

---

**UNIVERSIDAD DE EL SALVADOR  
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA  
ESCUELA DE MATEMÁTICA**



**Universidad de El Salvador**  
*Hacia la libertad por la cultura*

**“ANÁLISIS DE SERIES TEMPORALES  
CON OUTLIERS E INTERVENCIONES  
Y SUS APLICACIONES”**

**TRABAJO DE GRADUACIÓN PRESENTADO POR:  
JAIMES BERRIOS, ALMA AZUCENA.  
QUINTANILLA APARICIO, MARIA ISABEL**

**PARA OPTAR AL GRADO DE:  
LICENCIADA EN ESTADÍSTICA.**

**MARZO DE 2008.**

**SAN SALVADOR**

**EL SALVADOR**

**CENTRO AMERICA**

---

**UNIVERSIDAD DE EL SALVADOR  
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA  
ESCUELA DE MATEMÁTICA**

**TRABAJO DE GRADUACIÓN**

**“ANÁLISIS DE SERIES TEMPORALES  
CON OUTLIERS E INTERVENCIONES  
Y SUS APLICACIONES”**

**PRESENTADO POR:**

**JAIMES BERRIOS, ALMA AZUCENA.  
QUINTANILLA APARICIO, MARÍA ISABEL.**

---

**ASESOR:  
Msc. ROLANDO LEMUS**

**CIUDAD UNIVERSITARIA, SAN SALVADOR, MARZO DE 2008.**

## **UNIVERSIDAD DE EL SALVADOR**

**RECTOR:** **ING. RUFINO QUEZADA**

**SECRETARIO GENERAL:** **LIC. DOUGLAS VLADIMIR ALFARO CHAVEZ**

### **FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**

**DECANO:** **DR. RAFAEL ANTONIO GOMEZ ESCOTO**

**SECRETARIO:** **LIC. MARIA TRINIDAD TRIGUERO DE CASTRO**

### **ESCUELA DE MATEMÁTICA**

**DIRECTOR (INTERINO):** **M.SD. MARTA NOEMI MARTINEZ**

**SECRETARIO:** **LIC. ALBA IDALIA CÓRDOVA CUELLAR**

**CIUDAD UNIVERSITARIA, MARZO DE 2008.**



## **Agradecimientos**

A Dios todopoderoso, por habernos guiado por el camino correcto y darnos las fuerzas y valor para seguir adelante en los momentos difíciles de la carrera.

A nuestras familias por brindarnos todos su apoyo, estar siempre pendientes de nosotras a pesar de la distancia, por comprender que nos alejamos de ellas por terminar nuestras carreras, y por hacernos felices el poco tiempo que pasamos junto con ellos.

A nuestro asesor Msc. Rolando Lemus, por ayudarnos a dar el paso final de la carrera, gracias a su experiencia, conocimientos, ayuda y comprensión logramos terminar nuestra Tesis.

A Todos los docentes que nos transmitieron un poco de sus conocimientos, para llegar al final.

A todos nuestros amigos por confiar en nosotras, apoyarnos, y brindarnos ayuda cuando la solicitábamos.

**Azucena Jaimes e Isabel Quintanilla**

## **Agradecimientos**

A Dios todo Poderoso, por haberme dado la inteligencia necesaria y fuerzas para saber continuar después de un obstáculo. De igual forma por regalarme una familia súper especial en la que siempre encontraba abrigo y apoyo.

A mis papas: José Misael Quintanilla y Ana Vicente de Quintanilla, por su amor y apoyo incondicional, por esforzarse para que nunca me faltara nada.

A mis hermanos: Misael, Yesenia y Marisela, por ayudarme a distraerme en los momentos mas llenos de presión en mis estudios y por ser un ejemplo a ser una buena persona.

A Azucena por saber entenderme y ser una excelente compañera de tesis y muy buena amiga, es una hermana más, con la que puedo contar para todo.

A mi sobrinita Katherinne por darme su amor limpio y sin condición alguna.

A mis tíos, tías, abuela, por sus sabios consejos.

A mí cuñada Merlín por sus atenciones y cariño.

A Otoniel por compartir sus conocimientos con mí y demostrarme su apoyo en los momentos difíciles.

A mis primos, primas y amigos, por confiar en mí y compartir momentos de tristeza y alegría.

MARIA ISABEL QUINTANILLA APARICIO

## **Agradecimientos**

A Dios todopoderoso, por haberme guiado por el camino correcto y darme fuerzas y valor para seguir adelante en los momentos de prueba, gracias Dios por darme la oportunidad de cumplir una de las más grandes metas de mi vida. Además por poner a mi lado a las personas indicadas que me ayudaron a terminar mi carrera:

Mi madre Dora Luz Berrios, por sus consejos, apoyo moral, emocional y esforzarse día a día para poder brindarme su ayuda económica, gracias a ella supere obstáculos muy grandes que me impedían continuar y llegar al final de la meta. Con su ayuda he logrado terminar mi carrera en el menor tiempo posible, siempre ha sido mi ejemplo a seguir. Mi abuelita Maria Cristina Vda. de Berrios y mi tía Martha Berrios de Chicas que muchas veces hicieron el papel de mamá para mí, las quiero mucho a las tres como mis madrecitas. Muchas gracias por todo su amor y sabios consejos.

Maria Isabel, mi compañera y amiga que se ha convertido en mi hermana, por apoyarme cuando más lo necesitaba. Con su ayuda y conocimiento he logrado terminar este trabajo de investigación. Gracias a Dios que puso a esta excelente persona en mi camino para ser mi compañera de tesis.

Mis hermanos Jesús y Jeovanni por estar a mi lado en momentos de felicidad y tristeza, ayudarme a distraerme en los momentos de estrés y angustias.

Mis tías y tíos maternos: Gloria, Nelly, Miriam, Lorena, Cecilia, David y Armando. Por su ayuda económica y moral incondicional. Mis primos: Sergio, William, Esmeralda, Cristhina, Lisseth, Alex y Rene que junto con sus familias me ayudado durante toda mi carrera.

La familia Vásquez-Orellana por apoyarme durante dos años.

Todos mis amigos y compañeros de trabajo en especial Aída Gómez y Rolando Carrillo por su gran amistad, comprensión y cariño.

Todos los amigos sinceros y compañeros de universidad que conocí en San Miguel y San Salvador, que siempre estuvieron pendientes de mí, para apoyarme hasta el final.

ALMA AZUCENA JAIMES BERRIOS

## ÍNDICE

<b>Título</b>	<b>Pág.</b>
1. Introducción.....	12
2. Planteamiento del Problema.....	13
3. Antecedentes y Justificación.....	14
3.1 Antecedentes.....	14
3.2 Justificación.....	15
4. Fundamento Teórico.....	16
4.1. Análisis de Intervención.....	18
4.2. Modelos de Intervención .....	19
4.3. Identificación de los Modelos ARIMA con Intervención.....	24
4.4. Análisis de Outliers en Series Temporales.....	25
4.4.1. Tipos de Outliers.....	25
4.4.2. Efectos de Outliers.....	29
4.5. Detección y Tratamiento de Outliers.....	31
4.5.1. Procedimiento iterativo de detección de Outliers .....	31
4.5.2. Procedimiento de detección de outliers y estimación conjunta de los parámetros y efectos de los Outliers .....	36
5. Objetivos.....	39
6. Metodología.....	40
7. Aplicaciones de la Investigación.....	47
7.1. Aplicaciones de ST con Intervención.....	47
7.1.1. Serie Temporal del IVAE en comida Restaurante y Hoteles.....	47
7.1.2. Serie Temporal del IPC en el Consumo Individual de Salud.....	75
7.2. Aplicación de ST con Outliers.....	99
7.2.1. Serie Temporal del IPC en Prendas de Vestir y Calzado.....	99
8. Conclusiones.....	129
9. Recomendaciones.....	130
10. Referencias Bibliográficas.....	131
Anexos .....	132



## ÍNDICE DE TABLAS

Título	Pág.
Tabla 4.2.1 Valores de b, r y s.....	23
Tabla 7.1.1.1 Estadísticos descriptivos.....	49
Tabla 7.1.1.2 Medias y varianzas.....	51
Tabla 7.1.1.3 Estadístico de Levene.....	54
Tabla 7.1.1.4 Verificación de coeficientes para estacionariedad e invertibilidad.....	60
Tabla 7.1.1.5 Test de Normalidad.....	62
Tabla 7.1.1.6 Comparación de los de los estadísticos AIC y BIC.....	63
Tabla 7.1.1.7 Pronósticos con el modelo $IMA(1,2)xIMA(1,1)_{12}$ .....	64
Tabla 7.1.1.8 Test de normalidad.....	67
Tabla 7.1.1.9 Valores de pronósticos.....	68
Tabla 7.1.1.10 Estimación de los parámetros con intervención tipo comienzo gradual y duración permanente.....	69
Tabla 7.1.1.11 Estimación de los parámetros con intervención tipo comienzo brusco y duración permanente.....	70
Tabla 7.1.1.12 Estimación de los parámetros con intervención tipo comienzo brusco y duración temporal.....	72
Tabla 7.1.1.13 Pronósticos con el modelo identificado.....	74
Tabla 7.1.2.1 Estadísticos de la serie.....	77
Tabla 7.1.2.2 Medias y varianzas.....	78
Tabla 7.1.2.3 Estadísticos de Levene para las transformaciones.....	80
Tabla 7.1.2.4 Verificación de coeficientes para la estacionariedad e invertibilidad.....	84
Tabla 7.1.2.5 Test de normalidad.....	86
Tabla 7.1.2.6 Estadísticos AIC y BIC.....	86
Tabla 7.1.2.7 Valores de los pronósticos.....	88
Tabla 7.1.2.8 Estimación de los parámetros.....	90
Tabla 7.1.2.9 Verificación de coeficientes para estacionariedad e invertibilidad.....	94
Tabla 7.1.2.10 Valores de los pronósticos.....	98
Tabla 7.2.1.1 Test de Levene y significancia.....	103
Tabla 7.2.1.2 Modelo 1.....	107
Tabla 7.2.1.3 Modelo 2.....	107
Tabla 7.2.1.4 Modelo 3.....	107
Tabla 7.2.1.5 Verificación de la estacionariedad e invertibilidad.....	108
Tabla 7.2.1.6 Criterio AIC y BIC.....	112
Tabla 7.2.1.7 Pronósticos.....	115
Tabla 7.2.1.8 Detección de outliers y estimación conjunta de los parámetros.....	118
Tabla 7.2.1.9 Detección de outliers y estimación conjunta de los parámetros automáticamente.....	122
Tabla 7.2.1.10 Comparación de pronósticos y datos originales.....	126

## ÍNDICE DE GRÁFICOS

Título	Pág.
Gráfico 4.2.1 Comportamiento de la variable escalón.....	19
Gráfico 4.2.2 Comportamiento de la variable impulso.....	20
Gráfico 4.2.3 Diferentes formas del comportamiento de la variable escalón en una ST continua.....	21
Gráfico 4.2.4 Diferentes formas del comportamiento de la variable impulso en una ST continua .....	22
Gráfico 4.2.5 Comportamiento de la variable escalón e impulso en una misma ST continua .....	22
Gráfico 4.4.1.1 Comportamiento del outliers aditivo (AO) en una ST continua .....	27
Gráfico 4.4.1.2 Comportamiento del outliers innovacional (IO) en una ST continua .....	27
Gráfico 4.4.1.3 Comportamiento del outliers Cambio en Nivel (LS) en una ST continua .....	28
Gráfico 6.1 Esquema de la metodología Box-Jenkins.....	43
Gráfico 6.2 Comportamiento de ST de IVAE en Servicios de Comidas, restaurantes y Hoteles.....	45
Gráfico 6.3 Comportamiento de ST de IPC en la clasificación del consumo individual de Salud.....	46
Gráfico 6.3 Comportamiento de ST de IPC en la clasificación de prendas de vestir y calzado.....	46
Gráfico 7.1.1.1 Serie original.....	47
Gráfico 7.1.1.2 Función de Autocorrelación Simple.....	48
Gráfico 7.1.1.3 Función de Autocorrelación Simple Estacional.....	48
Gráfico 7.1.1.4 Función de Autocorrelación Parcial.....	49
Gráfico 7.1.1.5 Histograma de los Datos Originales.....	50
Gráfico 7.1.1.6 Serie menos el 5%.....	52
Gráfico 7.1.1.7 Serie del subconjunto de datos antes de la intervención.....	53
Gráfico 7.1.1.8 Serie del subconjunto de datos después de la intervención.....	54
Gráfico 7.1.1.9 Rango media.....	55
Gráfico 7.1.1.10 Función de Autocorrelación Simple parte Regular.....	55
Gráfico 7.1.1.11 Función de Autocorrelación Simple parte Estacional.....	56
Gráfico 7.1.1.12 serie diferenciada.....	56
Gráfico 7.1.1.13 FAS Regular Diferenciada .....	57
Gráfico 7.1.1.14 FAP regular diferenciada.....	57
Gráfico 7.1.1.15 FAS estacional diferenciada .....	58
Gráfico 7.1.1.16 FAP estacional diferenciada .....	58
Gráfico 7.1.1.17 Residuos con el modelo $IMA(1,2) \times IMA(1,1)_{12}$ .....	61
Gráfico 7.1.1.18 FAPR con el modelo $IMA(1,2) \times IMA(1,1)_{12}$ .....	61
Gráfico 7.1.1.19 Histograma residual.....	62
Gráfico 7.1.1.20 Probabilidad residual.....	62
Gráfico 7.1.1.21 Superpuesto de datos reales y datos generados con el modelo $IMA(1,2) \times IMA(1,1)_{12}$ .....	64

Gráfico 7.1.1.22 Residuos de la serie de enero 1990-marzo 2006.....	65
Gráfico 7.1.1.23 FASR de la serie de enero 1990-marzo 2006.....	66
Gráfico 7.1.1.24 Histograma residual.....	66
Gráfico 7.1.1.25 Probabilidad de los residuos.....	67
Gráfico 7.1.1.26 Superpuesto de datos reales y datos generados con el modelo identificado.....	68
Gráfico 7.1.1.27 Residuos con modelo de intervención.....	70
Gráfico 7.1.1.28 FASR.....	71
Gráfico 7.1.1.29 Histograma residual.....	71
Gráfico 7.1.1.30 Probabilidad normal de los residuos.....	72
Gráfico 7.1.1.31 Superpuesto de datos pronosticados y datos originales.....	73
Gráfico 7.1.2.1 ST del IPC en el consumo individual de salud.....	75
Gráfico 7.1.2.2 Función de Autocorrelación Simple.....	76
Gráfico 7.1.2.3 Función de Autocorrelación Simple estacional.....	76
Gráfico 7.1.2.4 Función de Autocorrelación parcial.....	77
Gráfico 7.1.2.5 Histograma de los dato originales.....	78
Gráfico 7.1.2.6 Serie menos el 20% de los datos.....	79
Gráfico 7.1.2.7 Rango-media.....	80
Gráfico 7.1.2.8 Serie Diferenciada Estacional y Regularmente.....	81
Gráfico 7.1.2.9 Función de Autocorrelación Simple Regular.....	81
Gráfico 7.1.2.10 Función de Autocorrelación Parcial Regular.....	82
Gráfico 7.1.2.11 Función de Autocorrelación Simple Estacional.....	82
Gráfico 7.1.2.12 Residuos con el modelo $ARI(1,1) \times ARI(1,1)_{12}$ .....	84
Gráfico 7.1.2.13 FASR del modelo $ARI(1,1) \times ARI(1,1)_{12}$ .....	85
Gráfico 7.1.2.14 Histograma Residual.....	85
Gráfico 7.1.2.15 Probabilidad normal de los residuos.....	86
Gráfico 7.1.2.16 Superpuesto de los datos reales y pronosticados.....	87
Gráfico 7.1.2.17 residuos del modelo.....	90
Gráfico 7.1.2.18 FASR.....	91
Gráfico 7.1.2.19 FAPR.....	91
Gráfico 7.1.2.20 Histograma Residual.....	92
Gráfico 7.1.2.21 Probabilidad Normal de los Residuos.....	92
Gráfico 7.1.2.22 Residuos con el modelo $IMA(1,1) \times I(1)_{12}$ .....	94
Gráfico 7.1.2.23 FASR.....	95
Gráfico 7.1.2.24 Histograma Residual.....	95
Gráfico 7.1.2.25 Probabilidad Normal de los residuos.....	96
Gráfico 7.1.2.26 Superpuesto de los datos Pronosticados y datos originales.....	97
Gráfico 7.2.1.1 Serie Prendas de Vestir y calzado.....	99
Gráfico 7.2.1.2 Serie prendas de vestir y calzado menos el 20%.....	100
Gráfico 7.2.1.3 Autocorrelaciones Simples Regulares.....	101
Gráfico 7.2.1.4 Autocorrelaciones Parciales Regulares.....	101
Gráfico 7.2.1.5 Autocorrelaciones Simples Estacionales.....	102
Gráfico 7.2.1.6 Autocorrelaciones Parciales Estacionales.....	102
Gráfico 7.2.1.7 Serie con dos diferenciaciones regulares.....	104
Gráfico 7.2.1.8 Autocorrelaciones Regulares con Dos Diferencias Regulare.....	104
Gráfico 7.2.1.9 Autocorrelaciones Estacionales con Dos Diferencias Regulares.....	105

Gráfico 7.2.1.10 Residuos del modelo 1.....	109
Gráfico 7.2.1.11 Residuos del modelo 2.....	109
Gráfico 7.2.1.12 Residuos del modelo 3.....	110
Gráfico 7.2.1.13 FASR del modelo 1.....	110
Gráfico 7.2.1.14 FASR del modelo 2.....	111
Gráfico 7.2.1.15 FASR del modelo 3.....	111
Gráfico 7.2.1.16 serie de los pronósticos y datos originales.....	116
Gráfico 7.2.1.17 residuos del modelos estimado con 11 outliers.....	120
Gráfico 7.2.1.18 FASMR del modelo estimado con 11 outliers.....	121
Gráfico 7.2.1.19 FAPMR del modelo estimado con 11 outliers.....	121
Gráfico 7.2.1.20residuos del modelo estimado con 5 outliers.....	123
Gráfico 7.2.1.21 FASMR del modelo estimado con 5 outliers.....	124
Gráfico 7.2.1.22 FAPMR del modelo estimado con 5 outliers.....	125
Gráfico 7.2.1.23 pronósticos y datos reales.....	127

## 1. INTRODUCCIÓN

Al analizar datos de Series Temporales (ST), muchas veces se observa que en determinados momentos la serie a estudiar presenta movimientos bruscos de importancia, que no es posible captar con una dependencia sistemática de su pasado; ya que dichos movimientos pueden tener su origen en intervenciones incontroladas ó inesperadas; es decir que si en un momento concreto ocurre una observación atípica ó anómala, que no guarda relación con el patrón de comportamiento de la serie, observado hasta esa fecha, tal movimiento no se va a explicar con un modelo ARIMA clásico (regular ó estacionario).

Los modelos ARIMA clásicos, sólo son validos para recoger una regularidad en el tiempo y explorarla con fines explicativos y predictivos. En caso de que se de algo anómalo, sin conexión con la experiencia previa, este tipo de modelos no estarán preparados para captarlo y solucionar este tipo de problema.

Surge así la necesidad de contar con una herramienta estadística para abordar el modelado de ST interrumpidas por ciertos sucesos externos tales como: vacaciones, huelgas, campañas promocionales y cambios de política. Sucesos a los que Box y Tiao, les denominó *Intervenciones y Datos Atípicos*.

En este sentido; este trabajo pretende realizar un estudio exhaustivo y comprensivo del análisis de ST con Datos Atípicos (Outliers) e Intervenciones, y sus Aplicaciones.

Las observaciones atípicas, deben tratarse con el análisis de outliers porque, no es adecuado sustituir los datos atípicos por la media. Por ejemplo, en un estudio de Análisis Multivariante las observaciones atípicas se sustituyen por la media de la serie.

Según Box y Tiao, los datos atípicos deben de analizarse, hasta conocer qué los originó, para luego tratarlos, tomando en cuenta sus efectos.

En caso de que una serie presente movimientos bruscos de importancia; estos movimientos pueden tratarse con uno de los tipos de intervenciones que se clasifican de acuerdo al comienzo y duración del efecto del suceso.

De esta forma, presentamos las ideas esenciales que serán desarrolladas en este trabajo y que nos ayudarán no sólo a explorar un área más de las estadísticas contemporáneas, sino, que permitirán aclarar las ideas que no han podido explicarse por falta de teoría y herramientas que las refuercen.

## 2. PLANTEAMIENTO DEL PROBLEMA

En el pasado los estadísticos, han tratado con modelos ARIMA regulares y estacionarios, con fluctuaciones similares, sin observaciones anómalas.

Para los cuales las técnicas clásicas de Modelización y Series Temporales pueden ser aplicadas. Series Temporales con movimientos bruscos de importancia y efectos de difícil explicación han sido consideradas como una de las curiosidades estadísticas ó como ST, que no son dignas de un estudio sistemático bajo una teoría conocida y aceptada.

De esta manera, el análisis de ST con outliers e intervenciones y sus aplicaciones provee un marco general que permite el estudio sistemático de ST interrumpidas.

Al revisar los gráficos de secuencia de las bases de datos (Las Remesas, y algunos de los Índices del Volumen de la Actividad Económica (IVAE)), que se utilizarán en la aplicación, nos encontramos ante situaciones anómalas que las herramientas de los modelos ARIMA tradicionales son insuficientes para modelarlas.

Por lo tanto la idea central de este trabajo se focaliza en facilitar los fundamentos que permitan modelar y predecir una ST con observaciones extrañas, sin dejar pasar la teoría de los modelos ARIMA que se estudia en el curso de series temporales.

Durante el desarrollo de este trabajo, se dará respuesta a las siguientes interrogantes, que se originan a partir de las observaciones extrañas que se observan en las series.

- ¿Qué es una Intervención?
- ¿Por qué se da una intervención?
- ¿Cuáles son los tipos de intervención?
- ¿Cuál es el efecto de las variables escalón e impulso en un modelo de intervención?
- ¿Qué es un Outliers?
- ¿Cuáles son los tipos y efectos de los outliers?
- ¿Cuáles son las herramientas estadísticas requeridas para poder realizar una aplicación de un Análisis de ST con intervenciones y outliers?
- ¿Cómo pueden ser aplicadas las técnicas estadísticas en el Análisis de ST con intervenciones y outliers?

### **3. ANTECEDENTES Y JUSTIFICACIÓN**

#### **3.1 ANTECEDENTES**

Históricamente, el interés en el análisis de ST ha sido estimulado por sus aplicaciones en las bases de datos de áreas como: Economía, Educación, y Salud.

Encontrar un modelo que se adecúe a los datos, permite que se pueda pronosticar, así mismo que se puedan hacer comparaciones con resultados anteriores.

Una situación similar sucede con el Análisis de Series Temporales con Outliers e Intervenciones y sus Aplicaciones. La reciente literatura muestra la variedad de campos donde se pueden tener series que presenten datos atípicos o intervenciones por algún suceso externo tales como: huelgas, vacaciones, cambios políticos, etc.

Las ST con Outliers e Intervenciones fueron concebidas aproximadamente en 1965, por los estadísticos Box y Tiao. Sus ideas fueron extendidas mas tarde fundamentalmente por Fox, quien en 1972 realizó el primer estudio sobre la detección de ST con outliers, donde dos modelos estadísticos, aditivo e innovacional, fueron introducidos. Otras referencias en este tópico fueron incluidas por Abraham y Box (1979), Martín (1980), Chang y Tiao (1983), Hillmer, Bell, y Tiao (1983), Tsay (1986), Chang, Tiao, y Chen (1988).

El estudio fue renovado a partir de 1975, siempre por Box y Tiao, quienes mostraban mucho interés por este estudio. Y fue fuertemente impulsado porque modelos de series temporales con intervenciones y outliers, pueden considerarse que tienen una variedad de aplicaciones. Por ejemplo en 1975 Box y Tiao, consideraron un estudio en bases de datos de políticas económicas y ambientales.

Otros estadísticos como: Wichern y Jones (1977), consideraron un estudio en bases de datos de dirección y control en la comercialización; y Zimring, (1975), realizó un estudio para los efectos de legislación del control de armas. Quienes mostraron mayor interés en las aplicaciones que se podrían realizar, en lugar de teoría estadística que se podría utilizar.

Desde la década del '70, este campo ha estado en la vanguardia de los estadísticos contemporáneos. Investigadores han explorado esta rama de la estadística con la ayuda de las computadoras y software modernos.

De los precursores en el análisis de ST con outliers e intervenciones y sus aplicaciones se pueden mencionar a los estadísticos Box y Tiao.

### 3.2 JUSTIFICACIÓN

El análisis de ST con outliers e intervenciones, son un concepto nuevo que surgió a principios de la década de los setentas, gracias a los trabajos de Box y Tiao. Sin embargo, las herramientas estadísticas usadas para su estudio no son nuevas. Los conceptos básicos que encontramos en ésta, como los modelos ARIMA, la teoría de Modelos Estacionarios, las funciones de Autocorrelación FAS y FAP, etc., existen desde mucho tiempo atrás.

A pesar de que las herramientas estadísticas utilizadas en el análisis de ST con outliers e intervenciones no son nuevas, nos parece, sin embargo una estadística complicada sobre todo por la interpretación de los parámetros de la parte de la intervención y los datos atípicos (outliers); ya que con ellos se obtienen modelos que no son fáciles de interpretar en una primera impresión.

Además se usan iteraciones de modelos, que en teoría parecen sencillos, al hacer composiciones de los mismos se generan unas estructuras estadísticas bastante complicadas.

La teoría de este campo llama mucho la atención, para ponerla en práctica en las aplicaciones que se realicen.

Por otra parte, el Análisis de ST con Outliers e Intervenciones y sus Aplicaciones es una rama de las estadísticas que todavía no ha sido desarrollada en la Escuela de Matemática de la Universidad de El Salvador, ya sea como asignatura ó como trabajo de investigación.

Por lo tanto surge la necesidad de realizar una investigación de carácter básico en torno a estos tópicos nuevos de la Estadística, que permita además sentar las bases para poder realizar futuras investigaciones más exhaustivas y profundas sobre este campo.

Este estudio podría coadyuvar al desarrollo académico de los estudiantes de la Facultad de Ciencias Naturales y Matemática de la Universidad de El Salvador, como también a los estudiantes de las Facultades de Ciencias Económicas y Medicina; los cuales en un momento pueden recurrir a aplicar el análisis de ST con outliers e intervenciones. Además la difusión de este trabajo podría ayudar a los esfuerzos de actualización y desarrollo de los docentes del área de Modelos Lineales y Modelos Lineales Generalizados como también reforzará a los docentes del área de Análisis Multivariante, y especialmente en la materia de Series Temporales.



#### 4. FUNDAMENTO TEÓRICO DEL ANÁLISIS DE SERIES TEMPORALES CON OUTLIERS E INTERVENCIONES Y SUS APLICACIONES

El estudio de Series Temporales con Intervenciones en principio toma referencia de los modelos ARIMA que se han estudiado, en cursos preliminares de Series temporales. También es importante hacer reseña a los Modelos de Regresión Dinámica y la Función de Transferencia; partiendo de un conocimiento apriorístico de la dirección de estos modelos.

