤ Condición de “Estacionaridad”

$$|-0.61381| < 1$$

Por lo tanto, la parte **SAR(1)₁₂** si cumple con la condición de Estacionaridad.

➤ Condición de “invertibilidad”

$$|0.823074| < 1$$

Por lo tanto la parte **SMA(1)₁₂**, si cumple con la condición de invertibilidad

Después de haber analizando cada uno de los anteriores supuestos y haber probado que cumplen, entonces se confirma que el modelo es confiable para

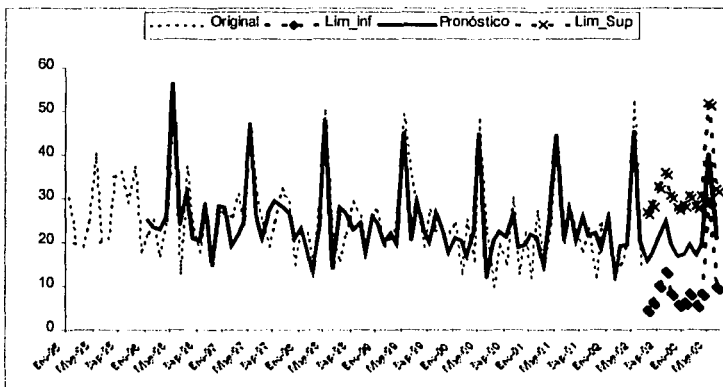
pronosticar los fallos de pilotaje automático para el STCM de la ciudad de México.

El Modelo es el siguiente: $SAR(1)_{12} * SMA(1)_{12}$

$$(I - \Phi_1 B^{12}) W_t = (I - \Theta_1 B^{12}) e_t + \delta \quad \text{donde } W_t = \nabla_{12}^1 Y_t$$

Con $\Phi = -0.61381$, $\Theta = 0.823074$ y e_t es ruido blanco. Ahora los datos originales se grafican junto con el pronóstico de fallos de trenes para la Línea 1 obtenido por la metodología Box y Jenkins, con sus intervalos del 95% confianza y observar si tienen un comportamiento parecido al dato real.

Gráfica 4
Serie original y el pronóstico con sus intervalos para fallos de la Línea 1.



Los datos de la anterior gráfica se encuentran en el anexo I, el comportamiento del pronóstico hacia atrás y hacia delante es muy parecido al original, y este pronóstico tiene un error cuadrático medio de: 5.37326

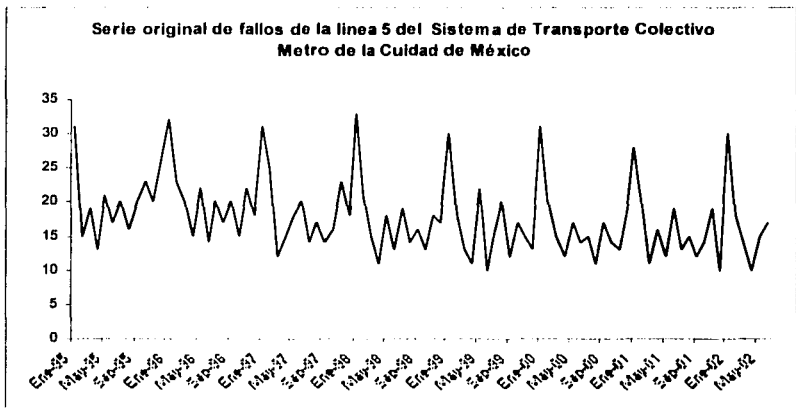
**TESIS CON
 FALLA DE ORIGEN**

4.6 La serie de tiempo de la Línea 5

Esta serie de tiempo representa los fallos de trenes del STCM de la Línea 5, para ajustar estos datos a un modelo ARIMA se agrupan en una forma mensual registrados desde enero de 1995 hasta junio 2002, es decir, se cuenta con noventa observaciones. La siguiente gráfica muestra los valores de los fallos de los trenes, los datos se encuentran en el anexo II

Gráfica 5

Fallos del Sistema de Transporte Colectivo Metro, para el periodo de enero 1995 a junio 2002.



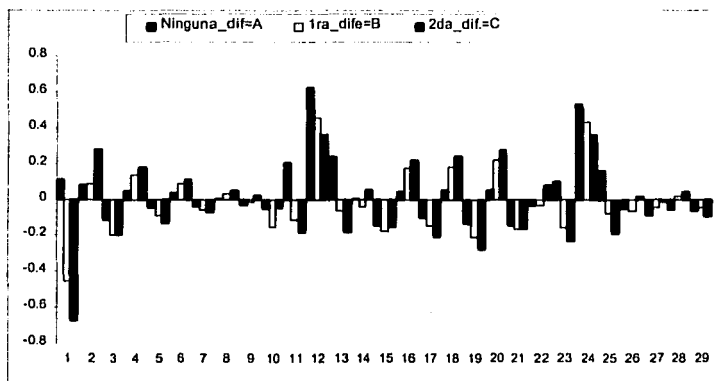
4.7 Identificación de la serie histórica para la Línea 5

Como puede observarse en la gráfica 5, esta presenta un comportamiento parecido a la Línea 1, la serie de tiempo vista anteriormente, *con fluctuaciones que no presentan variabilidad creciente o decreciente a través del tiempo, lo cual es lógico suponer ya que los trenes tienen mantenimiento, razón por la cual no se*

TESIS CON
FALLA DE ORIGEN

requeriré de una transformación estabilizadora de la varianza, por lo que no se utilizara ninguna transformación, ni la propuesta por Box-Cox en la sección (3.3). Adicionalmente se observa que la serie en estudio tiene algún comportamiento estacional por lo que pudiera ser un modelo **SARIMA** estacional. A continuación se estabiliza el nivel de la serie mediante la aplicación de diferencias ordinarias a la serie original sin transformar, como se explica en la sección (3.4). Para detectar cuál será la serie estacionaria se grafican los primeros 29 valores de **FAC**. Los datos se encuentran en el anexo II .

Gráfica 6
Autocorrelaciones de la serie original y la 1era, 2da diferencia ordinaria.



Las **FACs** graficadas corresponden a las siguientes diferencias sobre la serie sin transformar:

TESIS CON
FALLA DE ORIGEN

La serie A son las **FACs** de serie original sin transformada.

La serie B son las **FACs** de la primera diferencia ordinaria.

La serie C son las **FACs** de la segunda diferencia ordinaria.

Observando la gráfica 6 las autocorrelaciones más pequeñas corresponden a la serie A, entonces se elige esta serie. Por lo que ésta pasa a ser la serie estacionaria y no se le aplica ninguna diferencia, se contara con los 90 datos originales, para aplicar la metodología Box y Jenkins.

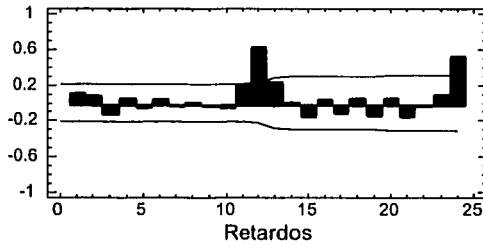
Otra herramienta para precisar si realmente la serie elegida ya esta estacionaria es calculando sus desviaciones estándar de las **FACs** de la serie original como también de la de primeras y segundas diferencias, los datos se encuentran en el anexo II .