El Modelo de Regresión Dinámica es similar al Análisis de Regresión; dado que produce resultados más realistas puesto que acentúa los efectos de ondulación que las variables de entrada pueden tener en la variable de salida. Este tipo de modelo puede predecir qué sucederá, si la variable explicativa cambia, por ejemplo, un cambio del precio hecho hoy, puede influir los volúmenes de venta en una variedad de maneras para muchos períodos futuros.

La Regresión Dinámica es un modelo de regresión que incluye valores retrasados de variables explicativas o de variables relacionadas o de ambas. La relación entre el pronóstico variable y la variable explicativa se modela usando un Proceso de Transferencia. Tomando en cuenta que para modelar la Regresión Dinámica se tiene que tener conocimientos de un modelo de Función de Transferencia.

Un modelo de Función de Transferencia consiste en tomar dos series esto para el caso más simple, porque se puede disponer de más de dos. Consideraremos una de esas series ( $Y_t$ ) como la variable de salida (output) y la otra como la variable de entrada ( $X_t$ ). Una vez aceptado este hecho, el nivel presente de  $Y_t$  puede escribirse como función de los valores presentes y pasados de  $X_t$ :<sup>1</sup>

$$Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + N_t$$

O bien:

$$Y_t = v(L)X_t + N_t$$

Donde:

$$v(L) = v_0 + v_1 L + v_2 L^2 + \dots$$

Siendo L el operador de retardos<sup>2</sup>.

Y  $N_t$  es una variable aleatoria independiente de  $X_t, X_{t-1}, X_{t-2}, \dots$ , que no tiene por qué ser necesariamente un ruido blanco.

Los coeficientes  $v_0, v_1, v_2, \dots$ , constituyen la denominada función impulso respuesta ( $v(L)$ ), para mantener la estabilidad de la variable input y la variable output, debe cumplirse que una variación finita en la variable ( $X_t$ ) produzca una variación también finita en la variable ( $Y_t$ ). Para introducir este hecho, indicaremos por b el número de períodos que transcurren antes de que  $X_t$  comience a tener efecto sobre la variable dependiente, debiendo escribir:

---

<sup>1</sup> Denotaremos a las series originales con letras mayúsculas, dejando la notación habitual de letras minúsculas a las series estacionarias.

<sup>2</sup> Véase métodos de predicción en economía (II) Pág. 34

$$Y_t = \frac{\omega(L)}{\delta(L)} X_{t-b} + N_t = \frac{\omega(L)L^b}{\delta(L)} X_t + N_t$$

Donde  $v(L)$  se aproxima mediante el cociente de dos polinomios (rationales) finitos, tal como se expresa a continuación:

$$v(L) = \frac{\omega(L)}{\delta(L)}$$

Donde:

$$\begin{aligned}\omega(L) &= \omega_0 - \omega_1 L - \omega_2 L^2 - \dots - \omega_s L^s \\ \delta(L) &= 1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r\end{aligned}$$

Debiendo caer todas las raíces de  $\omega(L) = 0$  y  $\delta(L) = 0$ , fuera del círculo unitario para cumplir la condición de estacionariedad.

Con lo anterior se ha tratado de construir un modelo de función de transferencia para lo que se requiere que tanto la variable dependiente como la explicativa sean estacionarias, obviamente no deberá cumplirse que el orden de diferenciación para ambas variables sea el mismo. Por otra parte, una vez lograda la estacionariedad en ambas variables, el proceso de  $N_t$  deberá ser un ARMA (p, q) donde p y q indican el orden de los parámetros de los polinomios autoregresivos y de medias móviles (Modelos estudiados en el curso de series temporales<sup>3</sup>), con lo cual el modelo de función de transferencia lo escribiremos como:

$$y_t = \frac{\omega(L)L^b}{\delta(L)} x_t + \frac{\theta(L)}{\phi(L)} u_t$$

Donde:

$$\begin{aligned}y_t &= (1-L)^{d'} Y_t \\ x_t &= (1-L)^d X_t\end{aligned}$$

Siendo  $d'$  el orden de diferenciación (transformación para alcanzar el suavizado de la serie en estudio en caso de no cumplirse la estacionariedad en media) de la serie  $Y_t$  y  $d$ , el de la serie  $X_t$ . Además  $u_t$  es una secuencia de perturbaciones aleatorias distribuidas idéntica e independientemente con media cero y varianza  $\sigma_u^2$  lo que se conoce como ruido blanco<sup>4</sup>.

Para relacionar este modelo de función de transferencia con las intervenciones, suponemos que tenemos una sola intervención que denominaremos  $I_t^{t_0}$ , siendo  $I_t^{t_0}$  una variable impulso ( $P_t^{t_0}$ ) o una escalón ( $S_t^{t_0}$ ).

De aquí el modelo de función de transferencia con una intervención puede escribirse como:

<sup>3</sup> Véase métodos de predicción en economía (II). Pag. 118-147, o trabajo de graduación titulado “Análisis univariante de series temporales”

<sup>4</sup> Véase métodos de predicción en economía (II). Pág. 31, o trabajo de gradación titulado “Análisis univariante de series temporales”

$$Y_t = v(L)I_t^{t_0} + \frac{\omega(L)}{\delta(L)(1-L)^{d'}} x_t + \frac{\theta(L)}{\phi(L)(1-L)^{d'}} u_t$$

Donde:

$$x_t = (1-L)^d X_t$$

$$v(L) = \frac{\omega(L)L^b}{\delta(L)}$$

Y

$$\omega(L) = \omega_0 - \omega_1 L - \omega_2 L^2 - \dots - \omega_s L^s$$

$$\delta(L) = 1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r$$

El modelo escrito anteriormente con una intervención puede también generalizarse al caso de k intervenciones, pudiéndose expresar entonces como:

$$Y_t = \sum_{j=1}^k v_j(L)I_{jt}^{t_0j} + \frac{\omega(L)}{\delta(L)(1-L)^{d'}} x_t + \frac{\theta(L)}{\phi(L)(1-L)^{d'}} u_t$$

Donde  $j$  indica el número de intervenciones que pueden ocurrir en la serie en estudio. Por lo tanto la notación  $jt$ , no indica intervenciones periódicas.

La identificación del modelo de función de transferencia con intervenciones puede efectuarse, de forma análoga a lo que ocurre con el modelo ARIMA con intervenciones, mediante dos métodos alternativos que posteriormente se especifican como un caso particular de variables binarias; es decir una variable de entrada y otra de salida.

Ese modelo particular que se puede dar de la Regresión Dinámica y de la Función de Transferencia, es el Modelo de Intervención.

#### 4.1 ANÁLISIS DE INTERVENCIÓN

Cuando se analizan los datos de series temporales, es corriente encontrar atípicos que tienen su origen en intervenciones incontroladas o inesperadas, Las series temporales se ven influenciadas por ciertos sucesos externos, tales como: vacaciones, huelgas, campañas promocionales y cambios de política. A estos sucesos externos Box y Tiao les denominaron: *intervenciones*.

El análisis de intervención tiene su técnica la cual consiste en evaluar el efecto de las intervenciones en el proceso de comportamiento de una serie temporal.

Para aplicar esta técnica se debe ser capaz de identificar dos características de los modelos de intervención, estas características son:

- a) El Período de comienzo de dichos sucesos externos (Intervenciones).
- b) La forma general del impacto de dichas intervenciones.

Para analizar detenidamente las intervenciones, es necesario detallar los modelos en los que se puntualizan las variables de intervención, que se pueden dar en una ST.

## 4.2 MODELOS DE INTERVENCIÓN

Los modelos de intervención son utilizados cuando algún suceso o evento ocurre y hace que los valores de la serie temporal en estudio aumenten (o disminuyan), a partir de la ocurrencia del suceso. La disminución de la serie temporal puede ser de forma continua o un cambio brusco, únicamente en el instante de ocurrencia del suceso.

Son dos tipos de variables de intervención los que, determinan si el cambio de la serie es continuo ó temporal.

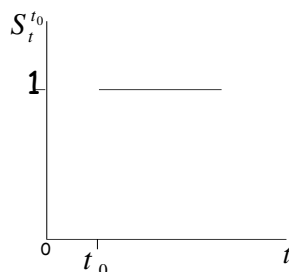
Una de estas variables de intervención es la **variable escalón**, tiene efectos a partir del tiempo  $t_0$  en el que ocurre, y se emplea solamente cuando los sucesos externos tienen una duración permanente.

La variable escalón se puede representar de la siguiente forma:

$$S_t^{(t_0)} = \begin{cases} 0, & t < t_0 \\ 1, & t \geq t_0 \end{cases} \quad (4.2.1)$$

El comportamiento de este tipo de variable se puede observar en el gráfico siguiente:

**Gráfico 4.2.1.** Comportamiento de la variable escalón.



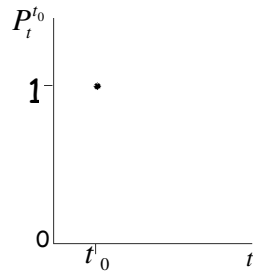
El otro tipo de las variables de intervención corresponde a la **variable impulso**, tiene efectos solo en el tiempo  $t_0$  de ocurrencia y se emplea cuando los sucesos externos tienen una duración temporal.

Este tipo de variable se representa de la siguiente forma:

$$P_t^{(t_0)} = \begin{cases} 1, & t = t_0 \\ 0, & t \neq t_0 \end{cases} \quad (4.2.2)$$

En el gráfico que se muestra a continuación, se observa el comportamiento de la variable impulso.

**Gráfico 4.2.2.** Comportamiento de la variable impulso.



Nótese, que las variables escalón e impulso se relacionan mediante la siguiente expresión:

$$P_t^{t_0} = (1 - L)S_t^{t_0}$$

Para demostrar la relación se esclarece que 1, es la función identidad y  $L$  es el operador de retardos ( $LS_t^{t_0} = S_{t-1}^{t_0}$ ), esto lleva a:

$$\begin{aligned} P^{t_0}(t) &= (I - L)S^{t_0}(t) \\ &= IS^{t_0}(t) - LS^{t_0}(t) \\ &= S^{t_0}(t) - S^{t_0}(t-1) \end{aligned}$$

Se ha definido que si  $t < t_0$ , se cumple que  $S^{t_0}(t) = 0$  y  $S^{t_0}(t-1) = 0$  de donde  $P^{t_0}(t) = 0$ . De igual forma, si  $t \geq t_0$  y  $(t-t_0) < 1$ , se tiene que  $S^{t_0}(t) = 1$  y  $S^{t_0}(t-1) = 0$ , de donde  $P^{t_0}(t) = 1$ . Y si  $(t-t_0) \geq 1$ , entonces  $S^{t_0}(t) = 1$ ,  $S^{t_0}(t-1) = 1$  y  $P^{t_0}(t) = 0$ . Con esto queda demostrado que las variables (Impulso y Escalón) se relacionan.

Por consiguiente un modelo de intervención puede representarse igualmente, bien con la variable escalón o con la variable impulso. El uso de una forma específica de estas variables es usualmente basado en la conveniencia de la interpretación en el modelo.

Algunos modelos pueden estructurarse, tomando en cuenta algunos criterios como los siguientes:

1. El modelo más simple sería aquel en que el impacto de una intervención es fijo y de magnitud desconocida. Así, dependiendo en el tipo de intervención, y tomando en cuenta que el modelo se puede generalizar transcurridos  $b$  períodos, el impacto es:

$$X_t = \omega L^b S_t^{(t_0)} \quad (4.2.3)$$

O

$$X_t = \omega L^b P_t^{(t_0)} \quad (4.2.4)$$

2. Si se tiene un impacto de una intervención con comienzo gradual, para los dos tipos de duración del efecto, con  $b$  períodos transcurridos después de la intervención se tiene para la variable escalón el modelo siguiente:

$$X_t = \frac{\omega L^b}{1 - \delta L} S_t^{(t_0)} \quad (4.2.5)$$

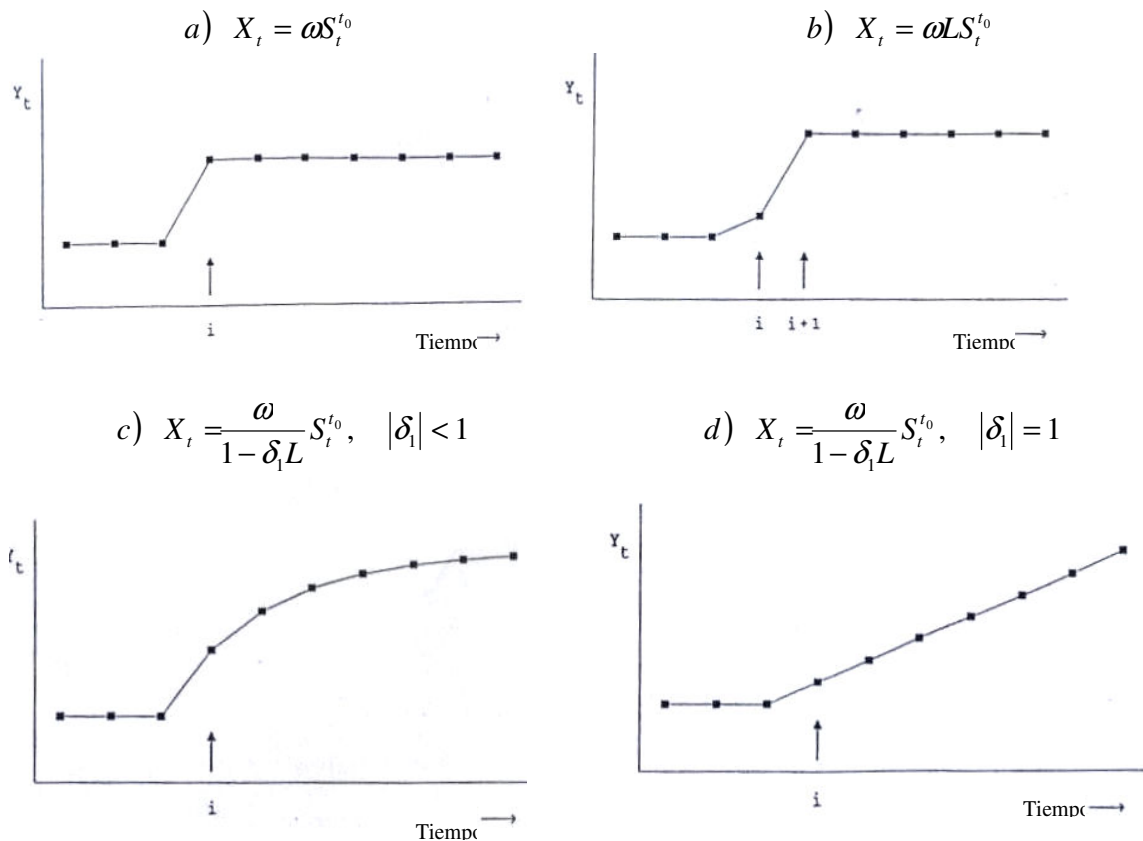
Y para la variable impulso se tiene:

$$X_t = \frac{\omega L^b}{1 - \delta L} P_t^{(t_0)} \quad (4.2.6)$$

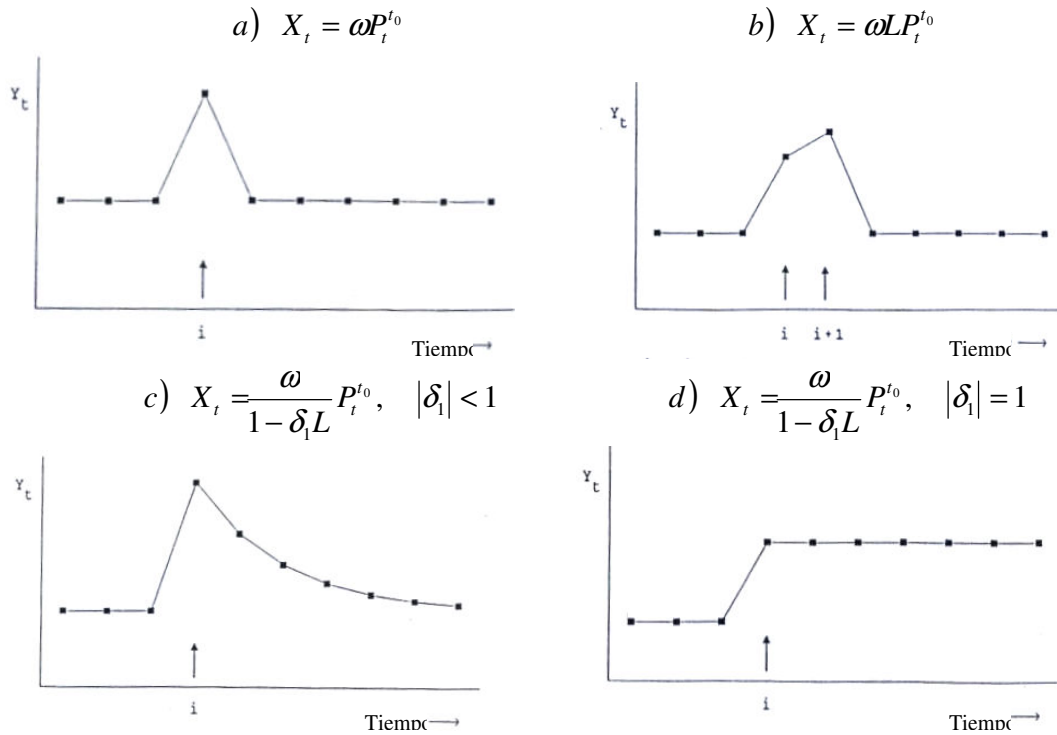
Donde  $0 \leq \delta \leq 1$ . Para  $\delta = 0$ , (4.2.5) y (4.2.6) reduce a (4.2.3) y (4.2.4), respectivamente. Si  $\delta = 1$ , el efecto de la intervención es permanente lo que lleva a que el modelo (4.2.6) se convierta en el modelo (4.2.3) teniendo en cuenta la relación de la variable escalón e impulso. Para más casos se tiene,  $0 < \delta < 1$ , siempre y cuando el comienzo es gradual.

Para mayor ilustración, en las siguientes figuras, se muestran las gráficas para ambas variables de intervención con  $b=1$  y  $0 < \delta < 1$ .

**Gráfico 4.2.3.** Diferentes formas del Comportamiento de la variable escalón en una serie temporal continúa.



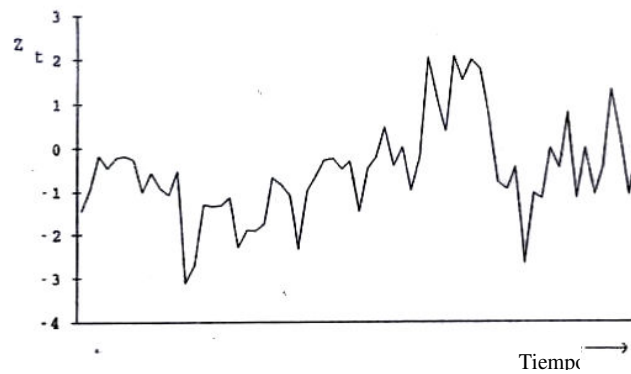
**Gráfico 4.2.4.** Diferentes formas del Comportamiento de la variable impulso en una serie Temporal continúa.



Debemos comentar que en ocasiones suelen combinarse variables escalón e impulso, en cuyo caso puede obtenerse el siguiente modelo de intervención, y resultará de utilidad cuando se espera que el efecto externo deje un efecto residual permanente después de una declinación gradual del impacto.

$$X_t = \left( \frac{\omega_1 L}{1 - \delta L} + \frac{\omega_2 L}{1 - L} \right) P_t^{t_0} \quad (4.2.7)$$

**Gráfico 4.2.5.** Comportamiento de la variable escalón e impulso en una misma serie temporal continúa.



Sin embargo, teniendo en cuenta la relación:  $P_t^{t_0} = (1-L)S_t^{t_0}$ , la expresión (4.2.7), puede tomar la siguiente estructura matemática:

$$X_t = \left( \frac{\omega_1 L}{1-\delta L} P_t^{t_0} + \omega_2 L S_t^{t_0} \right)$$

En conclusión, podemos señalar que todos los modelos de intervención discutidos en esta sección pueden especificarse mediante la siguiente expresión general:

$$X_t = v(L)I_t^{t_0}$$

Siendo  $I_t^{t_0}$  una variable impulso o escalón.

$$v(L) = \frac{\omega(L)L^b}{\delta(L)}$$

Donde:

$$\omega(L) = \omega_0 - \omega_1 L - \omega_2 L^2 - \dots - \omega_s L^s$$

$$\delta(L) = 1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r$$

En el cuadro siguiente, se recogen los valores que deberán tener b, r y s (subíndice de la expansión de cada uno de los parámetros) en la expresión anterior, para poder representar en  $X_t = v(L)I_t^{t_0}$  cada uno de los modelos expuestos.

**Tabla 4.2.1.** Valores de b, r y s.

	$I_t^{t_0} = S_t^{t_0}$	$I_t^{t_0} = P_t^{t_0}$
r = 0 b = 0	$X_t = \omega S_t^{t_0}$	Modelo no significativo
r = 0 b = 1	$X_t = \omega L S_t^{t_0}$	$X_t = \omega L P_t^{t_0}$
r = 1 b = 0	$X_t = \frac{\omega}{1-\delta L} S_t^{t_0}$	$X_t = \frac{\omega}{1-\delta L} P_t^{t_0}$
r = 1 b = 1	$X_t = \frac{\omega L}{1-\delta L} S_t^{t_0}$	$X_t = \frac{\omega L}{1-\delta L} P_t^{t_0}$
r = 2 b = 0	Modelo no significativo	$X_t = \frac{\omega}{1-\delta_1 L - \delta_2 L^2} P_t^{t_0}$



Obsérvese que en todos ellos el valor de  $s$  es cero porque,  $\omega(L) = \omega_0 = \omega$ . Dentro de esta expresión no cabe el modelo de intervención considerado, por la combinación de las variables escalón e impulso, dado que éste, como hemos visto es compuesto.

Un modelo para  $k$  intervenciones sería:

$$X_t = \sum_{j=1}^k v_j(L) I_{jt}^{t_0, j}$$

Donde:

$$v_j(L) = \frac{\omega_j(L)L^{bj}}{\delta_j(L)}$$

### 4.3 IDENTIFICACIÓN DE LOS MODELOS ARIMA CON INTERVENCIÓN

Dada una serie temporal  $\{y_t\}$ , en la que algún suceso externo produce modificaciones sustanciales en el comportamiento de la misma, será preciso tener en cuenta en su modelización la inclusión de alguna intervención (o varias). Veamos ahora como identificar un modelo ARIMA con intervención (intervenciones), para ello comenzamos obteniendo el modelo ARIMA univariante, como sigue:

$$\phi(L)(1-L)^d y_t = \theta(L)u_t$$

Donde:

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$$

$$\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$$

De manera que las raíces de ambos polinomios,  $\phi(L)$  y  $\theta(L)$ , deben caer fuera del círculo unitario para cumplir la invertibilidad; otra condición que deben de cumplir estos polinomios es la estacionariedad, teniendo en cuenta que  $(1-L)^d$ , es el operador de diferencias consecutivas donde el exponente  $d$  indica el numero de diferencias que se realizaron para suavizar la serie y  $u_t$  un ruido blanco.

Considerando que la media de  $\{y_t\}$  es igual a cero.

Obviamente, si el modelo fuese estacional multiplicativo ARIMA (p,d,q) x ARIMA (P,D,Q)<sub>s</sub>, La forma general de escribirla sería:

$$y_t = \frac{\Theta(L^s)\theta(L)}{\Phi(L^s)\phi(L)(1-L^s)^D(1-L)^d} u_t$$

Si quisiéramos incorporar ahora el modelo de intervención, la ecuación debería modificarse así:

$$y_t = v(L)I_t^{t_0} + \frac{\Theta(L^s)\theta(L)}{\Phi(L^s)\phi(L)(1-L^s)^D(1-L)^d} u_t$$

La ecuación puede generalizarse al modelo con  $k$  intervenciones, en cuyo caso escribiremos:

$$y_t = \sum_{j=1}^k v_j(L)I_{jt}^{t_0j} + N_t$$

Donde:

$$N_t = \frac{\Theta(L^s)\theta(L)}{\Phi(L^s)\phi(L)(1-L^s)^D(1-L)^d} u_t$$

Ahora, la tarea consiste en identificar el modelo recién escrito.

#### 4.4 ANÁLISIS DE OUTLIERS EN SERIES TEMPORALES

Cuando el tiempo y la causa de los factores externos creadores de observaciones atípicas son desconocidos, por ejemplo los errores inadvertidos de mecanografiar y la grabación, provocan repercusiones de interrupción y crean observaciones espurias que son inconsistentes con lo de la serie. Tales observaciones hacen referencia al análisis de outliers de series temporales

Justificada la relevancia del tratamiento de las observaciones atípicas (outliers) en el análisis de ST, se pueden describir los tipos de outliers, así como también los efectos de los mismos. La forma en que pueden detectarse y tratarse éstos.

##### 4.4.1 Tipos de Outliers.

Denominaremos por  $\{y_t\}$  la serie observada, y por  $\{z_t\}$  la serie libre de outliers. Suponiendo que  $\{z_t\}$  sigue un proceso estacional multiplicativo ARIMA (p,d,q) x ARIMA (P,D,Q)<sub>s</sub>, y considerando sin pérdida de generalidad que la media es cero, escribiremos:

$$\Phi(L^s)\phi(L)(1-L^s)^D(1-L)^d z_t = \Theta(L^s)\theta(L)u_t$$

Donde:

$$\Phi(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_p L^{ps}$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

$$\Theta(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_q L^{qs}$$

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

Los operadores anteriormente expandidos, son operadores estacionales e invertibles, sin factores comunes, y siendo  $\{u_t\}$  una secuencia de ruidos blancos.

En el caso particular en que  $z_t$  venga generada por un proceso ARMA (p,q), la ecuación anterior será igual a:

$$z_t = \frac{\theta(L)}{\phi(L)} u_t$$

De ahora en adelante supondremos que  $\{z_t\}$  viene generada por un proceso ARMA (p,q), (para simplificar la exposición y dado que la generalización resulta inmediata).

Respecto a Hillmer, Bell y Tiao, 1983, Muirhead, 1986, Tsay, 1988, Chen y Liu, 1990, Chen, Liu y Hudak, 1990, y Chen y Tiao, 1990, los cuatro tipos de outliers que se han considerado en la literatura son: el outliers aditivo (additive outliers, AO), el outliers innovacional (innovational outliers, IO), el cambio en nivel (level shift, LS) y el cambio temporal (temporary change, TC)<sup>5</sup>.

a) Outliers Aditivo (AO)

Un outliers aditivo es un suceso que afecta a una serie en un solo instante, temporal ( $t = t_0$ ), y se puede expresar como:

$$y_t = z_t + \omega I_t^{t_0}$$

Donde

$$I_t^{t_0} = \begin{cases} 1 & t = t_0 \\ 0 & t \neq t_0 \end{cases}$$

$I_t^{t_0}$  : Presencia o ausencia de Outliers en el período  $t_0$

$\omega$  : Efecto de dicho Outliers

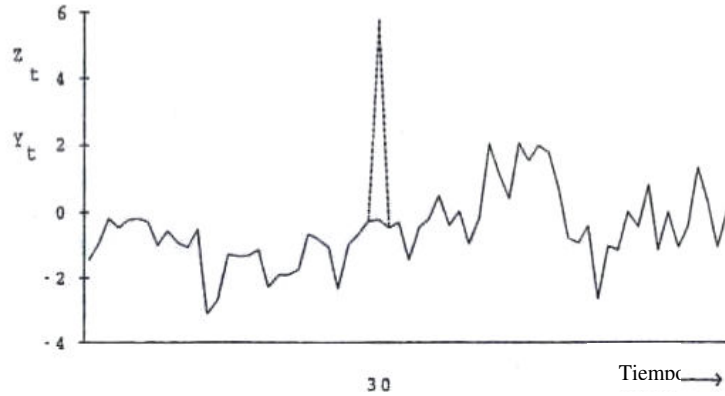
$z_t$  : Serie Libre de Outliers

$y_t$  : Serie Observada.

---

<sup>5</sup> Véase Métodos de Predicción en Economía (II), Pág. 329.

**Gráfico 4.4.1.1.** Comportamiento del Outliers Aditivo (AO) en una serie temporal continúa.



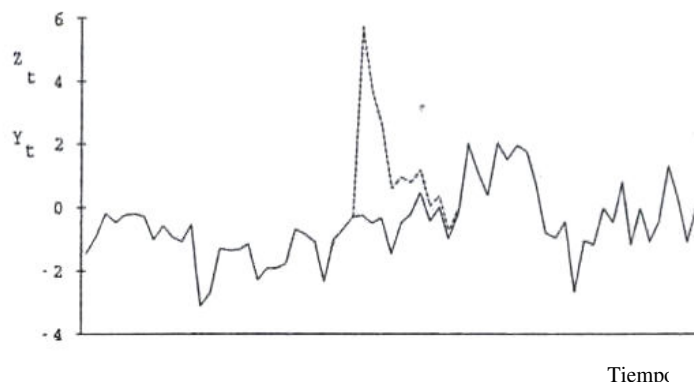
b) Outliers Innovacional (IO)

Un Outliers innovacional es un suceso cuyo efecto se propaga en conformidad con el modelo ARIMA del proceso, afectando a todos los valores observados después de su ocurrencia.

Se puede representar como sigue:

$$y_t = z_t + \frac{\theta(L)}{\phi(L)} \omega_t^{t_0}$$

**Gráfico 4.4.1.2.** Comportamiento del Outliers Innovacional (IO) en una serie temporal continúa.

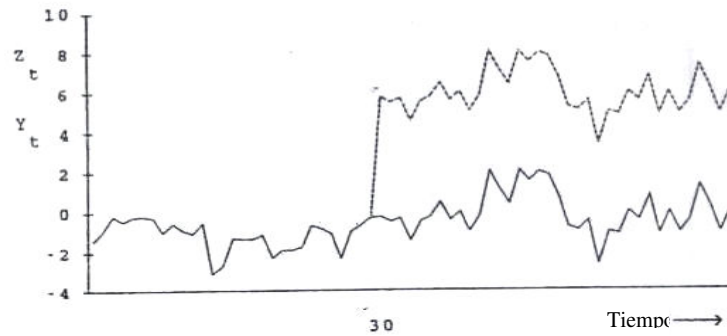


c) Cambio en Nivel (LS)

Un cambio en nivel (LS) es un suceso que afecta a una serie en un período dado y cuyo efecto es permanente, se puede representar como sigue:

$$y_t = z_t + \frac{1}{(1-L)} \omega I_t^{t_0}$$

**Gráfico 4.4.1.3.** Comportamiento del Outliers con Cambio en Nivel (LS) en una serie temporal continúa



d) Cambio Temporal (TC)

Un cambio temporal (TC) es un suceso que tiene un impacto inicial y cuyo efecto decae exponencialmente con un factor de amortiguación, que denotaremos mediante el parámetro  $\delta$  por lo que se puede expresar como:

$$y_t = z_t + \frac{1}{(1-\delta L)} \omega I_t^{t_0} \quad \text{Para } 0 < \delta < 1$$

Debe tenerse en cuenta que una serie temporal puede contener varios Outliers (por ejemplo  $k$ ) de diferentes tipos pudiéndose expresar el modelo general de Outliers como:

$$y_t = \sum_{j=1}^k \omega_j v_j(L) I_{t_j} + z_t$$

Donde:

$$z_t = \frac{\theta(L)}{\phi(L)} u_t$$

$$v_j(L) = \begin{cases} 1 & \text{Para un AO en } t = t_{0j} \\ \frac{\Theta(L)}{\phi(L)} & \text{Para un IO en } t = t_{0j} \\ \frac{1}{(1-L)} & \text{Para un LS en } t = t_{0j} \\ \frac{1}{(1-\delta L)} & \text{Para un TC en } t = t_{0j} \end{cases}$$

#### 4.4.2 Efectos de los Outliers.

Distinguiremos entre los efectos de los outliers en la serie original y en la serie residual. Los Efectos de los Outliers en la serie Original resultan inmediatos a partir de las expresiones de los Outliers Aditivos, Innovacional, Cambio en Nivel, y Cambio Temporal. En concreto cabe resaltar las siguientes observaciones respecto a los Outliers.

- 1) Los efectos de los Outliers son independientes de la estructura ARIMA de las series, salvo del caso de Outliers Innovacional.
- 2) Los Outliers AO y LS son caso límites del TC. En efecto, dada la definición del cambio temporal (TC) cuando  $\delta = 0$  estamos ante el OA y cuando  $\delta = 1$  ante el cambio en nivel.
- 3) El Outliers Aditivo Causa un efecto inmediato y único en la serie observada en el período  $t_0$ , de magnitud  $\omega$ .
- 4) En el caso de un Cambio Temporal (TC), el Outliers produce un efecto inicial  $\omega$  en el período  $t_0$ , y este efecto decae gradualmente en el tiempo, con un factor de amortiguación  $\delta$  (el valor de este parámetro puede especificarlo el analista, Chen y Liu (1990) aconsejan utilizar el valor  $\delta = 0.7$ ).
- 5) En el caso de un Cambio de Nivel, el Outliers introduce un cambio brusco y permanente en la serie observada de tipo escalón.
- 6) El efecto del Outliers Innovacional es el más complicado por cuanto depende de la estructura ARIMA de la serie. En concreto a partir de la expresión  $y_t = z_t + \frac{\theta(L)}{\phi(L)} \omega I_t^{t_0}$  podemos concluir que cuando un outliers IO ocurre en  $t = t_0$ , el efecto de ese outliers en  $t_0 + j$  para  $j \geq 0$  es igual a  $\omega \psi_j$ , donde  $\omega$  es el efecto inicial. Y  $\psi_j$  es el coeficiente j-ésimo del polinomio  $\psi(L)$ .

Para analizar los efectos Outliers en la serie residual, comenzaremos suponiendo el caso ideal (e irreal) de que los parámetros del modelo ARMA de la serie son conocidos. En este caso, se trata de examinar las series filtradas, que se escriben como:

$$\hat{u}_t = \pi(L)Y_t$$

Donde: 
$$\pi(L) = \frac{\theta(L)}{\phi(L)} = 1 - \pi_1 L - \pi_2 L^2 - \dots - \pi_g L^g$$

Considerando la expresión anterior y que tenemos un solo Outliers que se produce en el período  $t = t_0$ . Podemos reescribir  $\hat{u}_t$  de acuerdo con el tipo de Outliers presente como sigue:

$$AO : \hat{u}_t = \omega \pi(L) I_t^{t_0} + u_t$$

$$IO : \hat{u}_t = \omega I_t^{t_0} + u_t$$

$$LS : \hat{u}_t = \frac{\omega}{1-L} \pi(L) I_t^{t_0} + u_t$$

$$TC : \hat{u}_t = \frac{\omega}{1-\delta L} \pi(L) I_t^{t_0} + u_t$$

Pueden escribirse las expresiones anteriores de forma conjunta como sigue:

$$\hat{u}_t = \omega x_{it} + u_t$$

Donde  $i = 1, 2, 3, 4$

$$\text{Siendo } x_{it} = \begin{cases} x_{1t} = \pi(L) I_t^{t_0} \\ x_{2t} = I_t^{t_0} \\ x_{3t} = \frac{1}{1-L} \pi(L) I_t^{t_0} \\ x_{4t} = \frac{1}{1-\delta L} \pi(L) I_t^{t_0} \end{cases}$$

$$\text{Y donde: } x_{it} = \begin{cases} 0 & \text{para } t < t_0 \text{ y } \forall_i \\ 1 & \text{para } t = t_0 \text{ y } \forall_i \\ -\pi_j & \text{para } t = t_0 + j (j = 1, 2, \dots, T - t_0), i = 1 \text{ (AO)} \\ 0 & \text{para } t = t_0 + j (j = 1, 2, \dots, T - t_0), i = 2 \text{ (IO)} \\ 1 - \sum_{h=1}^j \pi_h & \text{para } t = t_0 + j (j = 1, 2, \dots, T - t_0), i = 3 \text{ (LS)} \\ -\delta^j - \sum_{h=1}^{j-1} \delta^{j-h} \pi_h - \pi_j & \text{para } t = t_0 + j (j = 1, 2, \dots, T - t_0), i = 4 \text{ (TC)} \end{cases}$$

Se observa que todos los outliers salvo el innovacional (IO) que acota sus efectos en los residuos en el instante temporal  $t = t_0$ , propongan su efecto en todos los residuos correspondientes a los períodos de tiempo posteriores a  $t_0$ , dependiendo además de dichos efectos de los coeficientes  $\pi_j$ , los cuales, recuérdese la expresión  $\hat{u}_t = \pi(L)Y_t$ , están en función de los parámetros del modelo ARMA identificado para la serie.

## 4.5 DETECCIÓN Y TRATAMIENTO DE LOS OUTLIERS.

La detección de outliers es importante, por cuanto como señalan Chen, Liu y Hudak (1990)<sup>6</sup>, un tratamiento adecuado de los mismos permite mejorar:

- 1) La comprensión de la serie objeto de análisis.
- 2) La modelización y estimación.
- 3) El análisis de intervención.
- 4) La calidad de las predicciones.

En la práctica, ni  $t_0$  y los valores de los parámetros  $(\phi_j, \theta_j, \sigma_u^2)$  son conocidos, por lo cual estos deben estimarse, es notorio además, que las existencias de outliers causa sesgos importantes en las estimaciones de estos parámetros, concretamente,  $\sigma_u^2$  tendera a estar sobreestimado.

Para detectar outliers en este caso se han propuesto procedimientos interactivos, como el de Chang y Tiao (1983) o el de Tsay (1986), los cuales se centran en detectar los posibles outliers, y especificar posteriormente, de acuerdo con el tipo de outliers identificado, un modelo de intervención, tal como se expresa en términos genéricos en la siguiente

expresión.  $y_t = \sum_{j=1}^k \omega_j v_j(L) I_{ij} + z_t$ , y finalmente estimar dicho modelo.

Alternativamente a estos procedimientos, más recientemente Chen y Liu (1990)<sup>7</sup> han diseñado un procedimiento que consiste en la detección de los outliers y la estimación conjunta de los parámetros del modelo y los efectos de los outliers.

A continuación describiremos cada uno de estos enfoques, centrándonos en el perímetro de ellos, en el procedimiento iterativo de Chang y Tiao (1983).

### 4.5.1 Procedimiento Iterativo de Detección de Outliers

Chang y Tiao (1983) han propuesto un procedimiento interactivo para detectar posibles outliers y el tipo de ellos, que consta de las siguientes etapas.

*Etapa 1:* Modelizamos la serie  $\{y_t\}$  suponiendo que no hay outliers y estimamos el modelo identificado:

$$\tilde{\phi}(L)y_t = \tilde{\theta}(L)\hat{u}_t$$

---

<sup>6</sup> Como la afirma Antonio Aznar y Francisco Javier Trivez, en su edición de Métodos de Predicción en economía (II), Pág. 328

<sup>7</sup> Véase Antonio Aznar y Francisco Javier Trivez, en su libro Métodos de Predicción en Economía (II), Pág. 337



O lo que es igual:

$$y_t = \frac{\tilde{\theta}(L)}{\tilde{\phi}(L)} \hat{u}_t$$

Donde:

$$\begin{aligned}\tilde{\theta}(L) &= 1 - \tilde{\theta}_1 L - \tilde{\theta}_2 L^2 - \dots - \tilde{\theta}_q L^q \\ \tilde{\phi}(L) &= 1 - \tilde{\phi}_1 L - \tilde{\phi}_2 L^2 - \dots - \tilde{\phi}_p L^p\end{aligned}$$

A partir de  $y_t = \frac{\tilde{\theta}(L)}{\tilde{\phi}(L)} \hat{u}_t$  los residuos son iguales a:

$$\hat{u}_t = \tilde{\pi}(L) Y_t$$

Donde:

$$\tilde{\pi}(L) = \frac{\tilde{\theta}(L)}{\tilde{\phi}(L)} = 1 - \tilde{\pi}_1 L - \tilde{\pi}_2 L^2 \dots$$

Y la estimación de  $\sigma_u^2$  será:

$$\tilde{\sigma}_u^2 = \frac{1}{T} \sum_{t=1}^T (\hat{u}_t - \bar{\hat{u}})^2$$

*Etapa 2:* Calculamos  $\tilde{\lambda}_{i,t}$  donde  $i = 1, 2, 3, 4$  utilizando el modelo estimado, esto es, teniendo en cuenta los estadísticos calculados antes.

$$\begin{aligned}\tilde{\lambda}_{1,t_0} &= \frac{\tilde{\tau} \tilde{\omega}_{AO}(t_0)}{\tilde{\sigma}_u} \\ \tilde{\lambda}_{2,t_0} &= \frac{\tilde{\omega}_{IO}(t_0)}{\tilde{\sigma}_u} \\ \tilde{\lambda}_{3,t_0} &= \frac{\tilde{\tau} * \tilde{\omega}_{LS}(t_0)}{\tilde{\sigma}_u} \\ \tilde{\lambda}_{4,t_0} &= \frac{\tilde{\tau} ** \tilde{\omega}_{TC}(t_0)}{\tilde{\sigma}_u}\end{aligned}$$

Entonces, definiendo:

$$\tilde{\lambda}_{t_0} = \text{Max}_i \text{Max}_t \{ \tilde{\lambda}_{i,t} \}$$

Donde  $t_0$  denota el período para el que tiene lugar el máximo valor de  $\tilde{\lambda}$ , si:

a)  $\tilde{\lambda}_{t_0} < c$

Donde  $c$  es una constante positiva predeterminada, a la que nos referiremos más adelante, diremos que no hay ningún outliers, finalizado el procedimiento.

$$\text{b.1) } \tilde{\lambda}_{t_0} = \left| \tilde{\lambda}_{1,t_0} \right| > c$$

Diremos que en el período  $t_0$  ocurre un outliers AO, cuyo efecto estimado es  $\tilde{\omega}_{AO}(t_0)$ .

En este caso, podremos modificar los valores de la serie, utilizando  $y_t = z_t + \omega I_t^{t_0}$ , como sigue:

$$\tilde{y}_t = y_t - \tilde{\omega}_{AO}(t_0)I_t^{t_0}$$

Y definir los nuevos residuos:  $\tilde{u}_t = \tilde{\pi}(L)\tilde{y}_t = \hat{u}_t - \tilde{\omega}_{AO}(t_0)\tilde{\pi}(L)I_t^{t_0}$

A partir de estos nuevos residuos ( $\tilde{u}_t$ ) se calculará una nueva estimación de  $\sigma_u^2$ :

$$\tilde{\sigma}_u^2 = \frac{1}{T} \sum_{t=1}^T (\tilde{u}_t - \bar{\tilde{u}})^2$$

$$\text{b.2) } \tilde{\lambda}_{t_0} = \left| \tilde{\lambda}_{2,t_0} \right| > c$$

Diremos que en el período  $t_0$  ocurre un outliers IO, cuyo efecto es  $\tilde{\omega}_{IO}(t_0)$ .

También en este caso puede eliminarse el efecto IO modificando la serie utilizando

$y_t = z_t + \frac{\theta(L)}{\phi(L)} \omega I_t^{t_0}$ , obteniendo:

$$\tilde{y}_t = y_t - \frac{\tilde{\theta}(L)}{\tilde{\phi}(L)} \tilde{\omega}_{IO}(t_0)I_t^{t_0}$$

Siendo los nuevos residuos

$$\tilde{u}_t = \tilde{\pi}(L)\tilde{y}_t = \hat{u}_t - \tilde{\omega}_{IO}(t_0)\tilde{\pi}(L)I_t^{t_0}$$

$$\text{b.3) } \tilde{\lambda}_{t_0} = \left| \tilde{\lambda}_{3,t_0} \right| > c$$

Diremos que en el período  $t_0$  ocurre un outliers LS, cuyo efecto es  $\tilde{\omega}_{LS}(t_0)$ .

También en este caso puede eliminarse el efecto LS modificando la serie utilizando

$y_t = z_t + \frac{1}{(1-L)} \omega I_t^{t_0}$ , obteniendo:

$$\tilde{y}_t = y_t - \frac{1}{(1-L)} \tilde{\omega}_{LS}(t_0)I_t^{t_0}$$

Siendo los nuevos residuos los cuales estimaremos nuevamente  $\sigma_u^2$

$$\tilde{u}_t = \tilde{\pi}(L)\tilde{y}_t = \hat{u}_t - \frac{\tilde{\omega}_{LS}(t_0)}{(1-L)} \tilde{\pi}(L)I_t^{t_0}$$

$$\text{b.4) } \tilde{\lambda}_{t_0} = \left| \tilde{\lambda}_{4,t_0} \right| > c$$

Diremos que en el período  $t_0$  ocurre un outliers TC, cuyo efecto es  $\tilde{\omega}_{TC}(t_0)$ .

También en este caso puede eliminarse el efecto TC modificando la serie utilizando

$$y_t = z_t + \frac{1}{(1-\delta L)} \omega I_t^{t_0}, \text{ obteniendo:}$$

$$\tilde{y}_t = y_t - \frac{1}{(1-\delta L)} \tilde{\omega}_{TC}(t_0) I_t^{t_0}$$

Siendo los nuevos residuos los cuales estimaremos nuevamente  $\sigma_u^2$

$$\tilde{u}_t = \tilde{\pi}(L) \tilde{y}_t = \hat{u}_t - \frac{\tilde{\omega}_{TC}(t_0)}{(1-\delta)} \tilde{\pi}(L) I_t^{t_0}$$

**Etapa 3:** Volvemos a calcular los  $\tilde{\lambda}_{i,t}$  donde  $i=1,2,3,4$  Basándonos ahora en los residuos  $\tilde{u}_t$  y, consecuentemente, en  $\tilde{\sigma}_u^{*2}$ , y repetimos la etapa 2 hasta que todos los outliers son identificados (esto es, hasta que en alguna iteración nos encontramos con el caso a) de la etapa 2). Debe tenerse en cuenta que las estimaciones iniciales en  $\pi(L)$  permanecen invariables.

**Etapa 4:** Supongamos que la etapa 3 finaliza con la detección de outliers en los períodos  $t_{01}, t_{02}, t_{03}, \dots, t_{0k}$ . Entonces, trataremos estos períodos como si fueran conocidos estimando los parámetros de los outliers  $\omega_1, \omega_2, \dots, \omega_k$  y los parámetros  $\phi_j$  y  $\theta_j$  simultáneamente utilizando el modelo

$$y_t = \sum_{j=1}^k \omega_j v_j(L) I_{t_j} + z_t$$

Donde:

$$v_j(L) = \begin{cases} 1 & \text{Para un AO en } t = t_{0j} \\ \frac{\Theta(L)}{\phi(L)} & \text{Para un IO en } t = t_{0j} \\ \frac{1}{(1-L)} & \text{Para un LS en } t = t_{0j} \\ \frac{1}{(1-\delta L)} & \text{Para un TC en } t = t_{0j} \end{cases}$$

Desde la estimación de  $y_t = \sum_{j=1}^k \omega_j v_j(L) I_{t_j} + z_t$  podemos calcular unos nuevos residuos como sigue:

$$\hat{u}_t^{(t)} = \tilde{\pi}^{(1)}(1) \left[ y_t - \sum_{j=1}^k \tilde{\omega}_j v_j(L) I_t^{t_{0j}} \right]$$

Obteniendo por tanto una nueva estimación de  $\sigma_u^2$

$$\tilde{\sigma}_u^{2(1)} = \frac{1}{T} \sum_{t=1}^T (\tilde{u}_t^{(1)} - \bar{\tilde{u}}^{(1)})^2$$

Las etapas 2 a 4 son repetidas sucesivamente hasta que todos los outliers son identificados y sus impactos son estimados simultáneamente.

El modelo de los outliers estimado será:

$$y_t = \sum_{j=1}^k \tilde{\omega}_j \tilde{v}_j(L) I_t^{t_0j} + \frac{\tilde{\theta}(L)}{\tilde{\phi}(L)} u_t$$

Donde  $\tilde{\omega}_j$ ,  $\tilde{\phi}(L)$  y  $\tilde{\theta}(L)$  son los estimadores obtenidos en la iteración final.

Respecto a este procedimiento de detección de outliers cabe efectuar dos observaciones:

1) El Valor de la constante positiva  $c$ , que constituye el punto crítico para aceptar o rechazar la hipótesis nula de ausencia de outliers, se recomienda a partir de determinados experimentos de simulación que debe tomarse valores comprendidos entre 3 y 4.5. Lo habitual es tomar el valor 3, sobre todo cuando el tamaño muestral disponible no es muy elevado.

2) Si bien el procedimiento se ha explicado considerando los cuatro tipos de outliers, solamente había sido desarrollado para los outliers AO e IO.

Finalmente hay que resaltar que si bien este procedimiento iterativo es efectivo cuando se trata de detectar las localizaciones y estimar los efectos de grandes outliers aislados, Chen y Liu (1990) encuentran tres problemas que surgen a partir del mismo:

- (a) La presencia de outliers puede hacer que el modelo no se especifique adecuadamente.
- (b) Incluso si el modelo está adecuadamente especificado, los outliers pueden producir sesgos importantes en los estimadores de los parámetros.
- (c) Puede que no se identifiquen algunos outliers "debido a un efecto de enmascaramiento.

Con el fin de paliar estos problemas (y más concretamente los dos últimos referidos), se ha propuesto el método alternativo de detección de outliers que describimos en el apartado siguiente:

#### 4.5.2 Procedimiento de detección de outliers y Estimación Conjunta de los Parámetros y los efectos de los Outliers.

Chen y Liu (1990) se centran en los problemas b) y c), diseñando un procedimiento que, según sus palabras, “este libre del efecto espúreo y ocultamiento de la detección de outliers, y sea capaz de estimar conjuntamente los parámetros del modelo y los efectos de los outliers”

El punto de partida para este procedimiento viene dado por el modelo general de los outliers

múltiples,  $y_t = \sum_{j=1}^k \omega_j v_j(L) I_{jt}^{t_0j} + z_t$ . Y la siguiente expresión de los residuos, que se deriva y

directamente a partir de  $\hat{u}_t = \pi(L)y_t$ :  $\hat{u}_t = \sum_{j=1}^k \omega_j \pi(L) v_j(L) I_{jt}^{t_0j} + u_t$ .

Esto es  $y_t = \sum_{j=1}^k \omega_j v_j(L) I_{jt}^{t_0j} + z_t$ , multiplicado en ambos lados por  $\pi(L)$ , y conociendo que

$\hat{u}_t = \pi(L)y_t$ , entonces tenemos:

$$\pi(L)y_t = \pi(L) \sum_{j=1}^k \omega_j v_j(L) I_{jt}^{t_0j} + z_t = \sum_{j=1}^k \omega_j \pi(L) v_j(L) I_{jt}^{t_0j} + z_t \pi(L) \Rightarrow \hat{u}_t = \sum_{j=1}^k \omega_j \pi(L) v_j(L) I_{jt}^{t_0j} + \hat{u}$$

Obviamente, si el efecto de un outliers y su localización se conocieran, se trataría de especificar adecuadamente la expresión:  $y_t = \sum_{j=1}^k \omega_j v_j(L) I_{jt}^{t_0j} + z_t$  y estimar los parámetros de

dicho modelo, lo habitual sin embargo, es que no se conocen, por lo cual debe diseñarse un procedimiento que permita tanto detectar los posibles outliers, como estimar los parámetros del modelo y los efectos de dichos outliers.

El procedimiento desarrollado con este fin por Chen y Liu (1990) consta de las tres etapas siguientes:

*Etapas 1:* Estimación de los parámetros y detección inicial de outliers.

1.1 Estimamos el modelo ARMA identificado utilizando las series observadas ajustadas más recientemente (el procedimiento comienza sin ajuste, tal como se señala en la etapa 1 del procedimiento iterativo de Chang y Tiao, 1993). Calculamos una serie de residuos.

1.2 Si descubrimos un outliers potencial, eliminamos su efecto de los residuos y de las observaciones, teniendo en cuenta el tipo de outliers detectado, y repetimos la búsqueda de outliers. En caso contrario pasamos la siguiente etapa.

1.3 Si no se ha encontrado ningún outliers, podemos concluir que la serie esta libre de outliers, ahora bien, si hemos encontrado outliers a partir de la estimación inicial de los parámetros, revisamos la estimación del modelo con los nuevos datos ajustados y repetimos las etapas iniciales. El procedimiento se repite hasta que no se encuentra ningún nuevo outliers.

*Etapa 2:* Estimación conjunta de efectos de outliers y parámetros del modelo.

2.1 Supongamos que hemos identificado  $k$  posibles outliers de diferentes tipos. Entonces, los efectos de los outliers,  $\omega_j$ 's, pueden estimarse conjuntamente utilizando el

modelo de regresión múltiple  $\hat{u}_t = \sum_{j=1}^k \omega_j \pi(L) v_j(L) I_{jt}^{t_0^j} + u_t$ .

2.2 Calculamos los estadísticos  $\tilde{\lambda}$  de los  $\tilde{\omega}_j$  estimados. Si el valor mínimo, en términos absolutos, de estos estadísticos ( $\tilde{\lambda}_j$ ) es inferior a un determinado valor  $\min_j |\tilde{\lambda}_j| = \tilde{\lambda}_L \leq c$  Crítico, siendo  $c$  el valor crítico utilizado en el procedimiento interactivo, habitualmente igual a 3, entonces el outliers del período  $t_{0L}$  es eliminado del conjunto de outliers identificados, volviendo a repetir la etapa 2.1 con los  $k-1$  outliers restantes. Repetimos 2.1 - 2.2 hasta que encontramos que el valor mínimo, en términos absolutos, del estadístico  $\tilde{\lambda}$  correspondiente a los outliers no eliminados, supera a  $c$ , en cuyo caso pasamos a 2.3.

2.3 Ajustamos la serie eliminando sólo los efectos de los outliers significativos, entendiendo por tales los así detectados a partir de 2.1 y 2.2, esto es los que no han sido eliminados.

2.4 Calculamos las estimaciones Máximo-verosímiles de los parámetros del modelo en base a la serie ajustada obtenida en 2.3, Si la variación en la estimación de  $\sigma_u$  con respecto a la estimación previa, es mayor que un determinado valor prefijado  $\lambda$  (habitualmente,  $\lambda=0.001$ ), volvemos a 2.1, para efectuar iteraciones adicionales, en otro caso pasamos a 3.1.

*Etapa 3:* Detección de outliers a partir de las estimaciones finales de los parámetros.

3.1 Calculamos los residuos filtrando la serie original, teniendo en cuenta las estimaciones de los parámetros calculadas en 2.4.