Serie	A	B	C
Desviación. Estándar.	0.180	0.188	0.228

Se obtiene que a través de este análisis también la serie de menor desviación estándar es la serie A de esta forma se ratifica que la serie A como estacionaria, ya que con las otras series con la aplicación de diferencias aumento su variabilidad. A esta serie estacionaria se le ajustara un modelo **ARMA** o **SARIMA**, para identificar que proceso sigue esta serie, ahora se analiza a continuación su **FAC** y **FACP**.

Los valores de estas funciones se encuentran en el anexo II, están los primeros 29 valores, se calculan con base en la fórmula mostrada en la ecuación (3.5.1) y la sustitución de estos valores en la expresión (2.4.9).

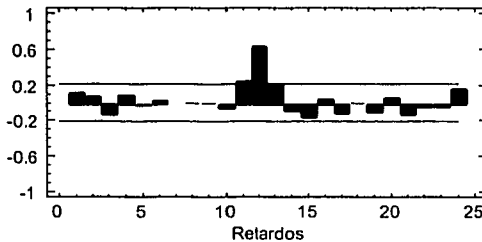
Función de Autocorrelación para la Línea 5



En la FAC hay valores significativos en $K=12$, $k=24$, que es la parte estacional de la serie, la parte estacionaria no hay algún retraso significativo, la serie sugiere la aplicación de diferencias de orden estacional.

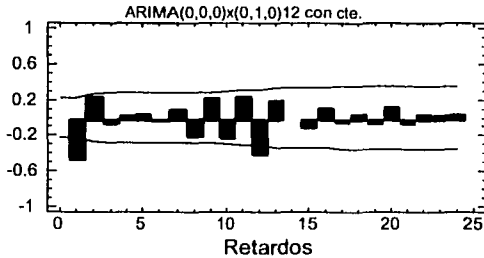
TESIS CON FALLA DE ORIGEN

Función de Autocorrelación Parcial para la Línea 5



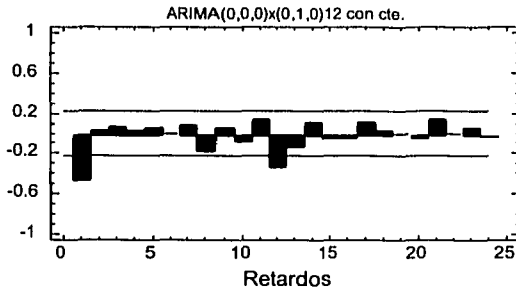
También la **FACP** tiene un valor significativo en el retraso $K=12$, ahora se le aplica diferencias de orden estacional.

Función de Autocorrelacion de los residuales para la Línea 5



Después de la aplicación de una diferencia de orden estacional la **FAC** de los residuales tiene un valor significativo en los retrasos $K=1$, $K=12$, y tienen un comportamiento decreciente entonces sugiere la aplicación de un modelo de orden **AR(1)** como se señala en la sección (2.12) y (2.13).

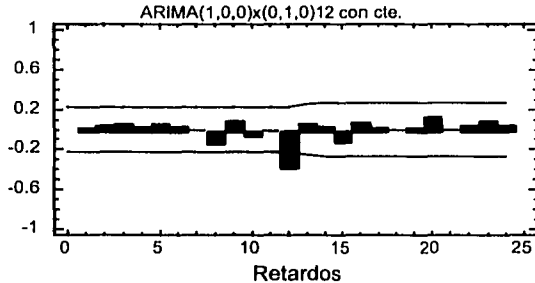
Función de Autocorrelacion Parcial de los residuales para la Línea 5



TESIS CON
FALLA DE ORIGEN

La **FACP** tiene un comportamiento que se corta después de p intervalos y tiene valores significativos en los retrasos $K=1, K=12$, lo que sugiere un **AR(1)** para la parte estacionaria como se indica en la sección (2.12) y (2.13).

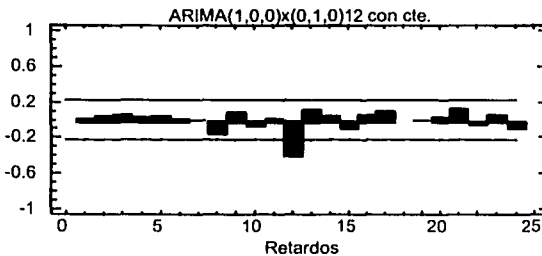
Función de Autocorrelación de los residuos para la Línea 5



TESIS CON FALLA DE ORIGEN

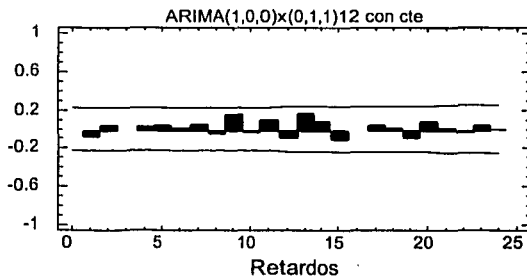
Después de haber sido propuesto un modelo estacionario **AR(1)**, ahora analizamos la parte estacional de la serie comenzando con la **FAC** de los residuales presenta valores significativos en el retardo $K=12$, ahora podría parecerse a un modelo de orden estacional **SMA(1)₁₂** por el comportamiento decreciente de las autocorrelaciones como lo sugiere la sección (2.12) y (2.13).

Función de Autocorrelacion Parcial de los residuos para la Línea 5



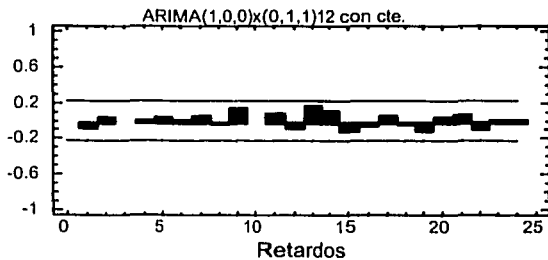
En la **FACP** de los residuales hay un valor significativo en el retraso $K=12$, y los demás valores de las autocorrelaciones son decrecientes lo que sugiere la gráfica un modelo de orden **SMA(1)₁₂**.

Función de Autocorrelación de los residuales para la Línea 5



Se observarse que los residuales se encuentran dentro de la banda de significancia de la **FAC**.

Función de Autocorrelación Parcial de los residuales para la Línea 5



TESIS CON
FALLA DE ORIGEN

También puede observarse que en la anterior gráfica los residuales de la **FACP** están dentro de la banda de significancia, entonces los residuales para ser que son ruido blanco, hay que verificar este modelo si es el óptimo, con los supuestos expuesto en la sección (3.7).

4.8 Estimación de parámetros para la Línea 5

Los resultados de los parámetros que arroja el paquete Statgraphics plus son:

Modelo ARIMA (1,0,0)x(0,1,1)¹² con cte.

Parámetro	Estimación	Error Stnd.	t	P-value
AR(1)	-0.41547	0.106576	-3.89834	0.000209
SMA(1)	0.863378	0.0408985	21.1102	0.000000
Media	-0.721461	0.0935557	-7.71157	0.000000
Constante	-1.02121			

Backforecasting: yes

Estimación varianza del ruido blanco = 7.92706 con 75 grados de libertad

Estimación desviación estandar del ruido blanco = 2.8155

Número de iteraciones: 6

4.9 Análisis de residuales para la Línea 5

Se inicia con la revisión de los supuestos para el análisis de residuales que se proponen en la sección (3.8), para aceptar si este modelo se ajusta a los fallos de trenes de la Línea 5.

Supuesto 1: e_t tiene media cero. Los valores de la media y desviación estándar

son: $M(e_t) = -0.052$ $S_{e_t} = 2.345$

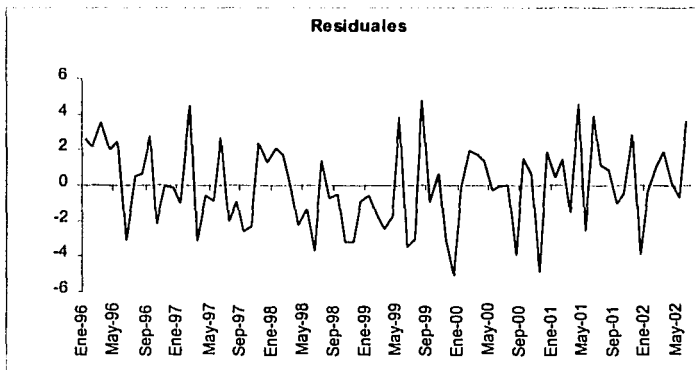
A continuación se calcula el cociente:

$$\left(\frac{N-d-p}{N} \right)^{1/2} M(e_t) / S_{e_t} = (90-0-1)^{1/2} \times (-0.052) / 2.345 = -0.97$$

Como si cumple que $-0.97 \leq 2$ si se aprueba el supuesto.

Supuesto 2: e_t tiene varianza constante. Se presenta a continuación la gráfica de la serie de los errores aleatorios calculados previamente, ya que a través de su revisión sabré si se cumple con el supuesto.

Gráfica 7
Serie de residuales de la Línea 5 del Metro de la Ciudad de México

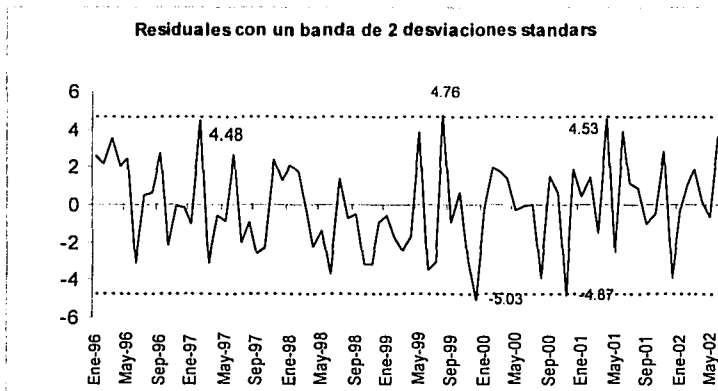


Como puede observarse, la gráfica 7 no presenta una tendencia creciente ni decreciente si no un comportamiento estable a través del tiempo por lo tanto la varianza si es constante y por lo tanto si se cumple el supuesto.

Supuesto 3: Las variables e_t son mutuamente independientes: Calculando las primeras 24 estimaciones de $r_1(e)$, se obtiene el siguiente valor de Q' que es de: 9.3898. Ahora el valor de tablas es de $X^2_{(0.025, 24)} = 39.3641$ dado que $X^2 > Q'$ no se rechaza el supuesto de independencia de las variables.

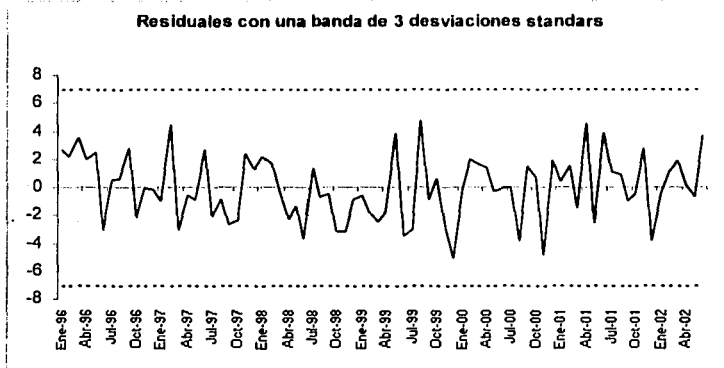
TESIS CON
FALLA DE ORIGEN

Supuesto 4: e_t se distribuye normal. Cuántos valores de la serie de choques aleatorios se encuentran fuera del intervalo por $(-2s, 2s) = (-2 \cdot 2.345, 2 \cdot 2.345) = (-4.7, 4.7)$ Los puntos fuera del intervalo con un 95% de confianza, son solo 3, que representa el 3.8% del total de las observaciones, como se muestra en la siguiente gráfica, por lo que se puede decir que la serie de residuales tienen un comportamiento de una distribución normal, por lo que se acepta el supuesto.



Supuesto 5: No existen observaciones aberrantes. Todos los residuales están dentro del intervalo: $(-3s, 3s) = (-3 \cdot 2.345, 3 \cdot 2.345) = (-7, 7)$, con un 99% de confianza, como lo muestra la siguiente gráfica; se acepta el supuesto.

**TESIS CON
FALLA DE ORIGEN**



Supuesto 6: El modelo es parsimonioso. Para verificar este supuesto se construyen los intervalos de confianza señalados en la sección (6.3.1).

$$\text{Parámetro } \beta \quad [\text{Var}(\beta)]^{1/2} \quad \left(\beta - 2[\text{Var}(\beta)]^{1/2}, \beta + 2[\text{Var}(\beta)]^{1/2} \right)$$

-0.41547	(0.106576)	(-0.628622	-0.202318)
0.863378	(0.0408985)	(0.781581	0.945175)
-0.721461	(0.0935557)	(-0.9085724	-0.5343496)

Ya que el intervalo de confianza para el parámetro del **AR(1)**, **SMA(1)₁₂** y la **media** no cubren el valor de cero se concluye que la hipótesis:

$$H_0: \beta_i = 0 \quad \beta_j = (\theta, \phi)$$

No se rechaza al nivel del 5%, es decir se requiere la inclusión de los parámetros para explicar el comportamiento de la serie de fallos de trenes, y el modelo en si resulta parsimonioso.

TESIS CON
 FALLA DE ORIGEN

Supuesto 7. Se analiza el resultado de los parámetros de la metodología Box y Jenkins, arrojados por el paquete Statgraphics Plus, que se encuentran en la sección (4.8)

➤ Análisis del estadístico "t" para el parámetro ϕ :

$$-3.89834 > 2$$

Entonces implica que ϕ si es significativo y por lo tanto no será eliminado.