3.2 Utilizamos los residuos obtenidos en 3.1 para volver a repetir las etapas 1 y 2, teniendo en cuenta: a) que las estimaciones de los parámetros utilizadas para repetir 1.2 - 1.4 son las fijadas en 2.4, b) que se omiten las etapas 2.3 y 2.4.

3.3 Los  $\tilde{\omega}_j$  estimados en 2.1, son las estimaciones finales de los efectos de los outliers detectados en la última iteración.

El procedimiento descrito puede aplicarse directamente a un modelo de intervención.

Las principales diferencias entre este procedimiento y el referido en el apartado 4.5.1 son las siguientes:

- a) La detección de outliers se efectúa interactivamente considerando tanto residuos como observaciones ajustadas.
- b) No se precisa llevar a cabo la estimación de los de intervención.
- c) Los outliers se detectan en base a estimaciones robustas de los parámetros del modelo, es decir, las estimaciones de los parámetros del modelo son poco sensibles a variaciones ocasionales en la muestra, así como a la presencia de outliers.
- d) Los efectos de los outliers se estiman conjuntamente utilizando un modelo de regresión múltiple.

Para finalizar, cabe señalar que si bien los dos procedimientos de detección de outliers descritos en esta sección son complejos, especialmente el último; pero la aplicabilidad de los mismos es inmediata, más cuando se conoce algún software en el que estén disponibles dichos procedimientos.

## 5. OBJETIVOS

### Objetivo General

- ✚ Estudiar los fundamentos de series temporales con datos atípicos (Outliers) e Intervenciones, y sus aplicaciones.

### Objetivos Específicos

- ✚ Estudiar los modelos de series temporales aplicados a cada Tipo de Intervención y Outliers.
- ✚ Identificar el modelo que se usará para cada tipo de intervenciones y outliers, en una base de datos.
- ✚ Aplicar los modelos de intervención y outliers a casos concretos de series temporales.



## 6. METODOLOGÍA

Con el fin de garantizar una mejora exitosa y armoniosa de ST con Intervenciones, primero revisaremos los tópicos de interés en la bibliografía a nuestro alcance, y elaboraremos síntesis escritas, que serán presentadas al asesor para recibir la realimentación oportuna.

Al analizar series temporales interrumpidas, las funciones de Autocorrelación y Autocorrelación parcial estarán distorsionadas por el efecto de las mismas. Por lo que no serán adecuadas para poder identificar el modelo generador de la serie. Para ello tendremos dos métodos que serán utilizados en la etapa de identificación.

1- Si el subconjunto de datos anteriores o posteriores a las intervenciones es suficientemente grande, podríamos identificar el modelo ARIMA como lo hacemos tradicionalmente, tomando como base de datos dicho subconjunto, incorporando posteriormente la modelización de las intervenciones. Debe tenerse en cuenta que el modelo supuesto para la intervención será tentativo, analizándose posteriormente en la etapa de chequeo cual de ellas resulta mas adecuada.

2- Cuando no dispongamos de suficientes datos antes o después de las intervenciones, o bien al inspeccionar la serie tenemos alguna intuición respecto a la forma en que han afectado al patrón de comportamiento de la serie, podemos especificar la forma de la intervención y estimarla, luego esta se elimina de la serie original teniendo así, los residuos de la serie, los que denotamos como los valores limpios, es decir una serie sin intervención. Con estos nuevos valores se identifica el modelo ARIMA de la forma tradicional. Posteriormente combinamos el modelo de intervención y el modelo residual para obtener el modelo de intervención global, el cual se estima conjuntamente por máxima verosimilitud.

Dado que en la práctica de lo que dispondremos será de unas determinadas series temporales, de datos de casos concretos adquiridas del Banco Central de Reserva, debiendo, por lo tanto, a partir de la misma averiguar qué modelo ARIMA  $(p,d,q)$  x ARIMA  $(P,D,Q)$ , es susceptible de haber generado dicha serie, es decir, identificar qué modelo representa adecuadamente el comportamiento de la misma, con el fin de utilizarlo para obtener predicciones de valores futuros de la serie en cuestión; por lo tanto para desarrollar una aplicación, seguiremos la metodología propuesta por Box y Jenkins (1970, 1976)

Antes de cualquier análisis se tomará un 20%, de los datos para efectos de comparación con las predicciones que se generen con el modelo encontrado.

Para iniciar la ejecución de la metodología Box y Jenkins se examina, si la serie temporal objeto de estudio es estacionaria o no. Para ello se utilizan las herramientas estadística como: gráfico de secuencia, funciones de Autocorrelación y Autocorrelación parcial de la serie original, además, se podrá hacer una intuición del modelo de intervención que mas se adecua, teniendo en cuenta el periodo de comienzo y la forma general del impacto.

En caso que la serie no fuese estacionaria en varianza las transformaciones que se van a tomar en cuenta son: la raíz cuadrada, logaritmo natural y raíz cuadrada inversa. El test que permitirá detectar la mayor o menor aproximación a la homocedasticidad será el Test de Levene (a menor estadístico y mayor significancia, mejor es la transformación) para cada una de las transformaciones, antes mencionadas.

Para la estacionariedad en media debemos diferenciarla previamente hasta que pueda admitirse la estacionariedad de la misma, para garantizar la estacionariedad en media se suavizará la serie, por medio de las diferenciaciones regulares (d), y en caso de que la serie presente estacionalidad se diferenciará estacionalmente (D). Lo habitual es hacer una o dos diferenciaciones. Para conocer si se ha eliminado la estacionariedad debemos conocer que la serie transformada y/o diferenciada tenga varianza constante y media cero.

Una vez que esta serie transformada es estacionaria (en media y en varianza), se obtienen las autocorrelaciones y autocorrelaciones parciales, (estas no deben tener un decaimiento lento a lo largo del tiempo) y según el número de picos fuera de los límites de confianza en cada una de estas gráficas se deducen algunos posibles modelos que pudieron generar la serie objeto de análisis. Resulta usual el hecho de que exista más de un modelo; en estos casos se cataloga todos los posibles modelos candidatos, discriminando respecto a ellos en las etapas posteriores.

La estimación de los coeficientes se obtiene mediante el método de Predicción de Mínimos Cuadrados Incondicional con valores iniciales automáticos para la estimación. Con el estadístico  $t$  podemos ver si cada uno de los parámetros es significativo puesto que contrasta la hipótesis de que sí difiera significativamente de cero.

Luego de obtener los valores de los parámetros se focaliza en conocer si los modelos cumplen la condición de estacionariedad y de invertibilidad.

A) Los modelos de medias móviles (MA (q)) siempre son estacionarios, mientras que los modelos autoregresivo (AR (p)) son estacionario si las raíces de la ecuación  $\phi(L) = 0$  caen fuera del círculo unitario.

B) Los modelos autoregresivo (AR (p)) siempre son estacionarios, Las medias móviles (MA (q)) son invertible, cuando pueden escribirse como un modelo AR ( $\infty$ ) siempre y cuando las raíces  $\theta(L) = 0$ , caigan fuera del círculo unitario.

C) El modelo ARMA (p, q) es estacionario e invertible siempre y cuando las raíces  $\phi(L) = 0$  y  $\theta(L) = 0$  caigan fuera del círculo unidad, adicionalmente supondremos que no hay raíces comunes.

Después de este análisis podemos descartar algunos modelos.

Enseguida nos centramos fundamentalmente en analizar si los residuos del modelo ( $\hat{u}_t$ ) tienen un comportamiento similar a las perturbaciones del mismo ( $u_t$ ), esto es, si puede afirmarse si son semejantes a un ruido blanco.

Otros criterios que nos ayudaran a seleccionar el mejor modelo son: El Criterio Akaike (AIC), El criterio de información Bayesiana (BIC), donde se tomará el modelo que tenga el menor valor de estos estadísticos y el criterio de parcimonia (modelo con menos parámetros).

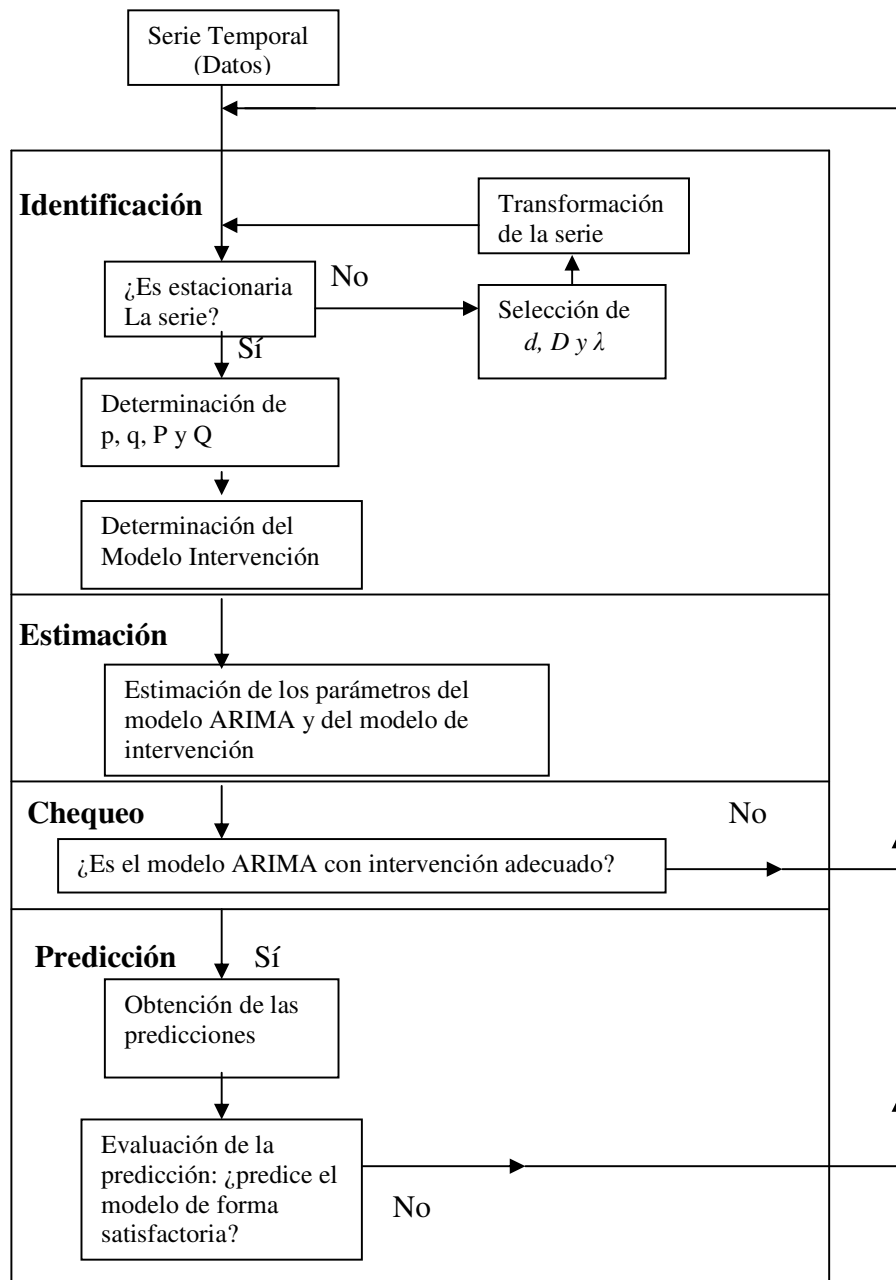
Los modelos que superen la etapa de Chequeo, tanto en la parte ARIMA clásica y la parte de Intervenciones, pueden ya ser utilizados para el fin básico de la etapa de **Predicción**; ya que se podrán encontrar valores futuros de la serie temporal (Predicciones puntuales y/o por intervalo), que se podrán comparar con el 20% de datos consignados para esto. Debe entenderse que un modelo es útil cuando cumple adecuadamente el objetivo para el que ha sido planteado (predecir adecuadamente).

Una vez obtenidas las predicciones del modelo con intervenciones se trata de volver a chequear la adecuación del mismo, pudiendo utilizar para ello métodos no paramétricos (Error Cuadrático Medio), como paramétricos (estadísticos de contenido informativo, exactitud y corroboración), este desarrollo lo podemos observar en el grafico 9.

Los métodos interactivos mencionados en el marco teórico, nos van a permitir estimar los efectos de los outliers teniéndose en cuenta que es necesario encontrar los valores de los parámetros del ARMA (caso mas simple, puede ser otro modelo), conocer la forma del impacto y el tiempo en que ocurrió el outliers. Una vez que se hayan encontrado los parámetros se procede como en las intervenciones con la única diferencia que estos datos serán solamente estimados.

Los software's que nos permitirán obtener los resultados son: STATISTICA 7.0, SPSS 13, TRAMO SEAT.

**Gráfico 6.1.** Esquema de la metodología BOX – JENKINS, con intervención.



Pasos a seguir con el método dos:

1. El análisis se inicia identificando el periodo y la forma de la intervención del gráfico original de la serie, luego se busca que la serie sea estacionaria en varianza y media, una vez ya se tiene estacionariedad, se observan los gráficos de autocorrelación estacionales y regulares. para poder plantearse los posibles modelos.
2. Estimación conjuntamente de los parámetros de cada modelo y se verifican la condición de estacionariedad e invertibilidad. También en la estimación se conocen los valores de los estadísticos AIC y BIC, permitiendo identificar el modelo que menor valor tenga y suponer que es el que debe cumplir mejor las siguientes etapas para aceptarlo como el que mas se adecua a los datos.
3. Análisis de residuos de los modelos que hayan pasado los criterios anteriores y se verifican si cumplen la condición de incorrelación y si se comportan como un ruido blanco. El modelo que se obtiene hasta aquí es un modelo ARIMA ESTIMADO.
4. Obtención de las predicciones con este modelo ARIMA ESTIMADO y se comparan los valores con los datos originales.
5. Al modelo ARIMA identificado se le incorpora el modelo de intervención y se obtienen residuos del modelo conjunto. Con estos residuos se repiten los pasos 1-4, para encontrar el verdadero modelo ARIMA con intervención

Para una aplicación de detección de Outliers se siguen los pasos siguientes:

1. Observar el grafico de la serie original que se analizara.
2. Identificar el modelo ARIMA que mas se ajusta al patrón de comportamiento de la serie teniendo en cuenta los criterios de:
  - estacionariedad en varianza y media
  - cumplimiento de estacionariedad e invertibilidad de los coeficientes de los parámetros.
  - Interpretación de los estadísticos AIC y BIC.
  - Análisis de los residuos.
  - Estabilidad de los parámetros.
3. Predecir con el modelo que se ha aceptado como el más adecuado al comportamiento de los datos.
4. Utilizar el modelo ARIMA que se ha seleccionado como adecuado, para detectar y estimar los parámetros y efectos de los datos atípicos.
5. Analizar los residuos que se obtienen incorporando el modelo de los outliers que se han detectado.
6. En caso que los residuos no cumplan satisfactoriamente las condiciones de incorrelación y de ser ruido blanco, se repite el proceso de selección del modelo ARIMA ordinario, tomando en cuenta la nueva serie con los efectos de los outliers que ya se han detectado. Con ese nuevo modelo que se encuentre se detectan outliers, el proceso para hasta que los residuos cumplan la condición de incorrelación y que sean ruido blanco.

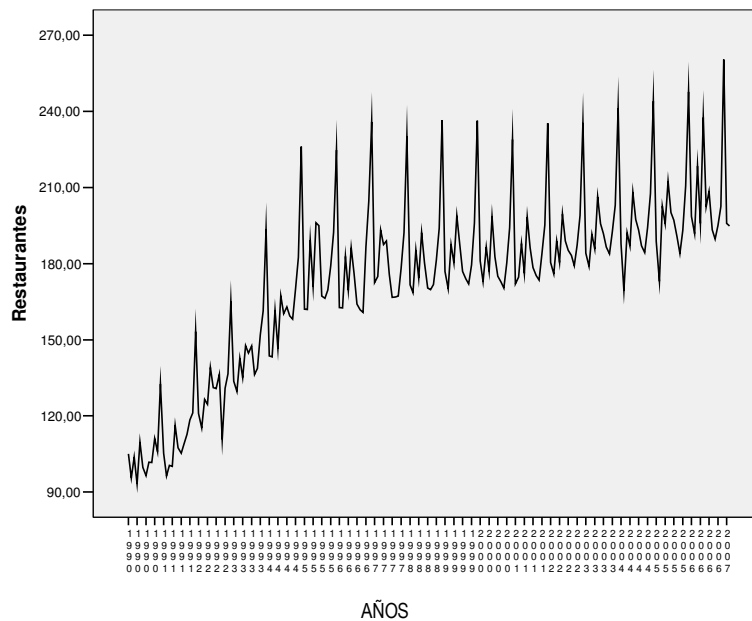
Finalmente se predice con el modelo ARIMA que se detecta en la última iteración, incorporando también el modelo de los Outliers que se han detectado con dicho modelo ARIMA.

Las bases de datos a utilizar para las aplicaciones son las siguientes:

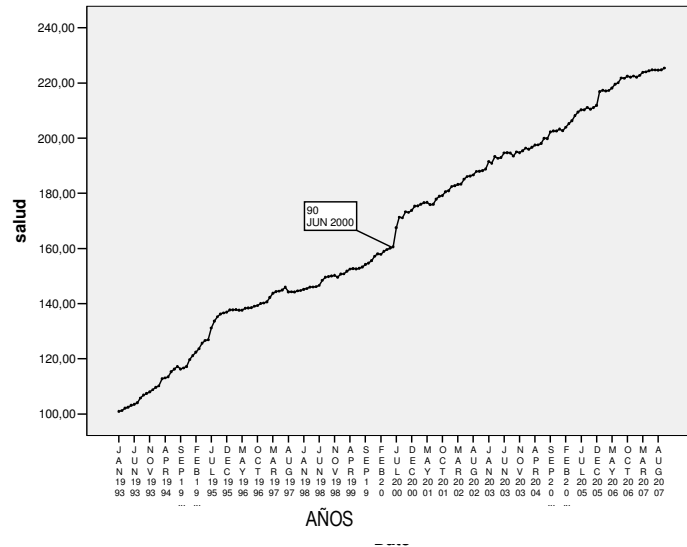
Índice del Volumen de la Actividad Económica (IVAE) en el área de Comida en Restaurantes y hoteles, Índice de Precios al Consumidor (IPC), en la clasificación del consumo individual de SALUD, como también en Prendas de Vestir y Calzado Las cuales se han obtenido del Banco Central de Reserva.

En los gráficos siguientes se observa el comportamiento de estas series temporales, donde por simple inspección se presume la existencia de intervenciones y outliers.

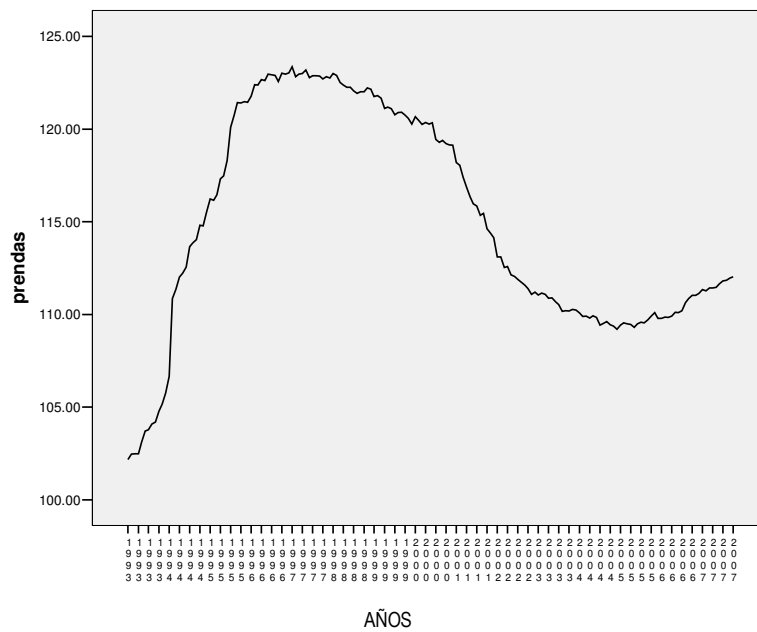
**Gráfico 6.2.** Comportamiento de la ST de IVAE en Servicios de Comidas, Restaurantes y Hoteles (Enero 1990- febrero2007).



**Gráfico 6.3.** Comportamiento de la ST del IPC, en la clasificación del consumo individual de SALUD (enero 1993- octubre 2007)



**Gráfico 6.4.** Comportamiento de la ST del IPC, en el área de Prendas de Vestir y Calzado (Enero 1993- octubre2007).



## 7. Aplicaciones de la Investigación

Con la teoría establecida, la atención se dirige a la puesta en práctica del análisis de intervención y outliers, con el fin de estimar, diagnosticar, y validar el modelo que mejor se adecue al patrón de comportamiento de la serie.

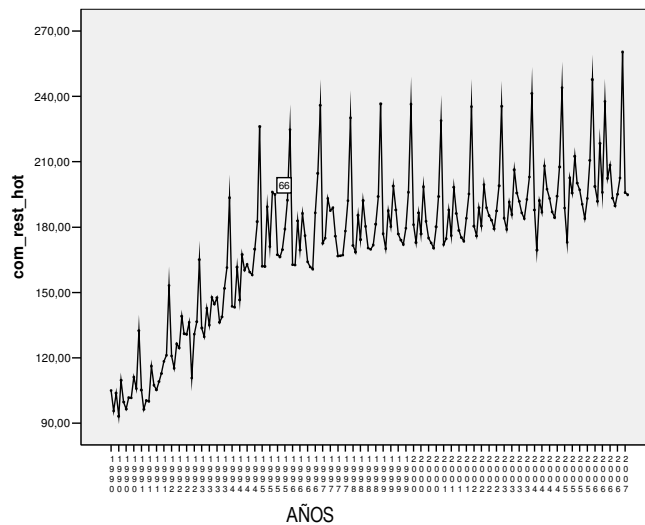
### 7.1 Aplicaciones de ST con Intervención

#### 7.1.1 Serie Temporal del IVAE en comida Restaurante y Hoteles

##### ANÁLISIS DESCRIPTIVO DE LA SERIE

La Serie Temporal de Índice del Volumen de la Activada Económica (IVAE) en Comida, Restaurantes y Hoteles, consta de 206 datos con periodicidad mensual desde enero de 1990 a febrero de 2007; Iniciamos el análisis de la serie observando e interpretando el gráfico de secuencia de la serie original.

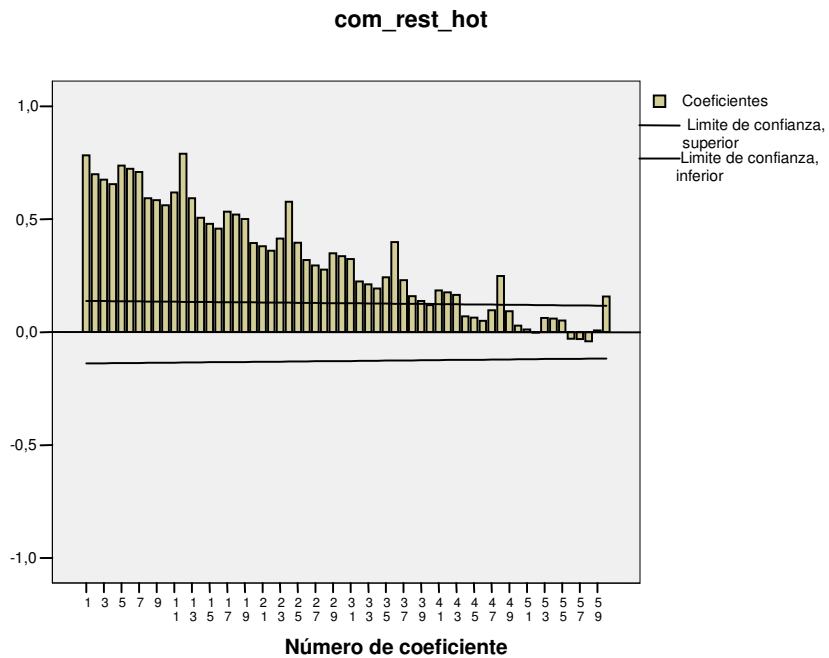
**Gráfico 7.1.1.1: Serie Original**



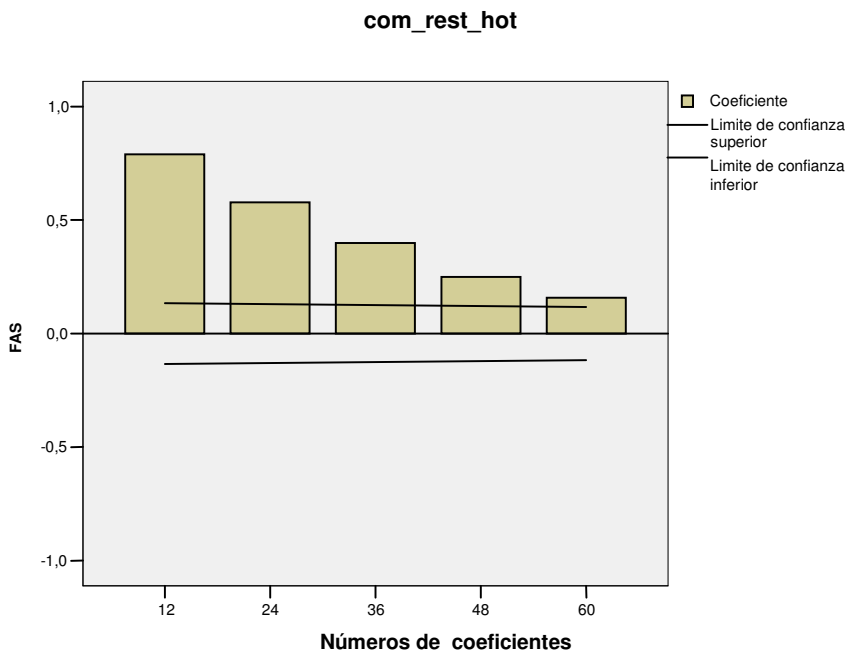
A partir de los valores de la serie temporal referida (ver anexo), así como de su representación gráfica se observa que el IVAE en comidas, restaurante y hoteles tiene un comportamiento distinto en el dato 66 (junio de 1995); esta observación anómala que presenta la serie, se debió a que en esa fecha el porcentaje del IVA aumento (del 11% al 13%).