➤ Para la *Media* se concluye:

$$-7.71157 > 2$$

Entonces μ si es significativa por lo tanto no se elimina la constante δ

➤ Análisis del estadístico de "t" para " Θ " :

$$21.1102 > 2$$

Entonces implica que Θ si es significativo y por lo tanto no será eliminado

➤ Condición de "Estacionaridad "

$$-0.41547 < 1$$

Por lo tanto, la parte **AR(1)** si cumple con la condición de Estacionaridad.

➤ Condición de "invertibilidad "

$$0.863378 < 1$$

Por lo tanto la parte **SMA(1)₁₂**, si cumple con la condición de invertibilidad

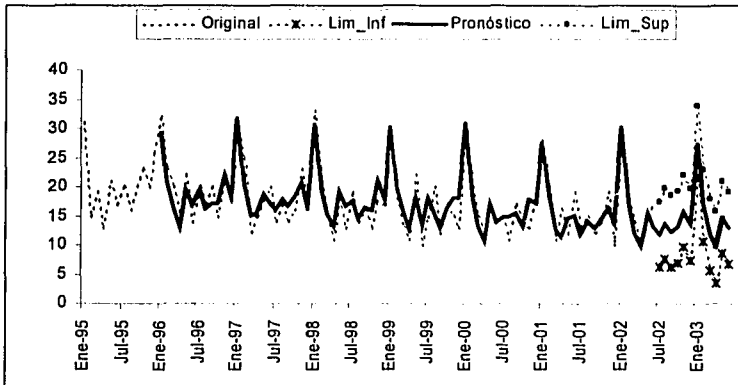
Después de haber analizando cada uno de estos anteriores supuestos y haber probado que cumplen, entonces se puede decir que es el modelo es confiable

para pronosticar los fallos de pilotaje automático para el STCM de la ciudad de México. El Modelo es el siguiente: $AR(1)*SMA(1)_{12}$

$$(I - \phi_1 B^1)W_t = (I - \Theta_1 B^{12})a_t + \delta \quad \text{donde: } W_t = \nabla^1 Y_t$$

Con $\phi = -0.41547$, $\Theta = 0.863378$ y a_t es ruido blanco. Ahora los datos originales los graficamos con el pronóstico de fallos para la Línea 5, con sus intervalos de confianza y observar si tienen un comportamiento parecido al dato real de la serie original, como lo muestra la siguiente gráfica.

Gráfica 8
Serie original y el pronóstico con sus intervalos para fallos de la Línea 5



Los datos de la anterior grafica se encuentran en el anexo II, el pronóstico obtenido para los fallos del STCM de la Línea 5 es muy parecido hacia atrás como hacia delante con la serie original, este pronóstico fue para el periodo de julio 2002 a diciembre 2002, con un error cuadrático de: 2.37623.

**TESIS CON
FALLA DE ORIGEN**

4.10 La disponibilidad de trenes en las Líneas 1 y 5

Para la Línea 1 y la Línea 5 del Sistema de Transporte Colectivo Metro, cuentan con 43 y 17 trenes de diferentes modelos respectivamente, son los que dispone para cubrir la demanda de usuarios para el año 2002, según sus propias estimaciones de la gerencia de ingeniería y desarrollo.

Se muestran en la siguiente tabla la afluencia estimada de usuarios de una manera mensual para estas dos líneas, para todo el año 2002.

Tabla 2
Proyección de afluencia de usuarios para el año 2002⁷

Mes	Línea 1	Línea 5
Enero	23,300,344	6,666,726
Febrero	20,853,649	5,970,546
Marzo	22,308,178	6,367,097
Abril	22,714,644	6,559,337
Mayo	23,207,585	6,714,593
Junio	22,316,471	6,445,810
Julio	23,443,726	6,831,904
Agosto	23,357,393	6,800,439
Septiembre	22,038,482	6,430,393
Octubre	23,367,756	6,867,636
Noviembre	22,219,102	6,503,059
Diciembre	22,773,404	6,705,261
Total	271,900,734	78,862,801

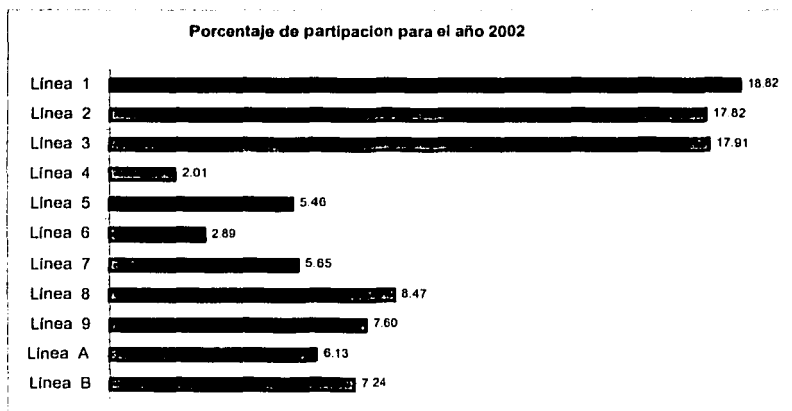
TESIS CON
FALLA DE ORIGEN

La falta de algunos de los trenes asignados para las líneas provocaría que los usuarios del STCM de la Ciudad de México se vieran afectados y arrojara este contratiempo, aglomeraciones en las horas picos, menos trenes están disponibles

⁷ Proyección realizada por el departamento de ingeniería y desarrollo del STCM de la Ciudad de México.

para cubrir la demanda diaria y los trenes que estuvieran disponibles realizarían más vueltas, teniendo con esto un mayor desgaste del tren en cada una de las dos Líneas. La gerencia de ingeniería y desarrollo estimo para todo el año 2002 la demanda servicio será en las 11 Líneas que comprende el STCM de la Ciudad de México será de 1,444,552,115 usuarios, de estos la Línea 1 tendrá una afluencia de 271,900,734 usuarios lo que representa el 19% del total de la red del sistema. La Línea 5 tendrá una afluencia de 78,862,801 usuarios para este periodo, esta cantidad de gente es el 5% del total estimado por del departamento de ingeniería y desarrollo, la siguiente gráfica muestra de una forma más clara la participación que espera que se de en el STCM para todas las 11 Líneas que comprende la red.

Gráfica 10
Porcentaje de afluencia de usuarios en toda la red del STCM



Para el sistema en su proyección del año 2002 la Línea 1 espera que tenga la

**TESIS CON
 FALLA DE ORIGEN**

mayor demanda de usuarios en todo este año ocupando el primer lugar en todas las 11 Línea de la red, y la Línea 5 está en un noveno lugar.