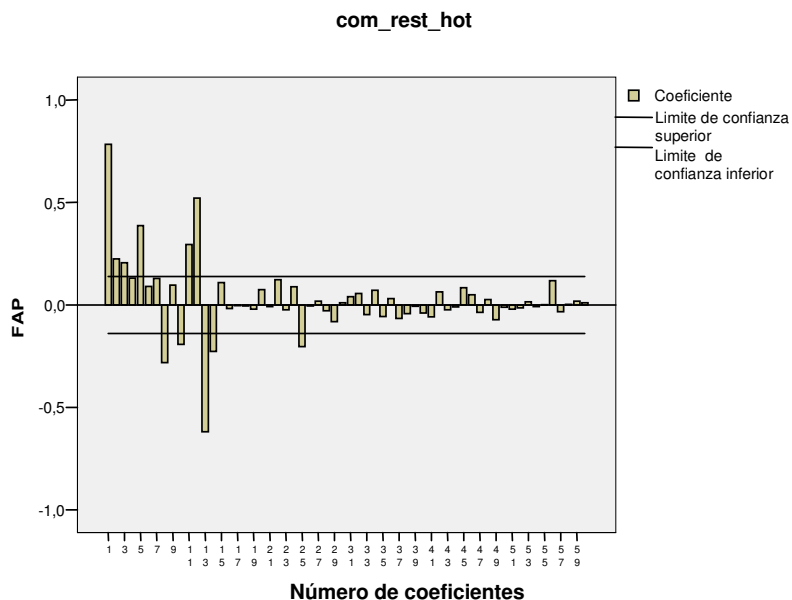
**Gráfico 7.1.1.2: Función de Autocorrelación Simple (FAS)**



**Gráfico 7.1.1.3: Función de Autocorrelación Simple Estacional**



**Gráfico 7.1.1.4: Función de Autocorrelación Parcial (FAP)**



Podemos observar que no se tiene estacionariedad en media tanto en la parte regular como en la estacional; por que no se mantiene el supuesto de que los datos varían alrededor de una única media; esto se confirma en el gráfico 7.1.1.2 y 7.1.13, de las funciones de Autocorrelación (FAS regular y estacional), ya que decaen lentamente a lo largo del tiempo, razón que nos lleva a tener que diferenciar regular y estacionalmente, ya que se detecta el componente estacional.

Antes de profundizar en el análisis, se exploraran los estadísticos descriptivos; para tener un bosquejo general del comportamiento de esta serie.

**Tabla 7.1.1.1: Estadísticos Descriptivos**

		Estadístico	Error Std.
com_rest_hot	Media	172,7371	2,38461
	Mediana	179,0150	
	Varianza	1171,392	
	Desviación Std.	34,22560	
	Asimetría	-,450	,169
	Curtosis	,119	,337

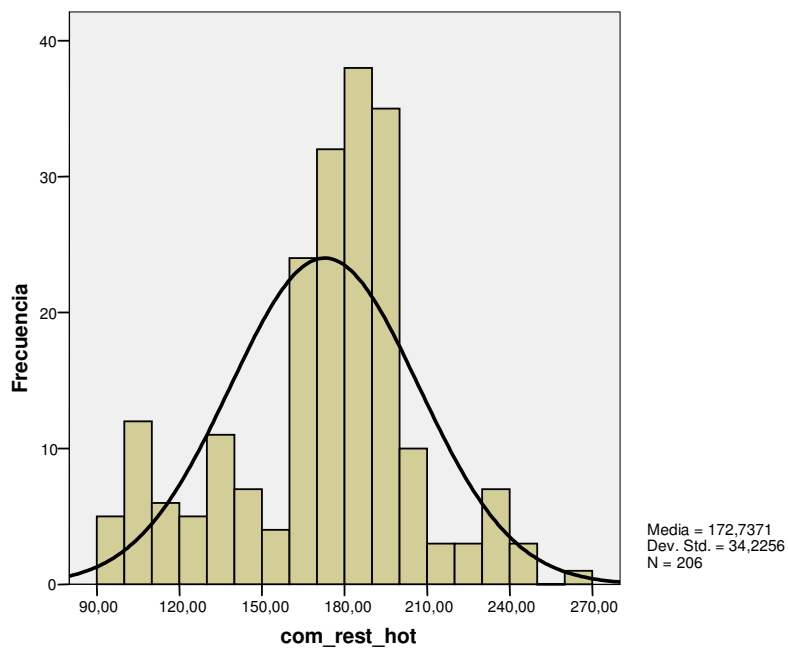
Analizando la desviación típica podemos decir que la variación esperada con respecto de la media y cada dato es 34.2256.

En nuestro caso la dispersión que hay entre cada uno de los datos con respecto a la media es de: 1171.392 (la varianza).

La Asimetría (Skewness) esto nos indica que los datos tienen los valores más extremos inferiores a la media, el valor absoluto de la curtosis es menor a uno, lo que indica una diferencia con la distribución normal.

El gráfico 7.1.1.5, nos afirma todo lo visto antes, con los resultados numéricos. Obsérvese que los datos en estudio no, presentan una semejanza a la distribución normal, se tiene una evidente diferenciación en la cola izquierda del histograma, en comparación con el histograma de la distribución normal en esta cola se tienen acumulados más datos.

**Gráfico 7.1.1.5: Histograma de los datos originales**



**Tabla 7.1.1.2: Medias y Varianzas**

com\_rest\_hot

AÑOS	Media	Varianza	Desviacion Std.
1990	104,6783	106,091	10,30004
1991	112,1408	225,382	15,01274
1992	130,6367	191,987	13,85594
1993	146,9150	292,809	17,11168
1994	165,1258	496,567	22,28378
1995	181,2425	353,798	18,80952
1996	179,4650	500,337	22,36822
1997	182,8442	317,435	17,81670
1998	183,0092	360,781	18,99424
1999	186,3617	330,989	18,19309
2000	184,9792	264,463	16,26230
2001	186,4217	311,188	17,64052
2002	190,3058	255,447	15,98270
2003	195,1208	274,706	16,57425
2004	196,0167	334,946	18,30154
2005	199,6525	348,135	18,65838
2006	207,8458	454,772	21,32537
2007	195,3550	,490	,70004
Total	172,7371	1171,392	34,22560

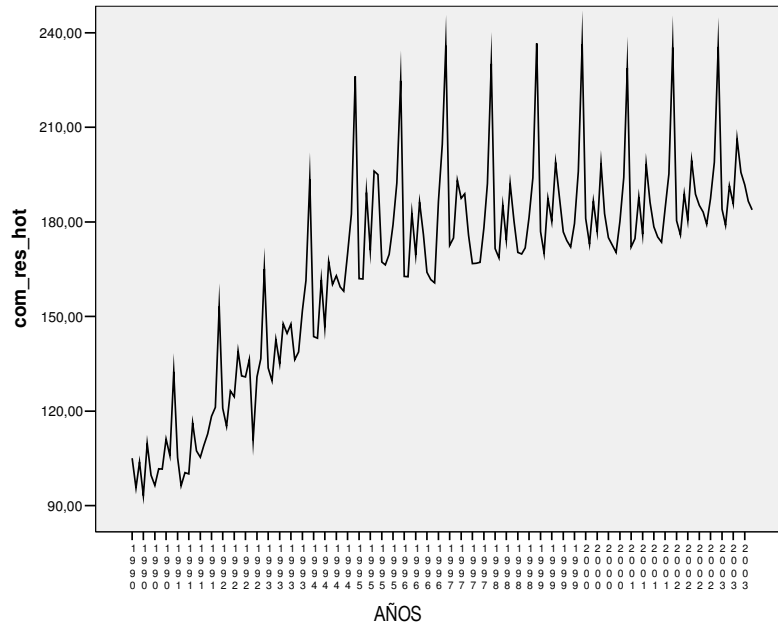
En la tabla 7.1.1.2, notamos que cada período (llamamos período a los 12 datos de cada año, siendo el período uno, los datos desde enero de 1990 hasta diciembre del mismo año.) tiene media distinta, entonces podemos decir que esta serie no tiene estacionariedad en media por lo que nos hace pensar que hay que transformarla.

Con estos indicadores del análisis exploratorio; pasamos al análisis ARIMA.

## IDENTIFICACIÓN DE LOS POSIBLES MODELOS

Para iniciar la identificación del modelo ARIMA, esclarecemos que se ha dejado el 5% de los datos, para efectos de comparación con los pronósticos generados a partir del modelo identificado. De ahora en adelante la serie consta de 196 observaciones, que corresponden al período de enero de 1990 – abril de 2006.

**Gráfico 7.1.1.6: Serie menos el 5%**



De aquí se puede asumir que las funciones de Autocorrelación y Autocorrelación parcial no serán un instrumento adecuado para identificar el proceso ARIMA generador de la serie por que estarán distorsionadas por la intervención que se detecta a partir de la observación antes detallada.

Como se ha mencionado en el marco teórico que las dos características de los modelos de intervención que se deben identificar son: el período de comienzo de las intervenciones y la forma general del impacto de las mismas. Pues bien, con respecto al período de comienzo de la intervención, ya hemos señalado que éste será junio de 1995, es decir  $t_0 = 66$ ; en cuanto a la forma general del impacto, supondremos que la intervención es de comienzo gradual y duración permanente tomando en cuenta un período de desfase temporal para el efecto (es decir,  $b=1$ ).

En consecuencia con lo anterior el modelo que se puede especificar para la intervención es:

$$X_t = \frac{\omega L}{1 - \delta L} S_t^{(t_0=66)}$$

Siendo  $S_t$  la variable escalón definida como:

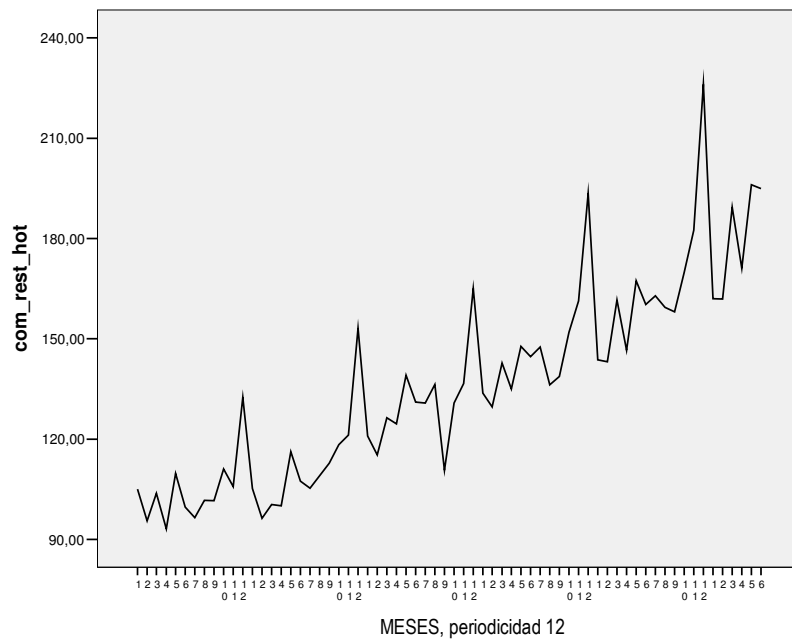
$$S_t = \begin{cases} 1 & \text{para } t \geq t_0 = 66 \\ 0 & \text{para } t < t_0 = 66 \end{cases}$$

O lo que es igual:

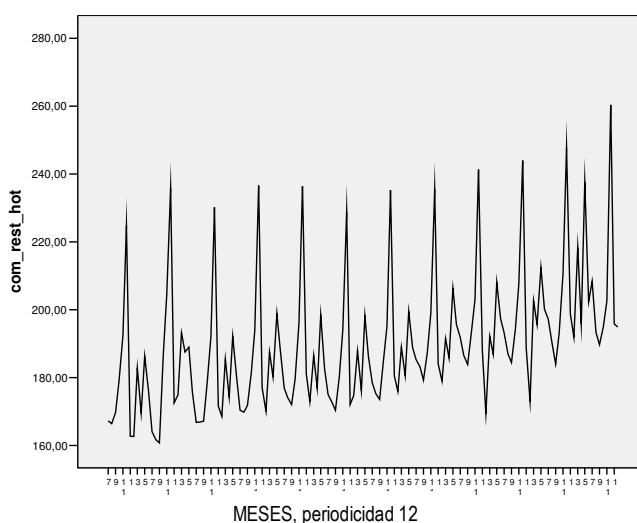
$$X_t = \frac{\omega}{1 - \delta L} S_t^{(t_0=67)}$$

Para identificar el modelo ARIMA que más se adecua a los datos, utilizaremos el primer método detallado en la metodología, el cual consiste en tomar el subconjunto de datos anterior o posterior a la intervención siempre que sea suficientemente representativo (mayor que el 60%). En este caso el porcentaje de datos antes de la intervención es el 32.03% de la serie completa y los después de la intervención representan el 67.96%.

**Gráfico 7.1.1.7: Serie del subconjunto de datos antes de la intervención**



**Gráfico 7.1.1.8: Serie con el subconjunto de datos después de la intervención**



En esta serie la mayor cantidad de observaciones se encuentran después de la intervención, se tomará entonces este subconjunto de datos para identificar el modelo ARIMA, los datos están comprendidos entre agosto 1995 y marzo 2006, dejando 10 observaciones de esta nueva serie, para efectos de comparación con las predicciones del mejor modelo ARIMA que se identifique.

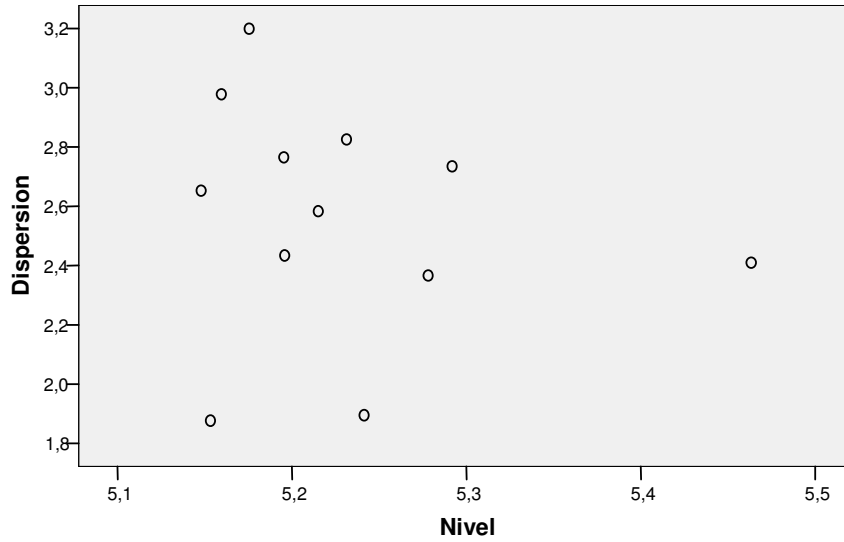
Observando el gráfico 7.1.1.8, notamos que presenta estacionariedad en varianza, por lo que podemos asumir que la serie no necesitará transformación., ya que no se detecta mucha dispersión de los datos; aunque aseguraremos analizando el test de Levene a continuación.

**Tabla 7.1.1.3: Estadístico de Levene**

Transformación	Estadístico, de Levene basado en la media	Sig
Sin transformación	0.916	0.527
Logaritmo Natural	1.314	0.225
Raíz cuadrada inversa	1.557	0.121
Raíz cuadrada	1.100	0.367

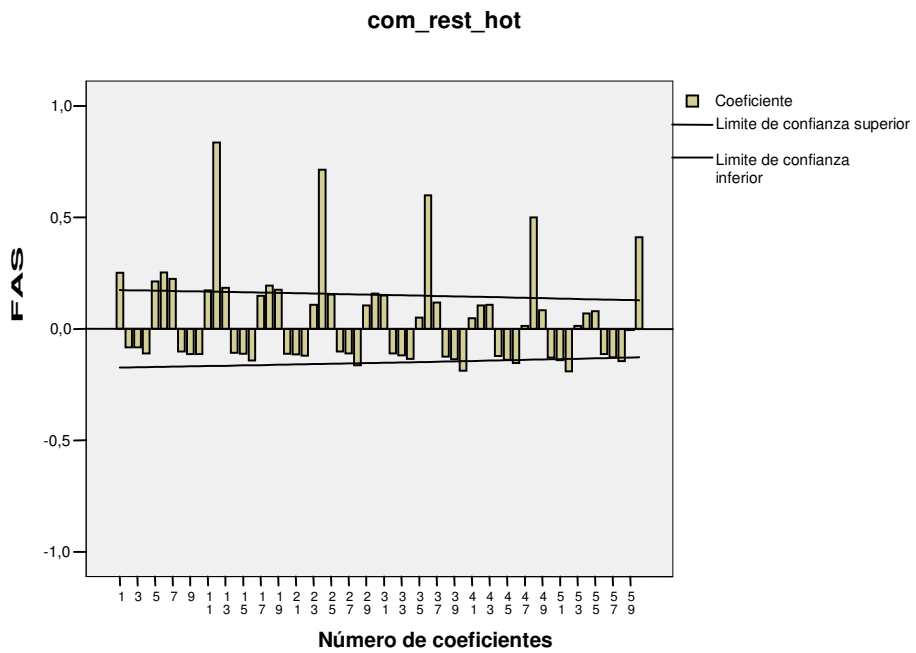
Según el estadístico de Levene y la significancia mostrada en la tabla 7.1.1.3, se concluye que no se tendrá que hacer ninguna transformación en la serie por que al observar el siguiente gráfico rango/media es plausible determinar que la variabilidad de la serie se mantiene constante a lo largo del tiempo y que además se tiene el Menor estadístico de Levene y mayor significancia en comparación con transformaciones como: Logaritmo Natural, Raíz Cuadrada Inversa, Raíz cuadrada.

**Gráfico 7.1.1.9: Rango-Media**



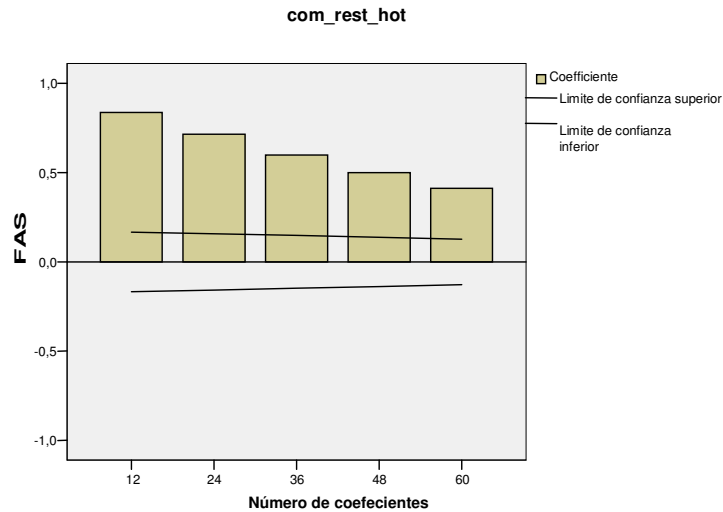
Slope = -,756 Power for transformation = 1,756

**Gráfico 7.1.1.10: Función de Autocorrelación Simple (FAS), parte regular.**



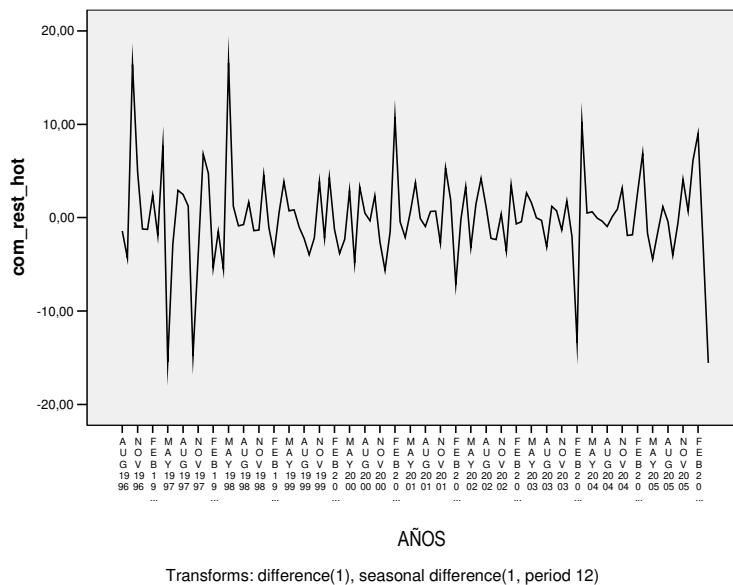


**Gráfico 7.1.1.11: Función de Autocorrelación Simple (FAS), parte estacional.**



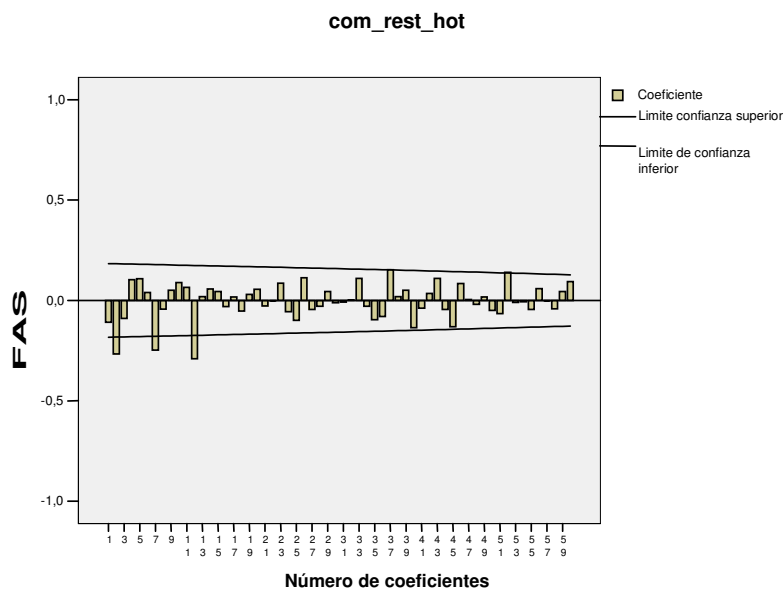
Interpretando los gráficos de autocorrelaciones, identificamos la no estacionariedad en media y se debe diferenciar la serie en la parte regular y estacional, permitiendo obtener el gráfico de secuencia (gráfico 7.1.1.12) y las funciones de Autocorrelación siguientes (gráficos 7.1.1.13-7.1.1.16):

**Gráfico 7.1.1.12: Serie Diferenciada**

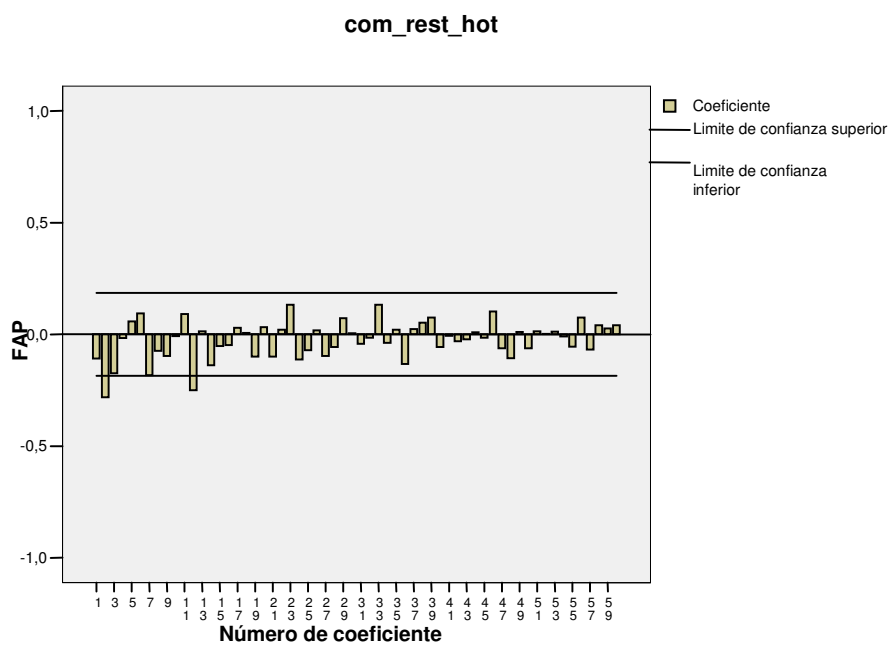


Ahora sí; la serie ya es estacionaria en media y en varianza. Por lo que ya se pueden analizar los gráficos de autocorrelación simple y parcial, con los cuales se plantearán los posibles modelos que se adecuan a los datos. Veamos estos gráficos:

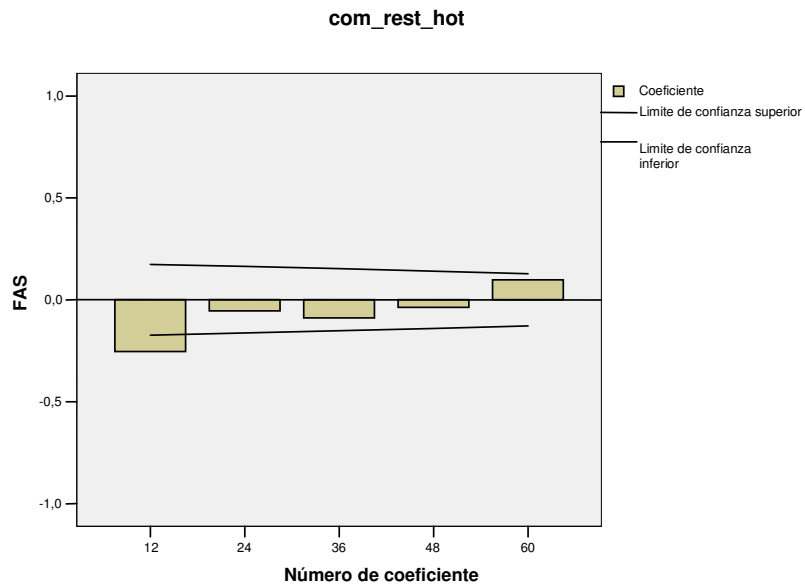
**Gráfico 7.1.1.13: FAS regular diferenciada**



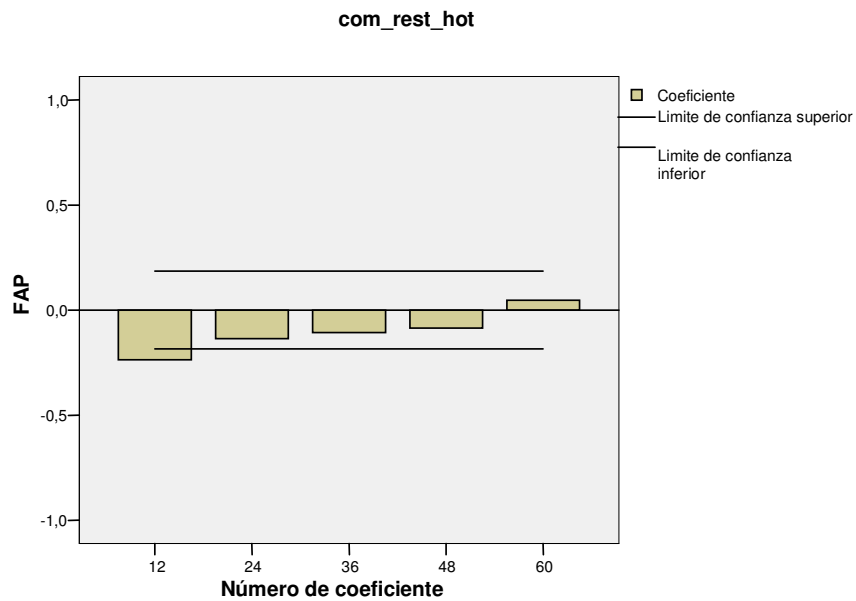
**Gráfico 7.1.1.14: FAP regular diferenciada**



**Gráfico 7.1.1.15: FAS estacional**



**Gráfico 7.1.1.16: FAP estacional diferenciada**



Los modelos que se planteen, posteriormente se examinarán en la etapa de Estimación, en el proceso algunos no cumplirán con los criterios de las fases y hasta ahí se analizarán, es decir, que no pasarán a la siguiente etapa, llegando al final con un solo modelo, esto permitirá tomar la decisión de elegir el **MODELO ADECUADO**.