Para cubrir toda esta demanda del servicio del STCM en el año los trenes tiene que realizar un numero determinado de vueltas, también el departamento de ingeniería y desarrollo estima que cantidad de vueltas debe realizar los trenes de las Líneas 1 y 5, como lo muestra la siguiente tabla:

Tabla 3
Proyección del Numero vueltas para el año 2002⁸

Mes	Línea 1	Línea 5
Enero	14,445	6,391
Febrero	12,948	5,749
Marzo	13,874	6,276
Abril	14,126	6,208
Mayo	14,445	6,391
Junio	13,915	6,155
Julio	14,625	6,422
Agosto	14,594	6,400
Septiembre	13,766	6,146
Octubre	14,625	6,422
Noviembre	13,915	6,155
Diciembre	14,265	6,360
Total	169,543	75,075

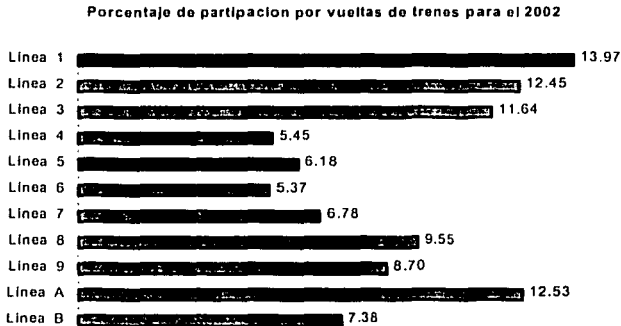
TESIS CON
FALLA DE ORIGEN

En todo el año 2002 todos los trenes del sistema van a realizar un total de 1,213,825 vueltas y para la Línea 1 que comprende su recorrido desde Observatorio a Pantitlán los trenes asignados realizarán un total de 169,543 vueltas durante solamente para el año 2002, para la Línea 5 que comprende desde Pantitlán a Politécnico sus trenes asignados harán 75,075 vueltas para este

⁸ Realizada por el departamento de ingeniería y desarrollo del STCM, para el año 2002

año también. Del total de vueltas que van realizar los trenes el STCM la Línea 1 representa el primer lugar con 14% de vueltas en toda la red, a diferencia que la Línea 5 es el 6%, ocupando un noveno lugar en toda la red del sistema como lo muestra la siguiente gráfica:

Gráfica 11
Porcentaje de vuelta por tren en toda la red del STCM



La cantidad de usuarios que puede ser transportados por un tren del STCM de la Ciudad de México, se puede calcular en base que cada tren cuenta 2 vagones con motriz que tienen cabina, 4 vagones con motriz que no tienen cabina, 2 vagones que son remolques y un vagón con pilotaje automático con remolque, si se suman estos vagones en su totalidad nos dan que un tren está compuesto por nueve vagones y la distribución de los usuarios en cada vagón se da de la siguiente manera⁹:

**TESIS CON
FALLA DE ORIGEN**

⁹ Datos proporcionados por el departamento de ingeniería

	Motriz con cabina (M)	Motriz sin cabina (N)	Remolque (R)
Pasajeros sentados	38	39	39
Pasajeros de pie	132	131	131

El total de personas que transporta un tren es de 1,530 usuarios por viaje suponiendo, que esta cantidad de usuarios aborda los trenes tanto para la Línea 1 como para la Línea 5 en sus terminales correspondientes, los usuarios llenaran por completo todos los nueve vagones del tren, la cantidad de usuarios anterior es considerando el recorrido del tren de terminal a terminal, para el caso de la Línea 1 es de observatorio a pantitlan, sus trenes hacen un recorrido de 18.18 Km. Y la Línea 5 es desde pantitlan a politécnico, estos trenes recorren en un viaje la distancia de 15.7 Km. Para dar servicio a esta cantidad de usuarios el STCM tendrá que llevar a cabo programas de mantenimiento preventivo y correctivo para tener la mayor cantidad de trenes disponible para cubrir la demanda mensual estimada.

Ya que los fallos en pilotaje automático pronosticados para el periodo julio 2002 a diciembre de 2002 obtenidos por el modelo de Box y Jenkins son los que se muestran en la siguiente tabla:

Tabla 4
Pronósticos de fallos de Pilotaje Automático

**TESIS CON
FALLA DE ORIGEN**

Meses	Pronóstico Línea 1			Pronóstico Línea 5		
	Lim Inferior	Probable	Lim Superior	Lim Inferior	Probable	Lim Superior
Jul-02	4	16	27	6	12	17
Ago-02	6	17	29	8	14	20
Sep-02	10	21	33	6	12	19
Oct-02	13	25	36	7	13	19
Nov-02	8	19	31	10	16	22
Dic-02	6	17	28	7	14	20

Si se considera que ocurra el evento más probable con un 95% de confianza y no él más optimista ni el más pesimista en ambas Líneas del STCM, para este periodo pronosticado, la gente que será afectada por la indisponibilidad de los trenes en la Línea 1 y 5 por la causa del fallo en el pilotaje automático es la que se muestra en la siguiente tabla:

Tabla 5
Usuarios Afectados por Fallos en PA

Meses	Línea 1	Línea 5
Julio	1,954,057	569,240
Agosto	1,372,800	485,791
Septiembre	1,050,256	535,552
Octubre	973,162	528,086
Noviembre	1,169,004	406,495
Diciembre	1,423,632	478,516
Total	7,942,931	3,004,103

El total de usuarios que serán afectado en estos seis meses pronosticados es 7,942,931 para la Línea 1 y 3,004,103 para la Línea 5, para estos seis meses pronosticados, el porcentaje participación de usuarios afectados en las dos líneas, es de:

Tabla 6
Porcentaje de usuarios afectados

Meses	Línea 1	Línea 5
Julio	8.34	8.33
Agosto	5.88	7.14
Septiembre	4.77	8.33
Octubre	4.17	7.69
Noviembre	5.26	6.26
Diciembre	6.25	7.14
Total	5.79	7.49

**TESIS CON
FALLA DE ORIGEN**

Los datos de la tabla 6 muestran que en la Línea 1 los meses con más alto porcentaje es julio, con un 8%, agosto y diciembre con un 6%, entonces de cada cien usuarios en estos meses al utilizar esta línea del STCM ocho se verán afectados en el mes de julio y seis en los meses de agosto y diciembre por el fallo en PA, esto tiende a crecer exponencialmente según la demanda del servicio.

En esta línea se realizan transbordos de usuarios con las líneas 2,3,4,7,8,9 y A. Donde el principal motivo para usar el Metro¹⁰ esta en primer lugar el desplazamiento del trabajador con un 40%, en segundo lugar el regreso a casa con un 16%, seguidos por motivos de diversión que representa un 12%. Las causas escolares el 10%, la realización de tramites un 5% y por último las compras con un 4.5%.

Esta línea como se menciona anteriormente tiene conexión con otras líneas y entre ellas están las que tienen la mayor cantidad de demanda de usuarios estimada para el 2002. Esta situación genera un exceso en la demanda de servicio para la Línea 1 sobre todo en los horarios denominados "hora pico" (de las 5.:00 a las 10:00 y de las 17:00 a las 21:00 horas) durante los cuales el STCM tiene la necesidad de dosificar el acceso de pasajeros en los andenes.