De la FAS y FAP, los modelos identificados son:

- 1-  $IMA(1,2) \times I(1)_{12}$   $(1-L)(1-L^2)y_t = (1-\theta_1L - \theta_2L^2)u_t$
- 2-  $ARI(2,1) \times I(1)_{12}$   $(1-\phi_1L - \phi_2L^2)(1-L)(1-L^2)y_t = u_t$
- 3-  $IMA(1,2) \times IMA(1,1)_{12}$   $(1-L)(1-L^2)y_t = (1-\theta_1L - \theta_2L^2)(1-\Theta_1L^2)u_t$
- 4-  $ARI(2,1) \times ARI(1,1)_{12}$   $(1-\phi_1L - \phi_2L^2)(1-\Phi_1L^2)(1-L)(1-L^2)y_t = u_t$
- 5-  $ARIMA(2,1,2) \times I(1)_{12}$   $(1-\phi_1L - \phi_2L^2)(1-L)(1-L^2)y_t = (1-\theta_1L - \theta_2L^2)u_t$
- 6-  $ARIMA(2,1,2) \times ARIMA(1,1,1)_{12}$   
 $(1-\phi_1L - \phi_2L^2)(1-\Phi_1L^2)(1-L)(1-L^2)y_t = (1-\theta_1L - \theta_2L^2)(1-\Theta_1L)u_t$

### ESTIMACIÓN DE LOS PARÁMETROS

Usando el paquete estadístico SPSS 13.0 se obtiene la estimación de coeficientes de los modelos identificados, mediante el método de predicción de Mínimos Cuadrados Incondicional, con valores iniciales automáticos para la estimación, los resultados son los siguientes:

- 1-  $IMA(1,2) \times I(1)_{12}$   $(1-L)(1-L^2)y_t = (1-0.266L - 0.370L^2)u_t$
- 2-  $ARI(2,1) \times I(1)_{12}$   $(1+0.337L^2)(1-L)(1-L^2)y_t = u_t$  con  $\phi_1 = 0$
- 3-  $IMA(1,2) \times IMA(1,1)_{12}$   $(1-L)(1-L^2)y_t = (1-0.369L - 0.431L^2)(1-0.493L^2)u_t$
- 4-  $ARI(2,1) \times ARI(1,1)_{12}$   
 $(1+0.364L^2)(1+0.398L^2)(1-L)(1-L^2)\sqrt{y_t} = u_t$  con  $\theta_1 = 0$
- 5-  $ARIMA(2,1,2) \times I(1)_{12}$   
 $(1-0.559L + .800L^2)(1-L)(1-L^2)y_t = (1-0.725L + 0.620L^2)u_t$
- 6-  $ARIMA(2,1,2) \times ARIMA(1,1,1)_{12}$   
 $(1-L)(1-L^2)y_t = (1-0.673L)u_t$  con  $\phi_1 = \phi_2 = \theta_1 = \theta_2 = \Phi_1 = 0$

Se puede observar que el último modelo se descarta, de ser el mejor, ya que de sus seis parámetros solamente uno es significativo.

Ver corridas en anexo.

**Tabla 7.1.1.4: VERIFICACIÓN DE COEFICIENTES PARA ESTACIONARIEDAD E INVERTIBILIDAD**

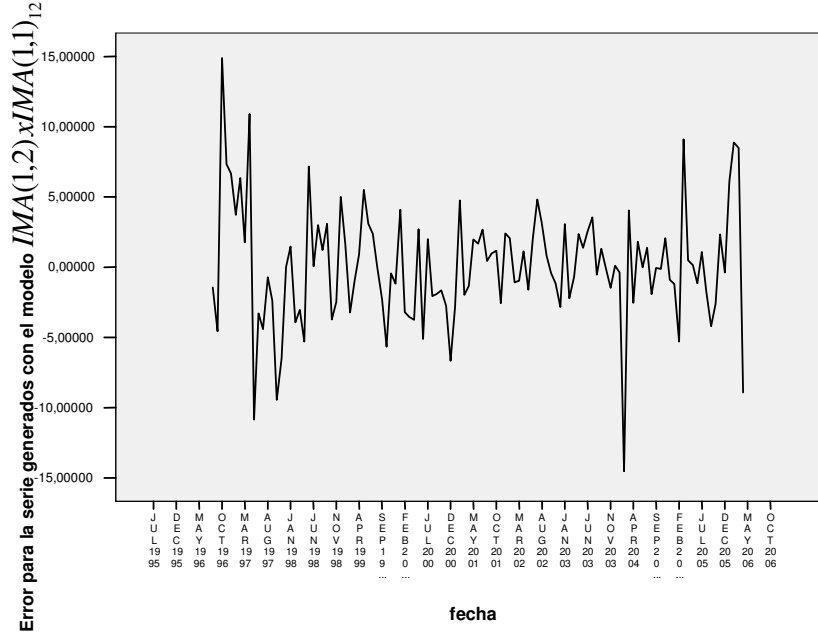
MODELO	ESTACIONARIEDAD	INVERTIBILIDAD
$IMA(1,2) \times I(1)_{12}$	El modelo es un MA y una de las características de estos modelos es que siempre es Estacionario	Debe cumplirse que $\theta_1 + \theta_2 < 1$ , $\theta_2 - \theta_1 < 1$ y $ \theta_2  < 1$ , siendo, $\theta_1 = 0.266$ y $\theta_2 = 0.370 < 1$ . El modelo es invertible.
$ARI(2,1) \times I(1)_{12}$	Debe cumplirse que $\phi_1 + \phi_2 < 1$ , $\phi_2 - \phi_1 < 1$ y $ \phi_2  < 1$ , siendo $\phi_1 = 0$ y $\phi_2 = -0.337$ . El modelo es estacionario.	El modelo es un AR y una de las características de estos modelos es que siempre es invertible
$IMA(1,2) \times IMA(1,1)_{12}$	El modelo es un MA y una de las características de estos modelos es que siempre es Estacionario	Debe cumplirse que $\theta_1 - \theta_2 < 1$ , $\theta_2 - \theta_1 < 1$ , $ \theta_2  < 1$ , y $ \Theta_1  < 1$ siendo, $\theta_1 = 0.369$ , $\theta_2 = 0.431$ y $\Theta_1 = 0.493 < 1$ . El modelo es invertible.
$ARI(2,1) \times ARI(1,1)_{12}$	Debe cumplirse que $\phi_1 + \phi_2 < 1$ , $\phi_2 - \phi_1 < 1$ , $ \phi_2  < 1$ , y $ \Phi_1  < 1$ , siendo $\phi_1 = 0$ , $\phi_2 = -0.398$ , $\Phi_1 = -0.364$ El modelo es estacionario.	El modelo es un AR y una de las características de estos modelos es que siempre es invertible
$ARIMA(2,1,2) \times I(1)_{12}$	Debe cumplirse que $\phi_1 + \phi_2 < 1$ , $\phi_2 - \phi_1 < 1$ y $ \phi_2  < 1$ , siendo $\phi_1 = 0.559$ , $\phi_2 = -0.800$ , . El modelo NO es estacionario.	Debe cumplirse que $\theta_1 + \theta_2 < 1$ , $\theta_2 - \theta_1 < 1$ y $ \theta_2  < 1$ , siendo $\theta_1 = 0.725$ , $\theta_2 = -0.62$ . El modelo es invertible.

Cuatro de los modelos cumplen estas condiciones, y pasan a la siguiente etapa, el modelo que se deja de analizar desde aquí es  $ARIMA(2,1,2) \times I(1)_{12}$ ; porque no cumple la condición de estacionariedad. Para poder elegir el mejor modelo se hace necesario el análisis de los residuos, esto es, verificar si se comportan como un ruido blanco (es decir se tienen un comportamiento normal).

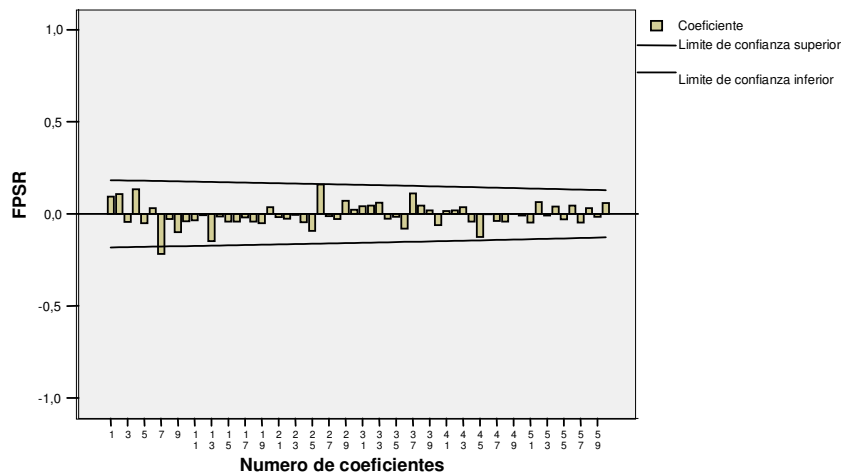
## ANÁLISIS DE LOS RESIDUOS

Sometiendo los modelos al análisis de los residuos en todos se tiene media cero, la varianza constante, esta condición tiene mayor exactitud para el modelo  $IMA(1,2) * IMA(1,1)_{12}$ , de igual manera la condición de incorrelación. A continuación se verifica observando los gráficos de secuencia y las autocorrelaciones residuales de dicho modelo.

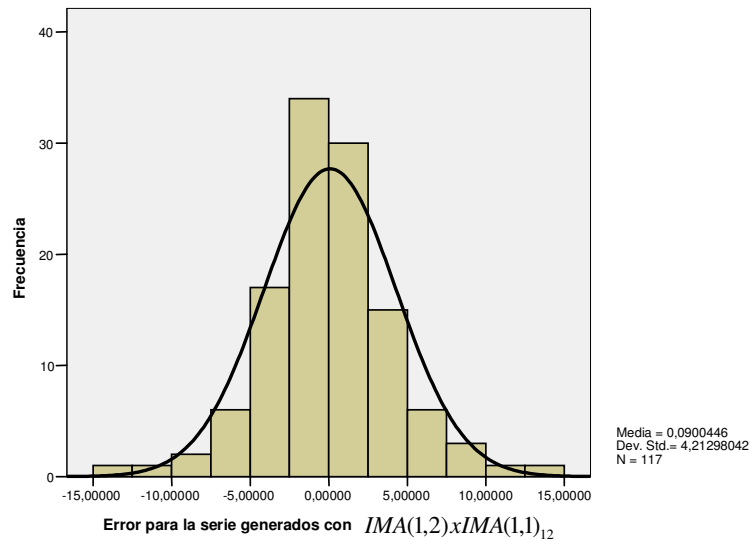
**Gráfico 7.1.1.17: Residuos con el modelo  $IMA(1,2) * IMA(1,1)_{12}$**



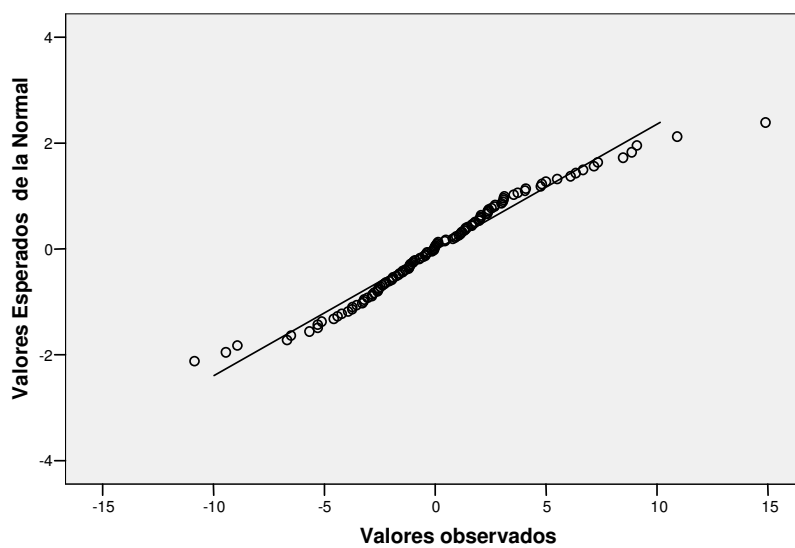
**Gráfico 7.1.1.18: FAPR con el modelo  $IMA(1,2) * IMA(1,1)_{12}$**



**Gráfico 7.1.1.19: Histograma Residual**



**Gráfico 7.1.1.20: Probabilidad Residual**



**Tabla 7.1.1.5: Test de normalidad**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	df	Sig.	Estadístico	df	Sig.
Error para com_rest_hot Generado con $IMA(1,2)xIMA(1,1)_{12}$	0,082	117	0,050	0,971	117	0,012

a. Lilliefors Significance Correction

Al observar el gráfico 7.1.1.17, notamos que los residuos de este modelo tiene un comportamiento constante a lo largo del tiempo, además en el gráfico 7.1.1.18, mas del 95% de los retardos no son significativos, aquí se cumple la incorrelación de los residuos.

Con el gráfico 7.1.1.19 y 7.1.1.20, verificamos que los residuos tiene un comportamiento normal, y así certificamos que los residuos de este modelo se comportan como un ruido blanco.

Al analizar los modelos con los estadísticos AIC (El Criterio Akaike) y BIC (El criterio de información Bayesiana); permite que se tenga un criterio más para identificar el modelo generador de la serie, ya que se ha dicho que el modelo adecuado será aquel que cumple mayores condiciones y criterios.

**Tabla 7.1.1.6: Comparación de los estadísticos AIC y BIC**

Modelo	AIC	BIC
$ARI(2,1) \times I(1)_{12}$	685.777	694.063
$IMA(1,2) \times IMA(1,1)_{12}$	684.475	697.761
$IMA(1,2) \times IMA(1,1)_{12}$	669.336	680.385
$ARI(2,1) \times ARI(1,1)_{12}$	677.931	688.980

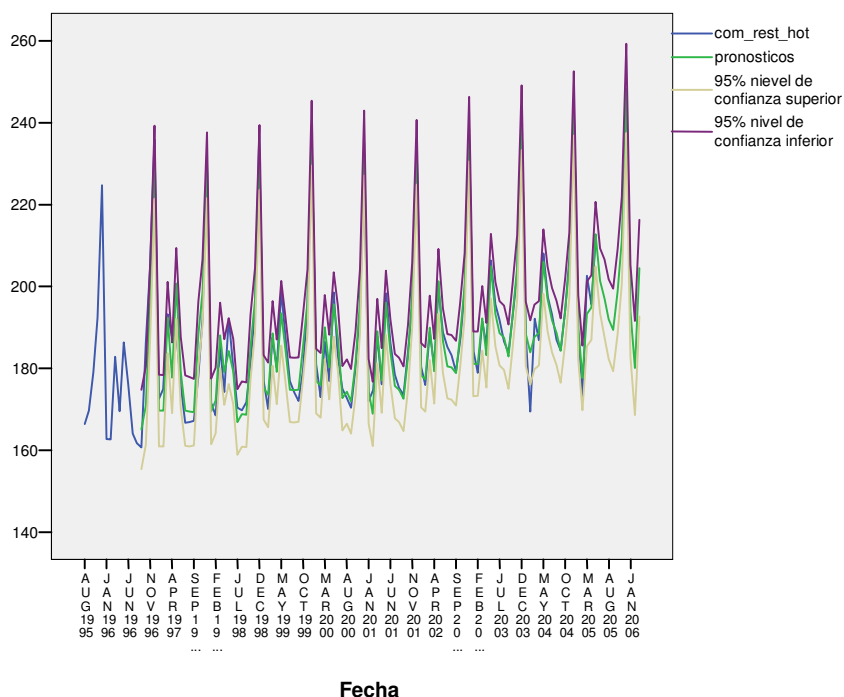
Para nuestro caso el modelo que cumple con esta condición, es el que ya se ha identificado en el análisis de residuo; por lo tanto estaríamos aceptando al modelo  $IMA(1,2) \times IMA(1,1)_{12}$ , como el que más se ajusta al comportamiento de los datos, pues cumple con la mayoría de los criterios para seleccionar el modelo.

### PREDICCIONES

A continuación se verificara si los pronósticos que genera este modelo, resultan adecuados, se pronostican 10 datos, desde junio del 2005 hasta marzo del 2006 y se comparan con los datos que se han dejado para este efecto.



**Gráfico 7.1.1.21: Superpuesto de datos reales y datos generados con el modelo  $IMA(1,2) \times IMA(1,1)_{12}$**



**Tabla 7.1.1.7: Pronósticos con el modelo  $IMA(1,2) \times IMA(1,1)_{12}$**

fecha	datos reales	pronósticos	error	límite superior 95%	límite inferior 95%
Jun-05	200,21	201.42278	-1,21	193.54605	209.29951
Jul-05	197,14	197.0751	0,07	187.60193	206.54826
Ago-05	190,55	192.00093	-1,45	182.23081	201.77104
Sep-05	183,94	189.37946	-5.43	179.3147	199.44422
Oct-05	193,11	198.82006	-5.71	188.46286	209.17725
Nov-05	210,62	211.20649	-0.58	200.55889	221.85409
Dic-05	247,73	248.36373	-0,63	237.42758	259.29987
Ene-06	198,64	194.01076	4.63	182.7878	205.23371
Feb-06	191,86	180.08412	11.78	168.57594	191.5923
Mar-06	218,38	204.47196	13.91	192.68003	216.26388

Observando el gráfico de los pronósticos junto con los datos reales, nuestro modelo sí se ajusta adecuadamente al subconjunto de datos analizados, numéricamente los pronósticos no se alejan mucho de los datos reales.

Debido a que nuestra serie, tiene comportamiento paulatinamente distinto en los dos subconjuntos truncados por la intervención, necesitamos conocer si este modelo es el generador de toda la serie, para ello analizamos, la etapa de chequeo y las predicciones del modelo  $IMA(1,2) \times IMA(1,1)_{12}$ , con las observaciones comprendidas desde enero 1990- marzo 2006.

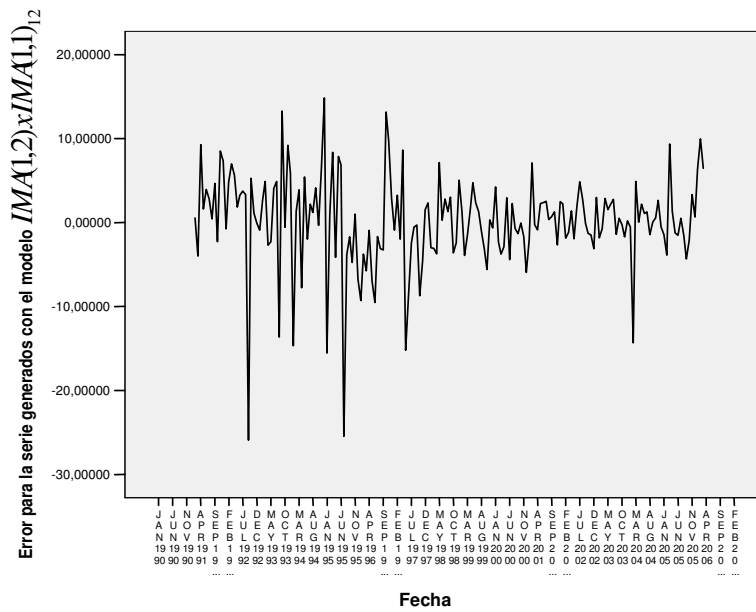
Estimación de los parámetros:

$$IMA(1,2) \times IMA(1,1)_{12} \quad (1-L)(1-L^{12})y_t = (1-0.336L-0.25L^2)(1-0.42L^{12})u_t$$

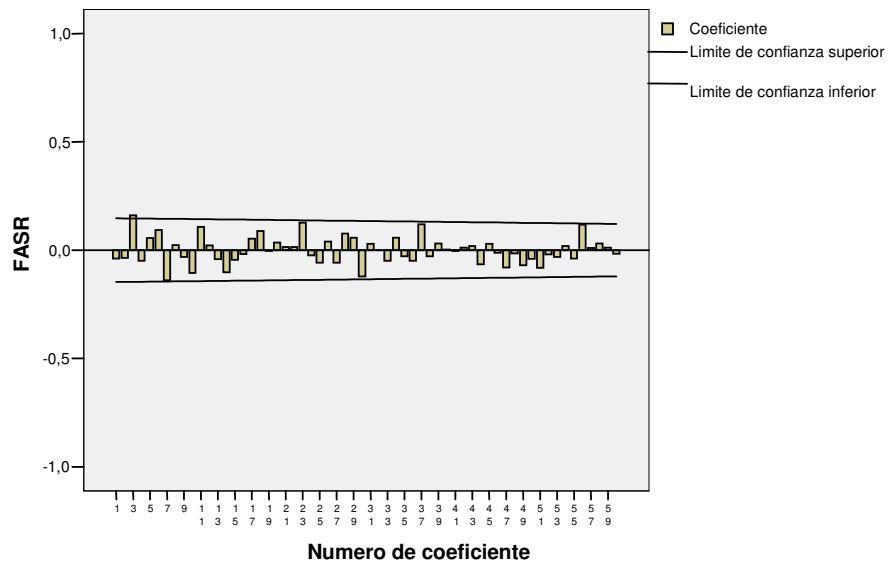
$$\theta_1 - \theta_2 < 1, \theta_2 - \theta_1 < 1, |\theta_2| < 1, \text{ y } |\Theta_1| < 1 \text{ siendo, } \theta_1 = 0.336, \theta_2 = 0.25 \text{ y } \Theta_1 = 0.42 < 1.$$

Cumple las condiciones de estacionariedad e invertibilidad.

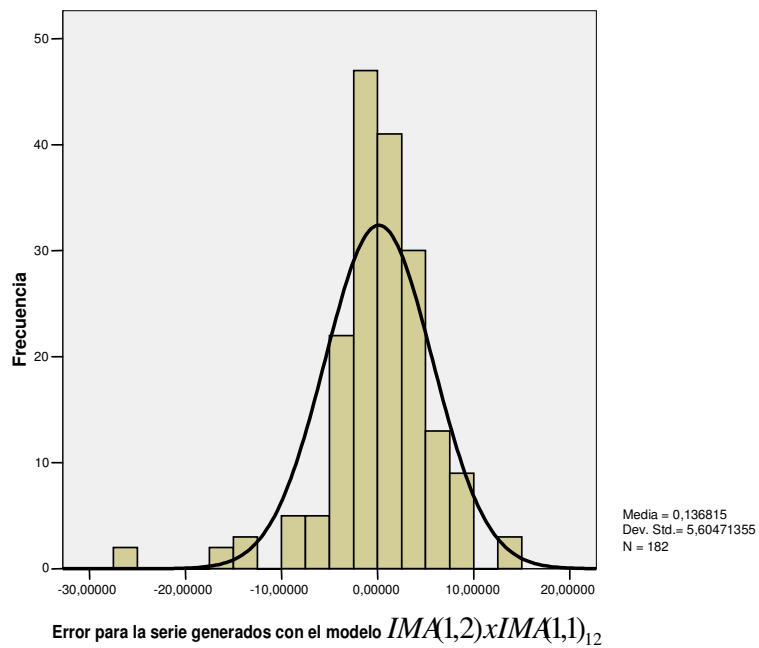
**Gráfico 7.1.1.22: Residuos de la Serie de Enero de 1990 a marzo de 2006**



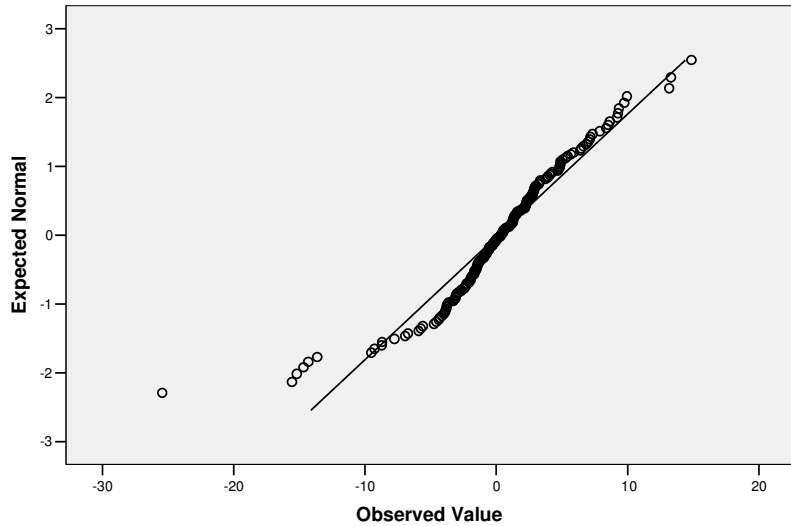
**Gráfico 7.1.1.23: FASR de la Serie de Enero de 1990 a marzo de 2006**



**Gráfico7.1.1.24: Histograma Residual**



**Gráfico7.1.1.25: Probabilidad de los residuos**



**Tabla 7.1.1.8: Test de Normalidad**

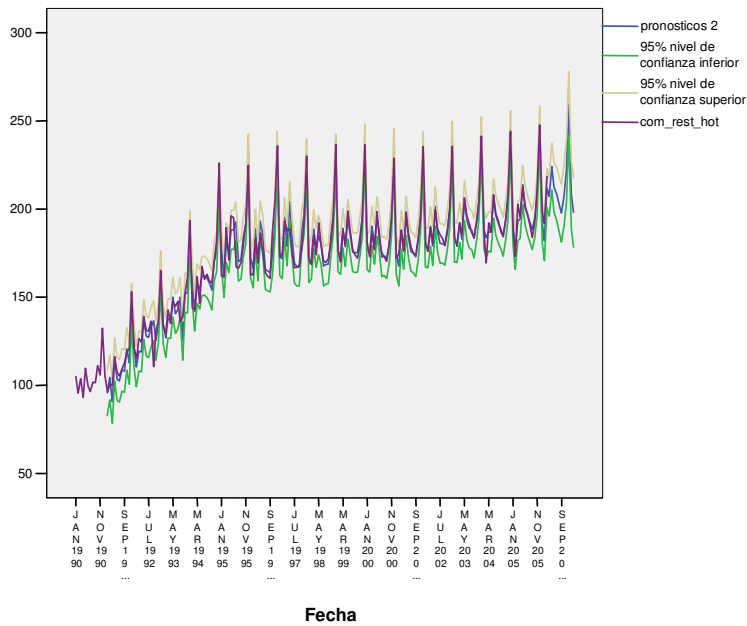
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	df	Sig.	Estadístico	df	Sig.
Error para com_rest_hot	,109	182	,000	,910	182	,000

a. Lilliefors Significance Correction

El modelo ha cumplido las condiciones del análisis de residuos; es decir pasa la condición de normalidad y por el gráfico 7.1.1.23, se dice que están incorrelados.

Verificamos ahora las predicciones:

**Gráfico 7.1.1.26: Superpuesto de datos reales y datos generados con el modelo identificado.**



**Tabla 7.1.1.9: Valores de Pronósticos**

fecha	datos reales	pronósticos	Error	limite inferior	limite superior
Abr-06	195,89	207.12064	-11.23	196.00218	218.2391
May-06	237,63	224.02664	13.6	210.64151	237.41176
Jun-06	202,36	212.34616	-9.98	198.15245	226.53987
Jul-06	208,41	208.72303	0.31	193.75559	223.69047
Ago-06	193,29	202.68563	-9.68	186.97417	218.39708
Sep-06	189,74	197.726	-7.98	181.2962	214.15581
Oct-06	195,03	206.99813	-11.96	189.87243	224.12384
Nov-06	202,55	222.35852	-19.8	204.55671	240.16032
Dic-06	260,37	259.44332	0.93	240.98305	277.90359
Ene-07	195,85	208.11434	-12.26	189.01141	227.21726
Feb-07	194,86	198.12757	-3.26	178.39625	217.85889

Observando el gráfico de los pronósticos junto con los datos reales, notamos que nuestro modelo sí se ajusta adecuadamente a los datos de la serie en estudio, numéricamente los pronósticos no se alejan mucho de los datos reales, con esto confirmamos que el modelo  $IMA(1,2) \times IMA(1,1)_{12}$  es el que mejor se ajusta a la ST del IVAE en comidas, restaurantes y hoteles.