Para este periodo pronosticado de seis meses del total estimado para la Línea 1 del STCM es el 6% de los usuarios se verán afectados por fallos en PA, como lo muestra la tabla 6, esta cantidad de personas durante su transcurso de usar este medio de transporte se percatara de aglomeraciones en los andenes, retardo en la velocidad de tren, perdida de su tiempo, el retardo en salir de una estación y

¹⁰ <http://www.df.gob.mx/ciudad/reportajes/metro/07.html> 6 de agosto 2002

estos fallos para el STCM representa un tren indisponible para cubrir la demanda de usuarios en la Línea 1. Esta situación se incrementara ya que esta Línea tiene trasbordo con la Línea 2 y 3 que son junto con la Línea 1 la de mayor participación de afluencia de usuarios para el STCM.

La Línea 1 tendrá que llevar a cabo programas de mantenimiento correctivo y preventivo en los meses de julio, agosto y diciembre, ya que son los meses más significativos, para asegurar la mayor cantidad de trenes disponibles y reducir la cantidad de usuarios que están afectados por los fallos en Pilotaje Automático en ese periodo.

Siguiendo ahora con la Línea 5 aquí la tabla 6 muestra el porcentaje de usuarios afectados por el fallo en PA, los meses más representativo son julio, septiembre y octubre con un 8%, esta línea cuenta con un número menor de trenes que le fueron asignados, para cubrir la demanda del servicio en el año de 2002, en comparación con la Línea 1, si alguno de los trenes que tiene la línea esta indisponible para prestar el servicio, los restantes tendrán que realizar una mayor cantidad de vueltas para cubrir la demanda del servicio, teniendo con ello un mayor desgaste y con ello aumenta la probabilidad que tenga más fallos y así estarán afectando a una mayor cantidad de usuarios que utiliza esta Línea.

En la Línea 5 se realizan trasbordos con las Líneas 1,3,4 y 6, y de estas cuatro Líneas la más representativa son las Líneas 1 y 3, ya que ocupa el primer lugar y la otra el segundo lugar en atención de usuarios con un 19% y 18% respectivamente para el año 2002, respecto a los otras dos Líneas la 4 y 6 si se suma su participación para este periodo se obtiene el 5%, que es el total de participación para la Línea 5 en este año, como se observa en la gráfica 10.

Para el periodo pronosticado también de seis meses, como lo muestra la tabla 6 en esta Línea es el 7.4% de los usuarios que utilicen este medio de transporte en estos meses con los motivos que anteriormente ya se mencionaron, estarán afectados por fallos.

Estos usuarios se percataran de las incidencias que ocasiona los fallos en los trenes, durante su recorrido al usar esta línea, esta demanda del servicio podría incrementarse ya que la Línea 5 tiene transbordo con otras Líneas, entonces el número de usuarios afectados se incrementara.

La Línea 5 tendrá que llevar acabo tareas de mantenimiento preventivo y correctivo en los meses de julio, septiembre y octubre, ya que estos meses por si solos son más significativos con un 8% con respecto del total en este periodo de seis meses estimados, como se puede observar en la tabla 6, para que el STCM asegure la mayor cantidad de trenes disponibles y reducir la cantidad de usuarios que son afectados al utilizar este medio de transporte cualquiera que sean sus motivos al abordarlo.

Considerando lo anterior expuesto para la Línea 1 y 5 del STCM de la Ciudad de México, la estructura de la demanda del servicio esta conformado por tres bloques bien diferenciados, por un lado las Líneas 4,5,6 y 7 que poseen una capacidad de atención de usuarios holgada. Frente al otro grupo que tiene una demanda real que son las Líneas 1,2 y 3. Y el otro grupo de las Líneas de demanda intermedia que comprende las Líneas 8,9, "A" y "B".

Esta estructura de demanda del servicio por el STCM pudiera ser una causa de que ciertas Líneas no fueran atendidas para su mantenimiento preventivo y

correctivo de los trenes, ya sea por la falta de materiales, recursos humanos, financieros, etc. Ya que ciertas Líneas tienen la prioridad de brindar el servicio al ciudadano que utiliza el STCM de la Ciudad de México, por la cantidad de usuarios que demandan este medio de transporte.

Conclusiones

Los pronósticos son una de las herramientas para la toma de decisiones dentro de las organizaciones tanto productivas como sin fines de lucro. Ya que uno de sus propósitos es reducir la incertidumbre acerca de lo que puede acontecer en el futuro proporcionando información cercana a la realidad, que permita tomar decisiones sobre los cursos de acción tanto en el presente como en el futuro. El enfoque de Box y Jenkins, para el análisis de series de tiempo es una herramienta para proporcionar pronósticos a corto y mediano plazo, los modelos ARIMA presentados por Box y Jenkins son prácticos pues la metodología que se sigue para modelar una serie histórica verifica la bondad del modelo seleccionado.

El STCM de la Ciudad de México para estas dos Líneas tiene asignado una determinada cantidad de trenes para cubrir la demanda del servicio durante este año, también la cantidad de vueltas debe realizar, si alguno de estos trenes esta indisponible para cubrir la demanda del servicio, los restantes tendrán que realizar más vueltas y con ello aumenta más la probabilidad que ocurra un fallo en PA durante su recorrido, y con esto los usuarios que utilicen este medio de transporte se verán afectados durante su recorrido dentro de la red del STCM, esta incidencia será más grave si ocurre en las horas pico, que tiene ya detectas el STCM. Las Líneas 1 y 5 fueron diseñadas considerando una determinada captación y transito de usuarios, sin embargo hay estaciones que con el tiempo la prevención fueron sobrepasadas por la demanda, y para esto el STCM tiene diseñado tres bloques uno de demanda alta donde esta la Línea 1 y con demanda holgada donde esta la Línea 5, y el otro grupo de demanda intermedia. Este

diseño en la captación de usuarios podría afectar en el momento de brindar mantenimiento preventivo o correctivo a los trenes, ya que hay Líneas que tienen una demanda alta y otras baja, entonces no son atendidas con la misma prioridad ya que unas tienen más trenes, más demanda de usuarios, la línea es más extensa, etc. Las Líneas de baja demanda, en este caso para la Línea 5 tienen relación con las de alta demanda del servicio y esto ocasionará que en las instalaciones de estas Líneas haya una considerable cantidad de usuarios, y estos estarán afectados por los fallos de los trenes.

Nos podemos guiar por el valor puntual obtenido, aunque debo señalar que podemos guiarnos por el intervalo de confianza pues el nos puede dar una idea amplia del evento optimista o pesimista que podría ocurrir en el futuro, cuando ocurran los fallos en PA, para saber en que momento se toma la decisión de implementar el mantenimiento preventivo o correctivo a los trenes, ya que estos fallos me dirán que cantidad de usuarios son los afectados.

El porcentaje de usuarios afectados para estos seis meses pronosticados es mayor para la Línea 5 en relación con la Línea 1, el STCM tendrá que considerar que esta Línea tiene transbordo de usuarios con las dos Líneas que tendrán más demanda del servicio durante el año 2002 que son la misma Línea 1, junto con la Línea 3. También el STCM para la Línea 1, tendrá que tomar en cuenta que esta Línea tiene conexión con las Líneas 2 y 3 que ocupan el 3 y 2 lugar de afluencia para el año 2002, según sus propias estimaciones.

El Sistema de Transporte Colectivo Metro de la Ciudad de México, tiene bajo su responsabilidad la atención de usuarios desde el momento de su ingreso a la

estación de origen, su transportación y su salida por la estación de destino, siendo esta la función principal para la cual fue creado.

Bibliografía

- Guerrero, M. Victor, *Análisis estadístico de series económicas*, México, Universidad Autónoma Metropolitana, 1991.
- Box, George E. P. y Gwilym M Jenkins, *Time Series Analysis. Forecasting and Control* U.S.A. Holden-Day, 1976.
- González Videgaray Ma. Carmen, *Modelos De Decisión con Procesos Estocásticos II* México ENEP Acatlan UNAM 1990.
- William W. S Wei, *TIME SERIES ANALYSIS*, Addison-Wesley Publishing Company, inc 1989
- Gujarati, Domodar N., *Econometria* México, Mc-Graw Hill, 1998
- Pérez López, César, *Econometría y análisis estadístico multivariable con Statgraphics*, España, RA-MA, 1996.
- Sistema de Transporte Colectivo Metro, *Plan de Empresa 1992-1996* México, STC Metro 1992.
- Sistema de Transporte Colectivo Metro, *Plan de Empresa 2000. -2006* México, STC Metro 2000.
- Martín P.F Javier, *Introducción a la estadística económica y empresarial. Teoría y Practica* España, A Editorial, 1995.
- <http://www.metro.df.gob.mx/> 2002-05-26
- <http://www.df.gob.mx/ciudad/reaportajes/metro> 2002-08-06

ANEXO I

Datos originales para la Línea 1

	1995	1996	1997	1998	1999	2000	2001	2002
enero	30	22	27	23	27	24	12	23
febrero	20	25	26	22	19	13	27	13
marzo	19	17	31	15	21	25	15	15
abril	25	25	24	30	25	16	22	20
mayo	40	51	47	50	49	48	44	52
junio	20	13	31	20	35	22	20	15
julio	21	37	24	16	28	10	28	
agosto	35	23	19	22	19	19	22	
septiembre	36	18	25	29	27	15	18	
octubre	29	27	32	26	23	30	23	
noviembre	37	15	29	19	22	13	12	
diciembre	18	27	15	26	22	22	24	

Pronósticos hacia atrás

	1995	1996	1997	1998	1999	2000	2001	2002
enero	25.1357	28.0498	23.3533	24.3044	20.8844	21.8354	26.0194	
febrero	23.4403	19.2055	18.3522	20.0112	20.232	21.1937	12.186	
marzo	23.0368	21.7547	13.3555	22.0257	16.7197	14.2878	19.1103	
abril	26.3185	24.6436	23.7019	19.6917	22.2582	25.2336	19.537	
mayo	56.3624	47.2201	48.1948	45.2311	45.0701	44.7606	45.6396	
junio	24.5027	25.3226	13.8368	20.2375	12.2006	20.4722	20.1746	
julio	31.1207	20.8983	27.9849	29.3333	20.2901	28.0764		
agosto	20.8731	27.1735	26.741	22.6191	22.3786	20.3392		
septiembre	20.3107	29.5088	22.9728	20.1423	21.1415	25.979		
octubre	28.6524	28.146	24.3172	26.8561	26.5737	21.4415		
noviembre	14.8108	26.9064	17.2418	22.2493	18.9221	21.957		
diciembre	28.1082	20.9462	25.8182	17.6568	19.4388	18.4503		

MISIS CON
 FALTA DE ORIGEN

ANEXO I

Residuales

	1995	1996	1997	1998	1999	2000	2001	2002
enero	-3.13574	-1.04978	-0.353346	2.69558	3.11556	-9.83544	-3.01937	
febrero	1.55972	6.79447	3.64782	-1.01116	-7.23204	5.80628	0.813994	
marzo	-6.03683	9.24527	1.64454	-1.02573	8.28026	0.712162	-4.11029	
abril	-1.31851	-0.643578	6.29813	5.30834	-6.2562	-3.23364	0.462988	
mayo	-5.36245	-0.220132	1.80523	3.76892	2.92994	-0.7606	6.36038	
junio	-11.5027	5.67742	6.16317	14.7625	9.79943	-0.472216	-5.17464	
julio	5.87929	3.1017	-11.9849	-1.33333	-10.2901	-0.0764141		
agosto	2.1269	-8.17348	-4.74097	-3.61909	-3.37856	1.66085		
septiembre	-2.31072	-4.50882	6.02722	6.85775	-6.14153	-7.97901		
octubre	-1.6524	3.85398	1.68281	-3.85613	3.42634	1.55846		
noviembre	0.189239	2.09359	1.75817	-0.249343	-5.92215	-9.95701		
diciembre	-1.10817	-5.94616	0.181795	4.34319	2.56118	5.5497		

Pronóstico 2002

	Limite inferior	Probable	Limite Superior
julio	4.35433	15.5727	26.791
agosto	6.13159	17.3499	28.5682
septiembre	10.0659	21.2842	32.5026
octubre	13.354	24.5723	35.7906
noviembre	8.14919	19.3675	30.5858
diciembre	5.54458	16.7629	27.9812

TESTE COM
 FALHA DE ORIGEM

Anexo I

Datos de la Función Autocorrelación
Para la Línea 1

Retrasos	ACF_ninguna	ACF_1era_difo	ACF_2a_dif
1	-0.0607328	-0.477954	-0.61617
2	-0.0611433	-0.0285255	0.129119
3	-0.0413043	-0.00571686	0.0215157
4	-0.0199437	-0.0461921	-0.100809
5	0.0856311	0.154572	0.19031
6	-0.108309	-0.157517	-0.16839
7	0.0300154	0.0500488	0.0472465
8	0.0380262	0.0624651	0.055142
9	-0.0725254	-0.0518891	-0.0424163
10	-0.0758096	-0.0561386	0.0199275
11	0.0385101	-0.138005	-0.210809
12	0.463681	0.429711	0.393057
13	-0.0286571	-0.202419	-0.246137
14	-0.0661056	-0.0317364	0.0497712
15	-0.05923	-0.0279294	-0.0263292
16	-0.00796354	0.0186862	0.0120953
17	0.00139771	0.0537936	0.0755798
18	-0.0964585	-0.14748	-0.166784
19	0.112584	0.149598	0.144261
20	0.00425619	-0.0202132	-0.0418743
21	-0.0365251	0.0102656	0.0248705
22	-0.124172	-0.0857768	0.0206923
23	-0.0182447	-0.197746	-0.289368
24	0.484232	0.465142	0.441107
25	0.0224401	-0.169182	-0.229189
26	-0.0492424	-0.0122705	0.0478054
27	-0.139292	-0.0918519	-0.0488044
28	-0.0200379	0.0332556	0.0152245
29	-0.00200008	0.039628	0.0404067

desviaciones
estandar

0.144

0.156

0.167

TESIS CON
FALLA DE ORIGEN

Anexo I: Pasajeros Afectados por los fallos en PA para la Línea 1

	USUARIOS estimados	Vueltas estimadas	Fallos	Vuletas no dadas por tren	Usuaños por vuelta	USUARIOS por un tren	USUARIOS afectados
Julio	23,443,726	14,625	12	1,219	1,603	1,530	1,954,057
Agosto	23,357,393	14,594	17	858	1,600	1,530	1,372,800
Septiembre	22,038,482	13,766	21	656	1,601	1,530	1,050,256
Octubre	23,367,756	14,625	24	609	1,598	1,530	973,182
Noviembre	22,219,102	13,915	19	732	1,597	1,530	1,169,004
Diciembre	22,773,404	14,265	16	892	1,596	1,530	1,423,632