## ANÁLISIS DE INTERVENCIÓN

Una vez que se ha identificado el mejor modelo ARIMA para el mayor subconjunto de datos; se incorpora el modelo de la intervención identificado en el análisis descriptivo de la serie al modelo ARIMA encontrado.

El modelo que se puede especificar es:

$$y_t = \frac{\omega}{1 - \delta L} S_t^{(t_0=67)} + \frac{(1 - \theta_1 L - \theta_2 L^2)(1 - \Theta_1 L^{12})}{(1 - L^{12})(1 - L)}$$

El análisis de los modelos de intervención se puede limitar a dos ya que observando el gráfico 7.1.1.1 claramente se notamos que la duración es permanente, por ello los dos modelos entre los que podría encontrarse el adecuado son: comiendo brusco y duración permanente; y comienzo gradual y duración permanente.

Al introducir este modelo en el programa STATISTICA 7.0, para la estimación de parámetros, y análisis de los residuos, se obtienen los resultados siguientes:

### COMIENZO GRADUAL Y DURACIÓN PERMANENTE

**Tabla 7.1.1.10: Estimación de los parámetros:**

Paramet.	Input: comidas, restaurantes y hoteles Transformaciones: D(1),D(1) (Interrupted ARIMA) Model:(0,1,2)(0,1,1) Seasonal lag: 12 MS Residual= 178,70							
	Param.	Asympt. Std.Err.	Asympt. t( 188)	p	inferior 95% Conf	superior 95% Conf	Interv. Case No.	Tipo Interv.
Constant	0.02404	0.048640	0.4942	0.621710	-0.0719	0.119990		
q(1)	1.49091	0.071941	20.7242	0.000000	1.3490	1.632825		
q(2)	-0.51892	0.073024	-7.1061	0.000000	-0.6630	-0.374864		
Qs(1)	-0.72069	0.039067	-18.4475	0.000000	-0.7978	-0.643625		
Omega(1)	-3.21153	4.818114	-0.6666	0.505874	-12.7160	6.292979	67	Grd/Perm
Delta(1)	-0.99379	0.024363	-40.7910	0.000000	-1.0418	-0.945729	67	Grd/Perm

El parámetro Omega no se incluye al modelo, resulta ser no significativo, el valor de t es menor que dos; esto estaría dejando fuera del análisis al modelo, ya que uno de los parámetros que aportan en él, en este caso no lo hace, Esto nos lleva analizar otros modelos de Intervención para verificar cual será el mejor y poder pronosticar.

## COMIENZO BRUSCO Y DURACIÓN PERMANENTE

**Tabla 7.1.1.11: Estimación de los Parámetros:**

Input: comidas, restaurantes y hoteles Transformaciones: D(1),D(1) (Interrupted ARIMA) Model:(0,1,2)(0,1,1) Estacional lag: 12 MS Residual= 170,28								
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 189)	p	inferior 95% Conf	superior 95% Conf	N. caso Int.	Tipo Interv.
Constant	0.0243	0.049519	0.4907	0.624187	-0.0734	0.12198		
q(1)	1.5075	0.069125	21.8085	0.000000	1.3711	1.64386		
q(2)	-0.5364	0.069967	-7.6662	0.000000	-0.6744	-0.39836		
Qs(1)	-0.7520	0.037669	-19.9638	0.000000	-0.8263	-0.67770		
Omega(1)	-22.1202	7.056834	-3.1346	0.001996	-36.0405	-8.19997	67	Abr/Perm

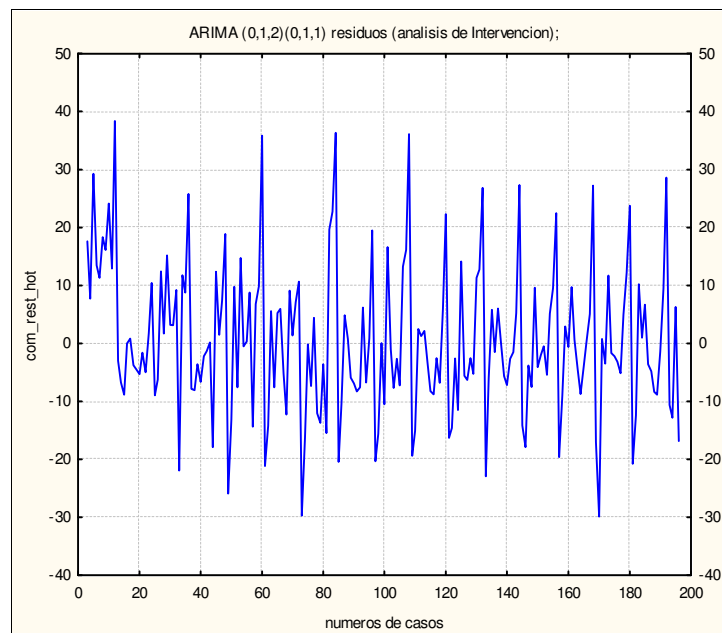
Obsérvese que todos los parámetros en excepción la constante aportan al modelo, por lo tanto este es un modelo candidato.

El modelo conjuntamente estimado sería entonces:

$$y_t = -22.1202S_t^{(t_0=67)} + \frac{(1 - 1.5075L + 0.5364L^2)(1 + 0.7520L^{12})}{(1 - L^2)(1 - L)}$$

Análisis de los Residuos del Modelo ARIMA con Intervención

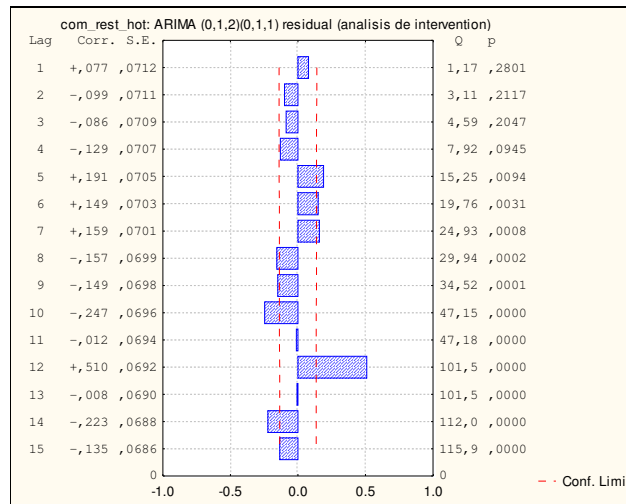
**Gráfico 7.1.1.27: residuos con modelo de intervención**



De este grafico podemos decir que los residuos cumplen con la condición de ser constantes en el tiempo.

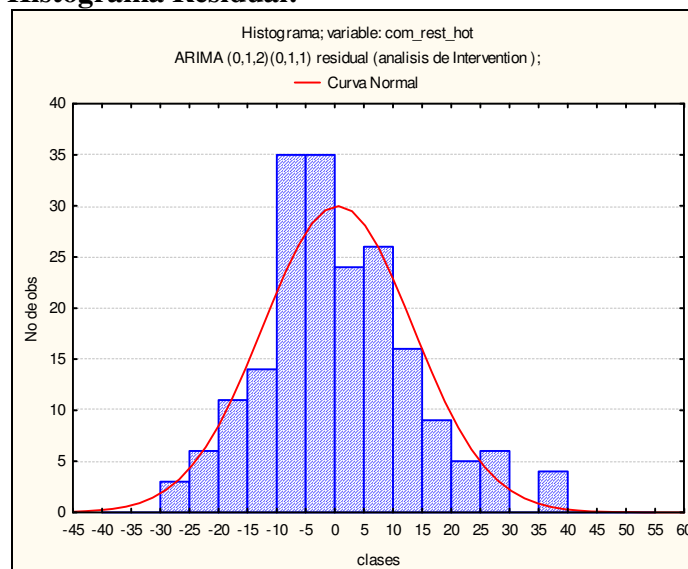
Para determinar la condición de si los residuos se comportan como un Ruido Blanco veamos los gráficos de autocorrelaciones siguientes:

**Gráfico 7.1.1.28: Función de Autocorrelación Simple Residual**



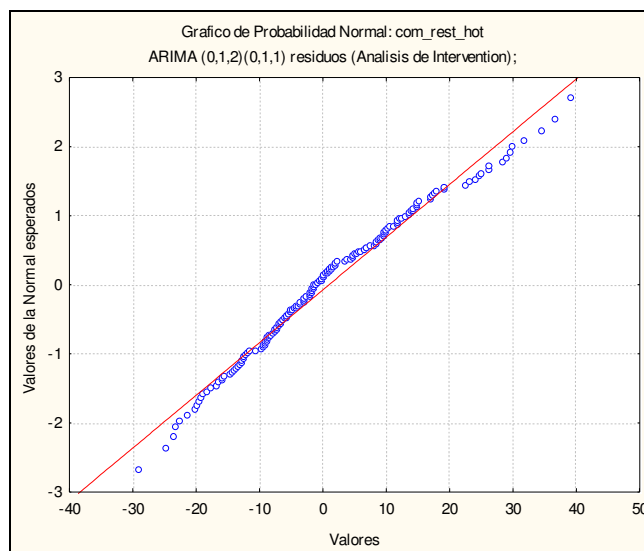
Observando el gráfico de autocorrelaciones, se afirma que la condición de incorrelación para los residuos es difícil asegurarla ya que hay más de un 5% de las observaciones fuera de los límites; para una afirmación más concreta analizamos el criterio de Normalidad con las representaciones siguientes:

**Gráfico 7.1.1.29: Histograma Residual:**





**Gráfico 7.1.1.30: Probabilidad Normal de los Residuos:**



**COMIENZO BRUSCO Y DURACIÓN TEMPORAL**

**Tabla 7.1.1.12: Estimación de los Parámetros:**

Input: variable com_rest_hotel								
Transformaciones: D(1),D(1) (Intervencion ARIMA)								
Model:(0,1,2)(0,1,1) Estacionall lag: 12 MS Residual= 178,42								
Paramet.	Param.	Std.Err.	t( 188)	p	superior 95% Conf	inferior 95% Conf	N de caso	Tipo Interv.
Constant	0.023979	0.048099	0.4985	0.618696	-0.07090	0.11886		
q(1)	1.473946	0.073209	20.1335	0.000000	1.32953	1.61836		
q(2)	-0.501226	0.074319	-6.7443	0.000000	-0.64783	-0.35462		
Qs(1)	-0.739093	0.040417	-18.2865	0.000000	-0.81882	-0.65936		
Omega(1)	4.072640	4.034927	1.0093	0.314106	-3.88691	12.03219	67	Bru/Temp
Delta(1)	-0.880371	0.128878	-6.8310	0.000000	-1.13460	-0.62614	67	Abr/Temp

El parámetro Omega no se incluye al modelo, resulta ser no significativo, el valor de t es menor que dos; esto estaría dejando fuera del análisis al modelo, ya que uno de los parámetros que aportan en él, en este caso no lo hace.

En nuestro caso en particular analizamos los tres tipos de intervención ya que ninguno de los anteriores cumple con las condiciones para ser el mejor modelo.

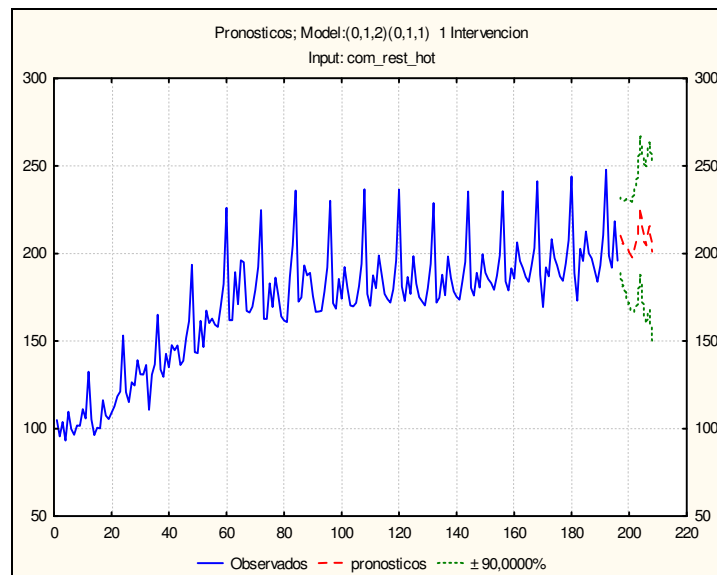
Ahora bien de los tres modelos que hemos interpretado NO cumplen exactamente todos los criterios de la parte de chequeo.

Escudriñando los modelos se detecta que el segundo modelo analizado (Comienzo brusco y duración Permanente), no cumple con el 100%, de los criterios entre estos la incorrelación; pero bien es el modelo con la mayoría de criterios aceptados. Por lo tanto el modelo de intervención que se considera el ADECUADO a los datos NO es el que se indico al inicio.

Señalamos que el tipo de intervención es Comienzo brusco y duración Permanente, ya que es el que mejor se adecua a los datos de comida, restaurantes y hoteles y que el efecto del cambio en el IVA, se mantiene en condiciones similares desde la fecha en que se aprobó; en vista de que el modelo de intervención que se ha incorporado en la modelización de la serie es permanente.

Ahora procedemos a las predicciones con este modelo. A continuación se presentan los pronósticos de este modelo acompañados con los datos originales:

**Gráfico 7.1.1.31: Superpuesto de los Datos Pronosticados y los Datos Originales:**



Al observar este gráfico, nos damos cuenta que los valores pronosticados tienen mucha similitud con el comportamiento de los datos originales.

**Tabla 7.1.1.13: Pronósticos con el modelo identificado**

fecha	datos reales	Pronósticos	error	Limite inferior	Limite superior
may-06	237,63	210,2397	27,39	189.2477	246.4804
jun-06	202,36	205,6624	-3,30	181.9129	245.1603
jul-06	208,41	204,2723	4,14	177.6963	247.0535
ago-06	193,29	200,6608	-7,37	171.1802	246.7507
sep-06	189,74	197,9236	-8,18	165.4960	247.3891
oct-06	195,03	200,8297	-5,80	165.3542	253.6825
nov-06	202,55	208,9062	-6,36	170.3354	265.2138
dic-06	260,37	227,1734	33,20	185.4599	287.0045
ene-07	195,85	208,9295	-13,08	163.9643	272.2918
feb-07	194,86	204,6271	-9,77	156.3151	271.5424

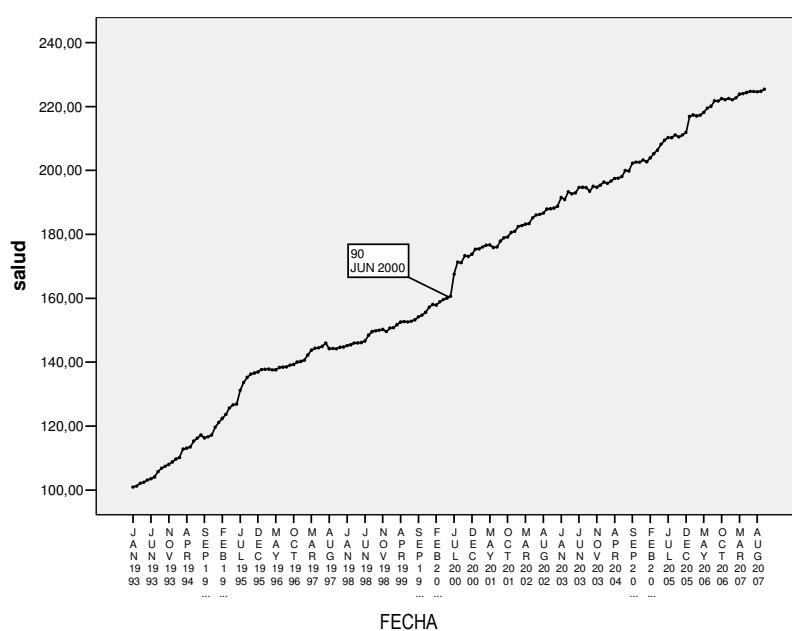
Si comparamos estos pronósticos con los obtenidos en el Análisis ARIMA, se cumple que los pronósticos con intervención son mejor que los del análisis sin intervención. Por lo tanto el modelo con intervención, se adecua mejor a los datos.

### 7.1.2 Serie Temporal del IPC en el Consumo Individual de Salud

#### ANÁLISIS DESCRIPTIVO DE LA SERIE

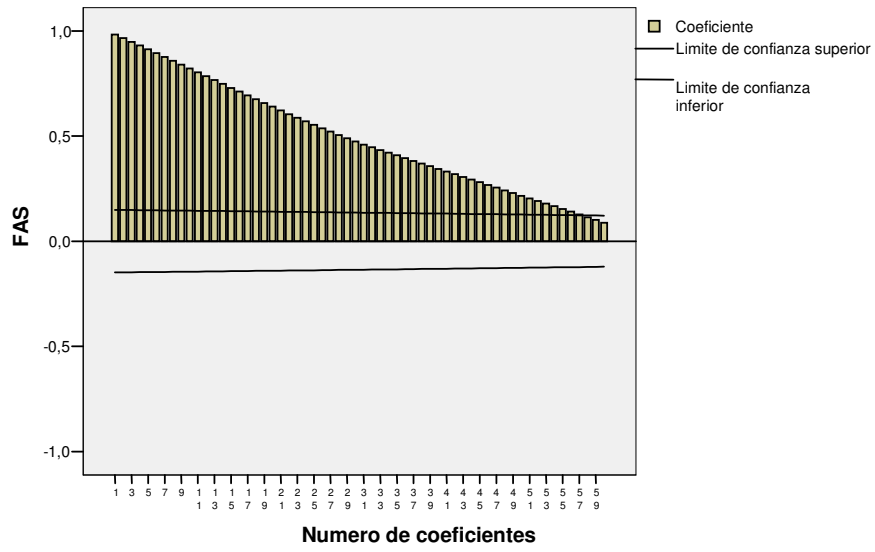
Índice de Precios al Consumidor (IPC), en la clasificación del consumo individual de SALUD. Se trata de una serie temporal de periodicidad mensual, que comprende 178 observaciones, las correspondientes al período enero 1993- octubre 2007, Iniciamos el análisis de la serie observando e interpretando el gráfico de secuencia de la serie original.

**Gráfico7.1.2.1: Serie Original**

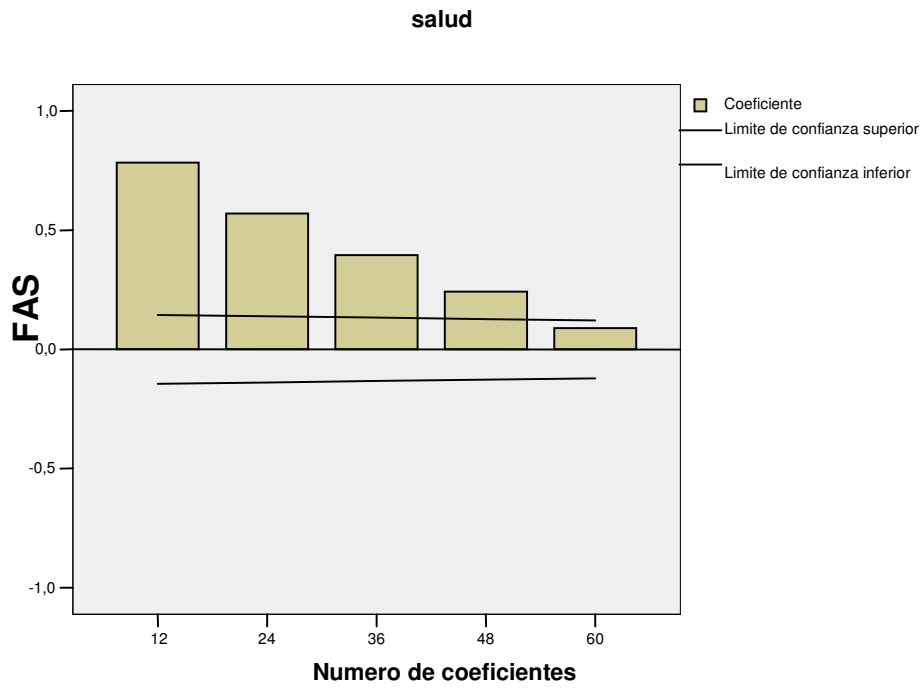


A partir de los valores de la serie temporal referida (ver anexo), así como de su representación gráfica se observa que existe un comportamiento bruscamente distinto a partir de Junio del 2000 ( $t_0=90$ ); los antecedentes que ocasionaron esta observación especial, surgen cuando se da un incremento del 5.1% a la actividad económica de los servicios en el país, incluyendo los servicios médicos. Debido a que conocemos el período y la forma general del impacto, analizaremos esta base de datos con Análisis de intervención, para obtener pronósticos interpretando este cambio de comportamiento en la serie.

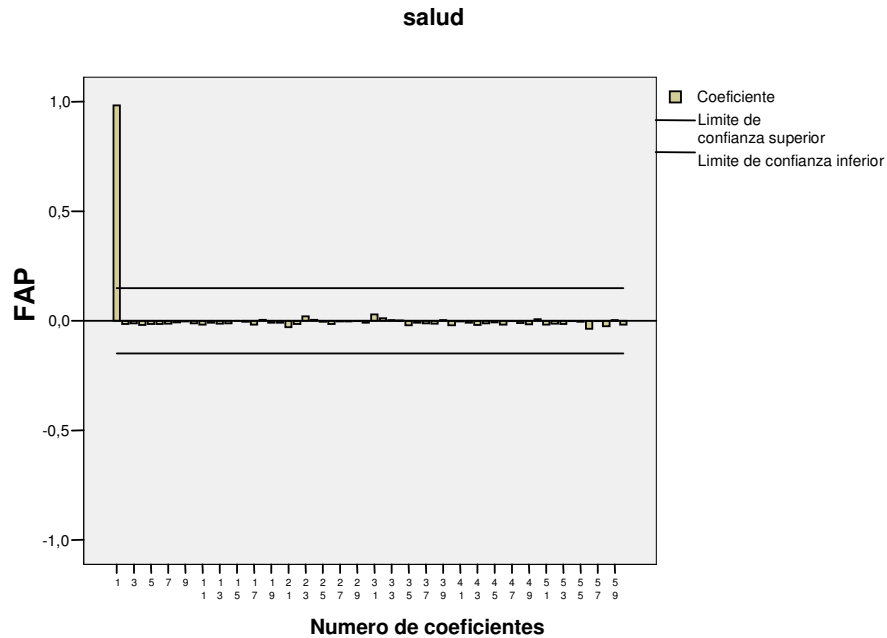
**Grafico 7.1.2.2: Función Autocorrelación Simple (FAS)**



**Gráfico 7.1.2.3: Función Autocorrelación Simple Estacional**



**Gráfico 7.1.2.4: Función de Autocorrelación Parcial (FAP)**



Podemos observar que no se tiene estacionariedad en media tanto en la parte regular como en la estacional; por que no se mantiene el supuesto de que los datos varían alrededor de una única media; esto se confirma en el gráfico de las funciones de Autocorrelación (FAS regular y estacional), que decaen lentamente a lo largo del tiempo, razón que nos lleva a tener que diferenciar regular y estacionalmente ya que se detecta el componente estacional. Antes de profundizar en el análisis es importante que exploremos los estadísticos descriptivos, esto con el fin de tener un bosquejo general del comportamiento de la serie. En la tabla se muestran los estadísticos importantes para este análisis exploratorio.

**Tabla 7.1.2.1: Estadísticos de la serie**

		Estadístico	Std. Error
salud	Media	158,66103	2,5477046
	Mediana	152,76500	
	Varianza	1012,565	
	Desviación Std.	31,820820	
	Asimetría	-,052	,194
	curtosis	-1,103	,386

Analizando la desviación típica podemos decir que la variación esperada con respecto de la media y cada dato es 31.820820.

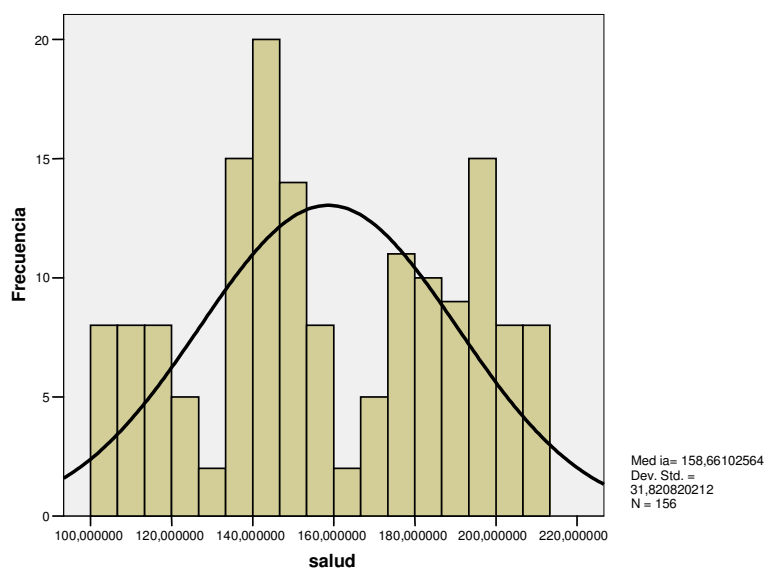
El valor obtenido en el *Coefficiente de Asimetría* (índice que expresa el grado de asimetría de la distribución), es de: - 0.052

- ➔ La asimetría negativa indica que los valores mas extremos se encuentran por debajo de la media.

El valor absoluto de la Curtosis es mayor que uno, lo que indica una total diferencia con la distribución normal. Según la teoría se sabe que si el valor de la Curtosis es negativo y mayor que uno, en una de las colas de la distribución hay acumulados más casos que en las colas de una distribución normal.

El histograma nos afirma todo lo visto antes con los resultados calculados, que los datos en estudio no presentan una semejanza a la distribución normal, y la mayor parte de los datos se encuentran acumulados en el lado derecho del histograma.