TESIS CON
 FALLA DE ORIGEN

Datos originales para la Línea 5

	1995	1996	1997	1998	1999	2000	2001	2002
enero	31	32	31	33	30	31	28	30
febrero	15	23	25	21	18	20	20	18
marzo	19	20	12	15	13	15	11	14
abril	13	15	15	11	11	12	16	10
mayo	21	22	18	18	22	17	12	15
junio	17	14	20	13	10	14	19	17
julio	20	20	14	19	15	15	13	
agosto	16	17	17	14	20	11	15	
septiembre	20	20	14	16	12	17	12	
octubre	23	15	16	13	17	14	14	
noviembre	20	22	23	18	15	13	19	
diciembre	26	18	18	17	13	19	10	

Pronósticos hacia atrás

	1995	1996	1997	1998	1999	2000	2001	2002
enero	29.3696	32.0316	30.8694	30.5548	31.1196	27.5893	30.3634	
febrero	20.8312	20.5218	19.2815	19.7415	18.0669	18.5562	16.9013	
marzo	16.4413	15.0754	15.2959	15.4807	13.2896	12.5021	12.1066	
abril	13.0024	15.5779	13.2313	12.7362	10.6468	11.4724	9.82334	
mayo	19.5766	18.8865	19.4061	18.1928	17.2762	14.5554	15.6779	
junio	17.0865	17.3055	16.6524	13.4703	14.0523	15.1013	13.3663	
julio	19.4974	16.0521	17.6588	18.0672	14.9651	11.8713		
agosto	16.3647	17.9231	14.6984	15.2437	14.8723	14.153		
septiembre	17.2832	16.6332	16.4986	12.9165	15.5093	13.0299		
octubre	17.1496	18.3276	16.1574	16.3667	13.3547	14.499		
noviembre	22.0837	20.6356	21.1838	18.0658	17.8721	16.1853		
diciembre	18.1516	16.6942	17.9288	18.0271	17.15	13.8887		

**TOP CON
ORIGEN**

ANEXO II

Residuales

	1995	1996	1997	1998	1999	2000	2001	2002
enero		2.6304	-1.0316	2.1306	-0.5548	-0.1196	0.4107	-0.3634
febrero		2.1688	4.4782	1.7185	-1.7415	1.9331	1.4438	1.0987
marzo		3.5587	-3.0754	-0.2959	-2.4807	1.7104	-1.5021	1.8934
abril		1.9976	-0.5779	-2.2313	-1.7362	1.3532	4.5276	0.1767
mayo		2.4234	-0.8865	-1.4061	3.8072	-0.2762	-2.5554	-0.6779
junio		-3.0865	2.6945	-3.6524	-3.4703	-0.0523	3.8987	3.6337
julio		0.5026	-2.0521	1.3412	-3.0672	0.0349	1.1287	
agosto		0.6353	-0.9231	-0.6984	4.7563	-3.8723	0.8470	
septiembre		2.7168	-2.6332	-0.4986	-0.9165	1.4907	-1.0299	
octubre		-2.1496	-2.3276	-3.1574	0.6333	0.6453	-0.4990	
noviembre		-0.0837	2.3644	-3.1838	-3.0658	-4.8721	2.8147	
diciembre		-0.1516	1.3058	-0.9288	-5.0271	1.8500	-3.8887	

Pronóstico 2002

	Limite inferior	Probable	Limite Superior
julio	6.22647	11.8353	17.444
agosto	7.65782	13.7314	19.805
septiembre	6.24474	12.395	18.5453
octubre	7.08206	13.2455	19.4089
noviembre	9.69641	15.8621	22.0278
diciembre	7.47387	13.6399	19.806

1997 CON
 VALLA DE ORIGEN

Anexo II

Datos de la Función Autocorrelación
Para la Línea 5

Retrasos	ACF_ninguna	ACF_1era_dife	ACF_2a_dif
1	0.10948	-0.455163	-0.675085
2	0.0793118	0.0846909	0.276343
3	-0.11825	-0.194921	-0.194951
4	0.0497159	0.131776	0.179677
5	-0.0477947	-0.09244	-0.137458
6	0.037535	0.0839604	0.107971
7	-0.0425734	-0.0582578	-0.0735783
8	0.000140762	0.0309252	0.0476526
9	-0.0323411	-0.0179355	0.0227793
10	-0.0507074	-0.150755	-0.0490471
11	0.204828	-0.114029	-0.184831
12	0.623621	0.454183	0.360399
13	0.234896	-0.0659635	-0.182419
14	0.00217137	-0.0392516	0.0551637
15	-0.14729	-0.181928	-0.156072
16	0.0384451	0.170553	0.215531
17	-0.101233	-0.145637	-0.21034
18	0.0521989	0.179326	0.240155
19	-0.13482	-0.211786	-0.279132
20	0.0539798	0.217946	0.275418
21	-0.145993	-0.165599	-0.166244
22	-0.0361417	-0.0368511	0.0771634
23	0.0985817	-0.162646	-0.232219
24	0.532945	0.432216	0.3634
25	0.159407	-0.075346	-0.191495
26	-0.0544125	-0.0665448	0.0175503
27	-0.0897542	-0.047034	-0.0141174
28	-0.0592262	0.0138234	0.0420382
29	-0.0663624	-0.0455353	-0.098088

desviaciones
estandar

0.180

0.188

0.226

TESIS CON
FALLA DE ORIGEN

Anexo II: Pasajeros Afectados por los fallos en PA para la Linea 5

	USUARIOS estimado	Vueltas estimadas	Fallos	Vuletas no dadas por tren	Usuarios por vuelta	USUARIOS por un tren	USUARIOS afectados
Julio	6,831,994	6,422	12	535	1,064	1,530	569,240
Agosto	6,800,439	6,400	14	457	1,063	1,530	485,791
Septiembre	6,430,393	6,146	12	512	1,046	1,530	535,552
Octubre	6,867,636	6,422	13	494	1,069	1,530	528,086
Noviembre	6,503,059	6,155	16	385	1,057	1,530	406,945
Diciembre	6,705,261	6,360	14	454	1,054	1,530	478,516

DE ORIGEN