**Gráfico 7.1.2.5: Histograma de los datos originales**



**Tabla 7.1.2.2: Medias y Varianzas**

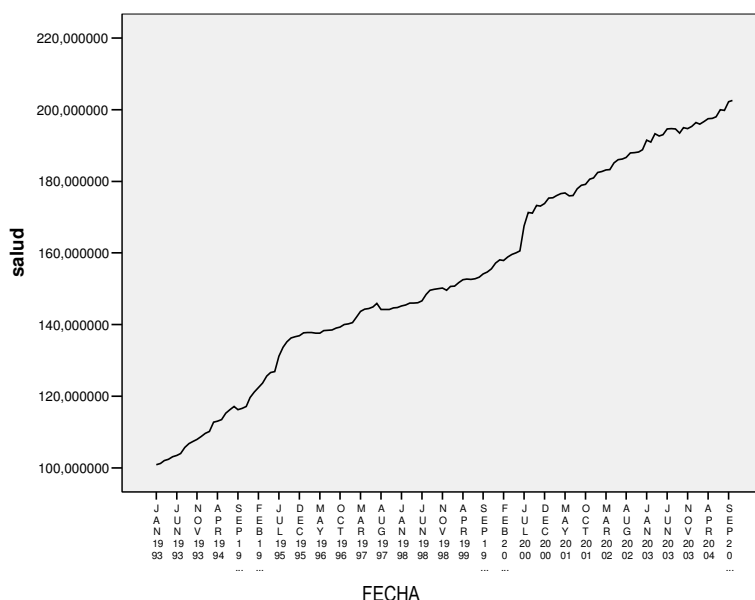
AÑOS	Media	Varianza	Deviation Std.
1993	104,5042	7,504	2,739384331
1994	114,8025	9,079	3,013162413
1995	129,6817	35,316	5,942706505
1996	138,5275	,865	,930250455
1997	144,0317	1,901	1,378832400
1998	147,7492	4,052	2,012905897
1999	153,2258	3,651	1,910775465
2000	165,4317	45,633	6,755250932
2001	177,4642	3,940	1,984968323
2002	185,7258	5,332	2,309161431
2003	193,6650	2,061	1,435494847
2004	199,3750	7,357	2,712464495
2005	208,4092	9,613	3,100476381
Total	158,6610	1012,565	31,820820212

En la tabla anterior notamos que cada período (llamamos periodo a los 12 datos de cada año, siendo el periodo uno, los datos desde enero de 1993 hasta diciembre del mismo año.) tiene media distinta, entonces podemos decir que esta serie no tiene estacionariedad en media por lo que nos hace pensar que hay que transformarla.

## IDENTIFICACIÓN DE LOS POSIBLES MODELOS

Para iniciar la identificación del modelo ARIMA, esclarecemos que se ha dejado el 20% de los datos, para efectos de comparación con los pronósticos generados a partir del modelo identificado. De ahora en adelante la serie consta de 142 observaciones, que corresponden al período de enero de 1993 – octubre de 2004.

**Gráfico 7.1.2.6: Serie menos el 20% de los datos**



Debido a que no disponemos de un subconjunto de datos que sobrepase el 60% de las observaciones, antes o después de la intervención, y al observar el gráfico de secuencia original de la serie tenemos una intuición respecto a la forma en que afecta la intervención. Utilizaremos el método 2 para la identificación del modelo ARIMA con intervención enunciado en la metodología.

Se inicia calculando un modelo ARIMA obviando la intervención identificada, este será un modelo ESTIMADO.

Desde el análisis descriptivo identificamos la Heteroscedasticidad, razón q permite analizar el test de levene las transformaciones como: logaritmo natural, raíz cuadrada inversa, raíz cuadrada. Según este test, cuanto menor sea el estadístico de Levene y mayor significancia, tanto mayor será la aproximación a la homocedasticidad. Por lo tanto se contrasta la hipótesis nula ( $H_0$ ) de que "Existe homocedasticidad" (es decir, la varianza es constante), se rechaza cuando el nivel de significación es menor que .05. El valor de Levene confirma que existe una significativa dispersión.



## Eliminación de la Heteroscedasticidad

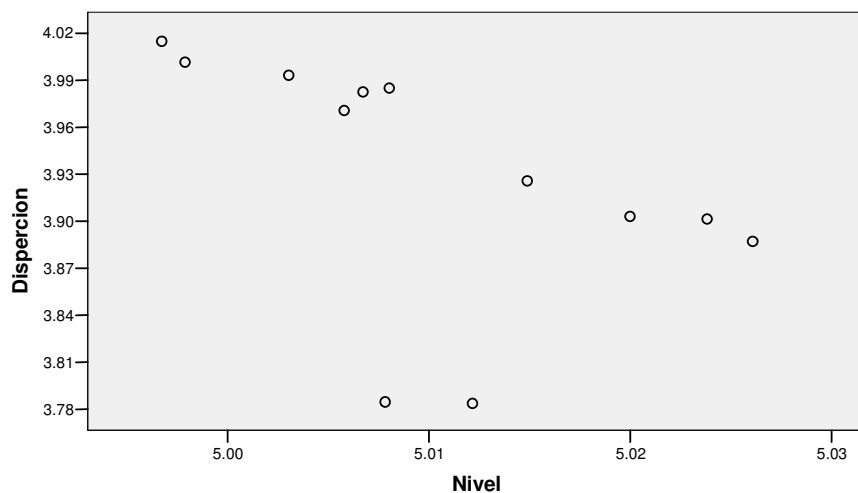
**Tabla 7.1.2.3: Estadístico de Levene para las transformaciones.**

Transformación	Estadístico, de levene basado en la media	sig
Sin transformación	0.31	1.00
Logaritmo Natural	0.33	1.00
Raíz cuadrada inversa	0.41	1.00
Raíz cuadrada	0.29	1.00

Según la tabla 7.1.2.3, la serie no se debe transformar, ya que para todas las transformaciones se tiene condiciones similares (el valor de la significancia es igual a 1 en todos los casos).

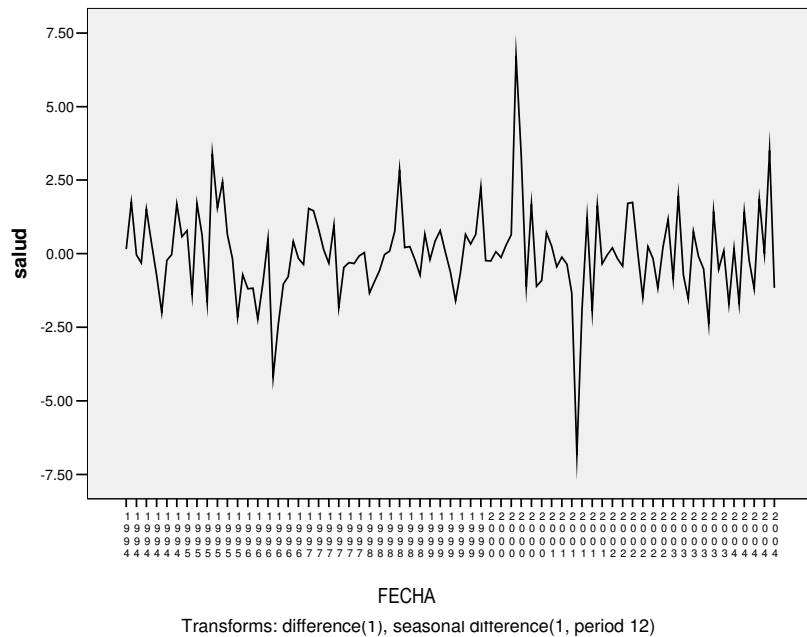
A continuación se muestra el gráfico rango-media que genera la serie sin ninguna transformación (para las transformaciones el gráfico es similar), debido a que no podemos decidir por ninguna transformación se opta por analizar la serie sin transformación.

**Gráfico 7.1.2.7: Rango-Media**



Ahora bien para lograr la estacionariedad en media se debe diferenciar la serie en la parte regular y estacional, permitiendo obtener el gráfico de secuencia y las funciones de Autocorrelación siguientes:

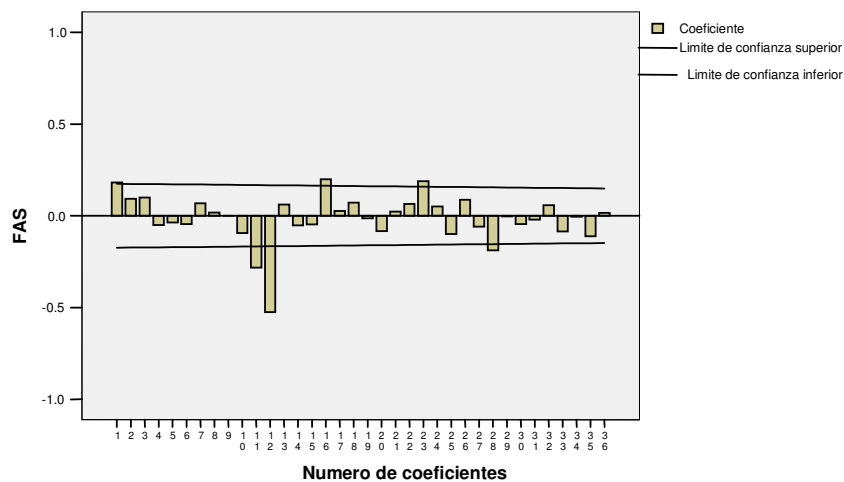
**Gráfico 7.1.2.8: Serie diferenciada estacional y regularmente**



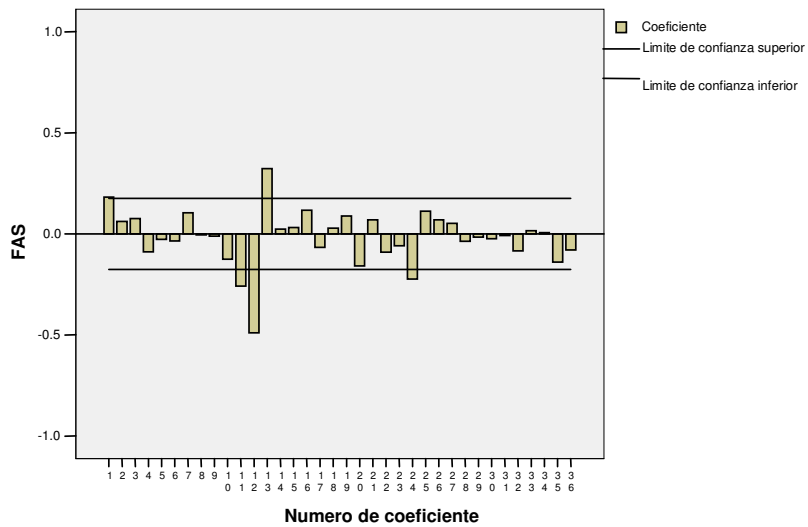
Se logra estacionariedad en media con una diferencia regular y una estacional, se nota una distorsión relevante en el centro de la grafica, que se debe a la intervención.

Ahora sí; la serie ya es estacionaria en media y en varianza. Por lo que ya se pueden analizar los gráficos de autocorrelación simple y parcial, con los cuales se plantearan los posibles modelos que se adecuan a los datos. Veamos estos gráficos:

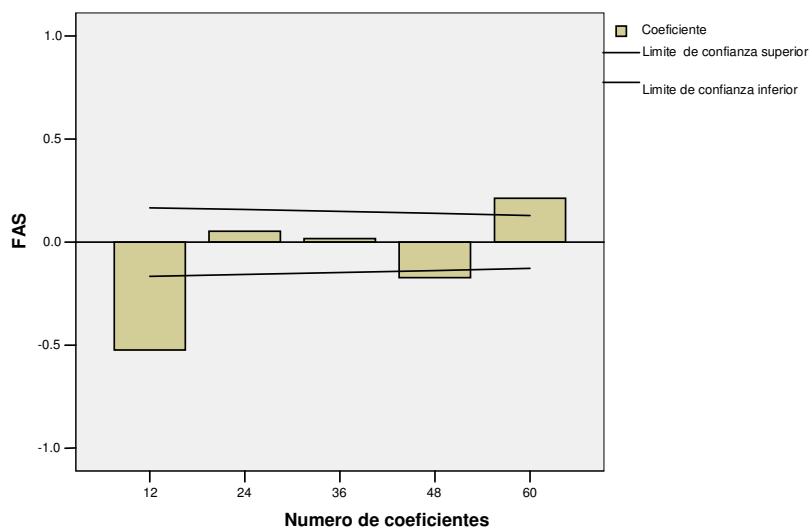
**Gráfico 7.1.2.9: Función de Autocorrelación Simple (FAS) de la parte regular**



**Gráfico 7.1.2.10: Función de Autocorrelación Parcial (FAP) de la parte regular**



**Gráfico 7.1.2.11: Función de Autocorrelación Simple (FAS) de la parte estacional**



A partir de la representación gráfica de las autocorrelaciones, simples y parciales de la parte regular y estacional, se identifican algunos modelos ARIMA, los que se analizarán en las etapas siguientes, permitiéndonos tener un solo modelo, que es el que identificaremos como el modelo que se ajusta mejor a los datos y que permitirá obtener mejores pronósticos. Los modelos planteados son los siguientes:

- 1-  $IMA(1,1) \times I(1)_{12}$   $(1-L)(1-L^{12})y_t = (1-\theta_1 L)u_t$
- 2-  $I(1) \times IMA(1,1)_{12}$   $(1-L)(1-L^{12})y_t = (1-\Theta_1 L^{12})u_t$
- 3-  $ARI(1,1) \times I(1)_{12}$   $(1-\phi_1 L)(1-L)(1-L^{12})y_t = u_t$
- 4-  $I(1) \times ARI(1,1)_{12}$   $(1-\Phi_1 L^{12})(1-L)(1-L^{12})y_t = u_t$
- 5-  $IMA(1,1) \times IMA(1,1)_{12}$   $(1-L)(1-L^{12})y_t = (1-\theta_1 L)(1-\Theta_1 L^{12})u_t$
- 6-  $ARI(1,1) \times ARI(1,1)_{12}$   $(1-\phi_1 L)(1-\Phi_1 L^{12})(1-L)(1-L^{12})y_t = u_t$
- 7-  $ARIMA(1,1,1) \times ARIMA(1,1,1)_{12}$   $(1-\phi_1 L)(1-\Phi_1 L^{12})(1-L)(1-L^{12})y_t = (1-\theta_1 L)(1-\Theta_1 L)u_t$

### ESTIMACIÓN DE LOS PARÁMETROS

Usando el paquete estadístico SPSS se obtiene la estimación de coeficientes de los modelos identificados, mediante el método de predicción de Mínimos Cuadrados Incondicional, con valores iniciales automáticos para la estimación, los resultados son los siguientes:

- 1-  $IMA(1,1) \times I(1)_{12}$  ningún parámetro es significativo
- 2-  $I(1) \times IMA(1,1)_{12}$  el parámetro  $\Theta_1$  no es significativo, por que su t-valor no es mayor que dos.
- 3-  $ARI(1,1) \times I(1)_{12}$   $(1-0.182L)(1-L)(1-L^{12})y_t = u_t$
- 4-  $I(1) \times ARI(1,1)_{12}$   $(1+0.550L^{12})(1-L)(1-L^{12})y_t = u_t$
- 5-  $IMA(1,1) \times IMA(1,1)_{12}$   $(1-L)(1-L^{12})y_t = (1-0.945L^{12})u_t$ , el parámetro de la parte regular no es significativo. Y se convierte en el modelo 2.
- 6-  $ARI(1,1) \times ARI(1,1)_{12}$   $(1+0.554L^{12})(1-L)(1-L^{12})y_t = u_t$  el parámetro de la parte regular no es significativo. Y se convierte en el modelo 4
- 7-  $ARIMA(1,1,1) \times I(1,1,1)_{12}$  ninguno de los parámetros es significativo en este modelo.

Ver corridas en anexo.

Con la estimación de parámetros se descartan tres modelos, por que sus parámetros no son significativos y dos porque se convierten en otro modelo ya analizado, los dos restantes pasan a la siguiente etapa.

**Tabla 7.1.2.4: VERIFICACIÓN DE COEFICIENTES PARA ESTACIONARIEDAD E INVERTIBILIDAD**

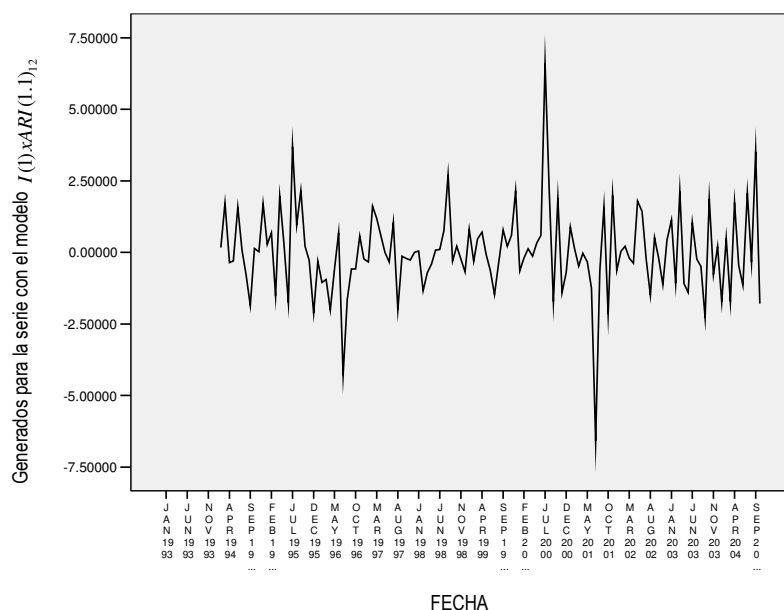
MODELO	ESTACIONARIEDAD	INVERTIBILIDAD
$ARI(1,1) \times I(1)_{12}$	Debe cumplirse que $ \phi_1  < 1$ , es decir, $ 0.182  < 1$ , el modelo es estacionario.	El modelo es un AR y una de las características de estos modelos es que siempre es invertible
$I(1) \times ARI(1,1)_{12}$	Debe cumplirse que $ \Phi_1  < 1$ , es decir, $ -0.550  < 1$ , el modelo es estacionario.	El modelo es un AR y una de las características de estos modelos es que siempre es invertible

Estos dos modelos, cumplen la condición de estacionariedad e invertibilidad, para poder elegir el mejor modelo se hace necesario el análisis de los residuos, verificar si se comportan como un ruido blanco.

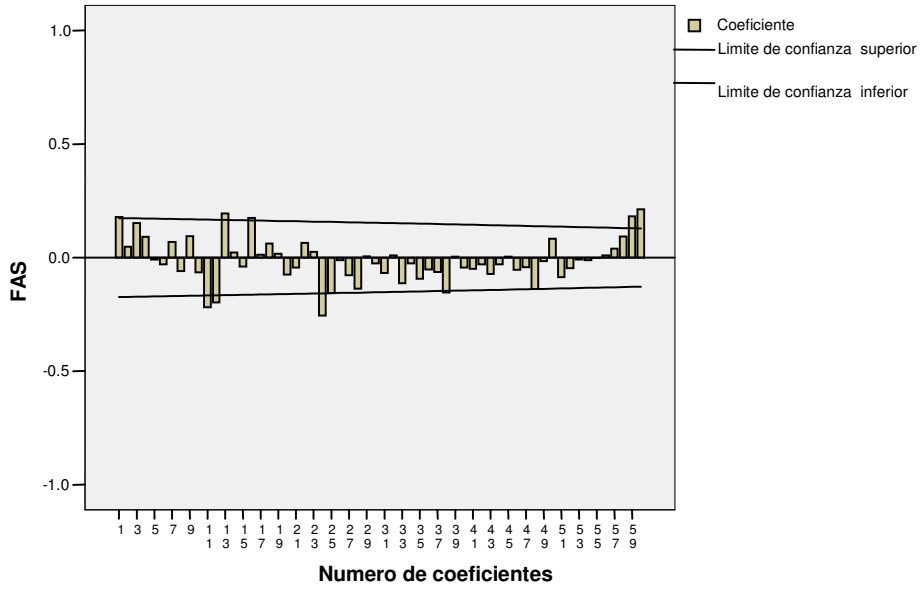
#### ANÁLISIS DE LOS RESIDUOS

Sometiendo los modelos al análisis de los residuos, las condiciones de media cero y varianza constante, tienen un mayor realce, para el modelo  $I(1) \times ARI(1,1)_{12}$ , de igual manera la condición de incorrelación. A continuación se verifica; observando el gráfico de secuencia y de las autocorrelaciones de los residuos de dicho modelo.

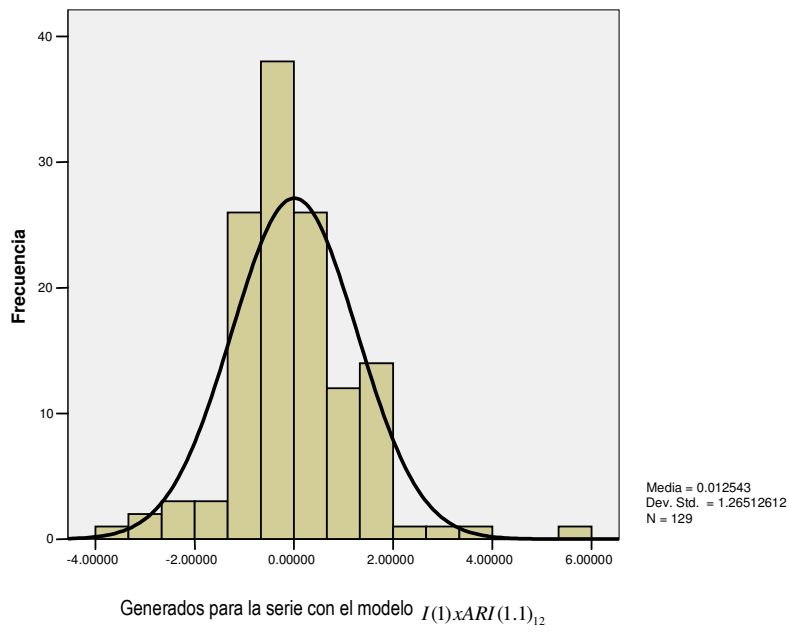
**Gráfico 7.1.2.12: Residuos con el modelo  $I(1) \times ARI(1,1)_{12}$**



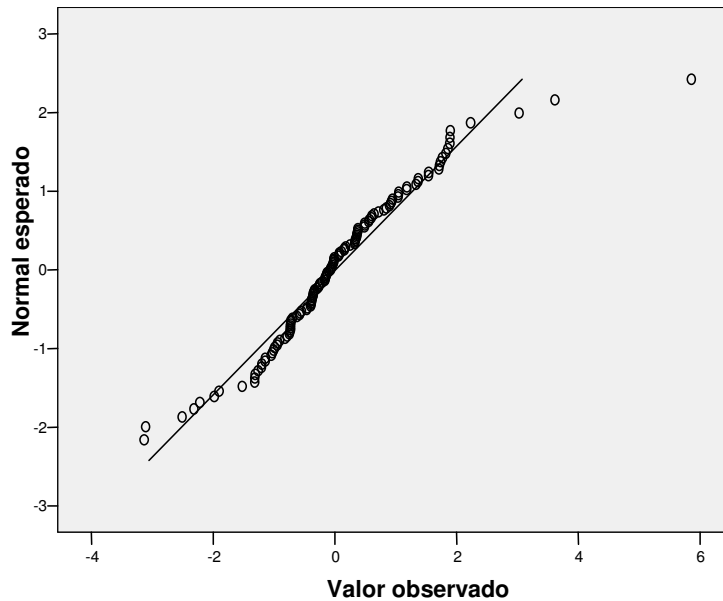
**Gráfico 7.1.2.13: FAS de los residuos con el modelo  $I(1) \times ARI(1.1)_{12}$**



**Gráfico 7.1.2.14: Histograma Residual**



**Gráfico 7.1.2.15: Probabilidad Normal de los Residuos**



**Tabla 7.1.2.5: Test de Normalidad**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	df	Sig.	Estadístico	df	Sig.
Error generados con el modelo $I(1) \times ARI(1.1)_{12}$	.100	129	.003	.924	129	.000

a. Lilliefors Significance Correction

Obsérvese que la normalidad no se cumple; ya que la significancia de los estadísticos en la tabla 7.1.2.5, son inferiores al 0.05. La incorrelación de los residuos se ve afectada por la presencia de la intervención y el cambio que ocasiona.

Al analizar los modelos con los estadísticos AIC (El Criterio Akaike) y BIC (El criterio de información Bayesiana); permite que se tenga un criterio más para identificar el modelo generador de la serie, ya que se ha dicho que el modelo adecuado será aquel que cumple mayores condiciones y criterios.

**Tabla 7.1.2.6: Estadísticos AIC y BIC**

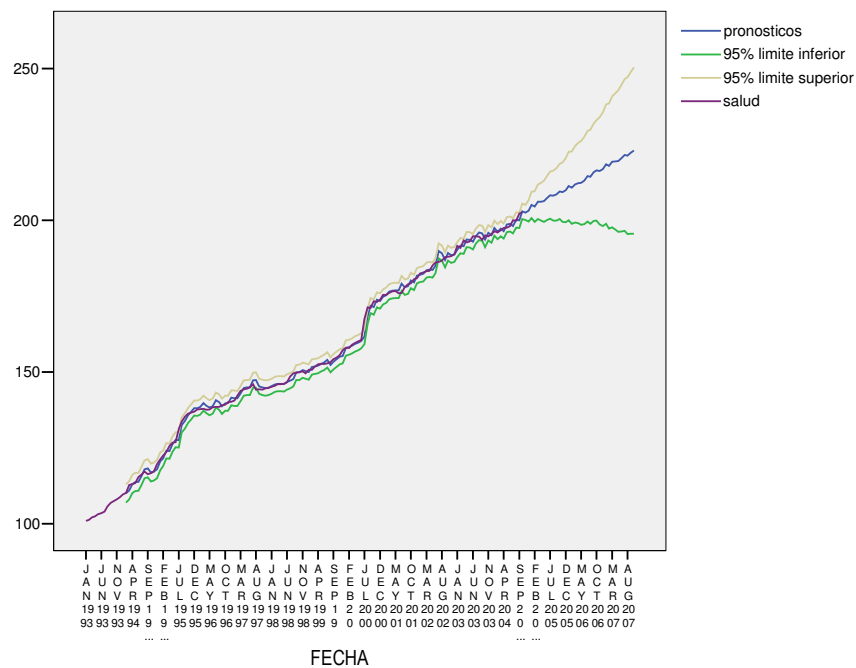
modelo	AIC	BIC
$ARI(1,1) \times I(1)_{12}$	471.804	477.524
$I(1) \times ARI(1.1)_{12}$	431.508	437.228

Según las condiciones anteriores el mejor modelo que genera los datos es  $IMA(1,1) \times IMA(1,1)_{12}$  pero este modelo se descarta de ser el mejor que el parámetro de la parte regular no es significativo, y se convierte en otro modelo ya analizado que no cumple otras condiciones, por esta razón tomamos el modelo  $I(1) \times ARI(1.1)_{12}$  como el mejor se ajusta a los datos. Verifiquémoslo con los pronósticos siguientes:

### PREDICCIONES

A continuación se verificará si los pronósticos que genera este modelo, resultan adecuados, se pronostican 36 datos, desde noviembre del 2004 hasta octubre del 2007 y se comparan con los datos que se han dejado para este efecto.

**Gráfico 7.1.2.16: Superpuesto de datos reales y pronosticados**





**Tabla 7.1.2.7: Valores de los pronósticos**

fecha	Datos reales	pronósticos	error	Limite inferior	Limite superior
nov-04	202,56	202.55129	0.01	200.05277	205.04980
dic-04	203,26	203.13621	0.12	199.58826	206.68415
ene-05	202,7	205.04173	-2.34	200.67871	209.40476
feb-05	203,93	204.50267	-0.57	199.44433	209.56102
mar-05	205,25	206.11171	-0.86	200.43363	211.78978
abr-05	206,33	206.11092	0.22	199.86616	212.35567
may-05	208,17	206.26627	1.90	199.49456	213.03799
jun-05	209,45	207.38041	2.07	200.11281	214.64802
jul-05	210,28	208.25322	2.03	200.51482	215.99162
ago-05	210,24	208.10312	2.14	199.91469	216.29154
sep-05	211,12	208.53328	2.59	199.91234	217.15422
oct-05	210,49	209.49093	1.00	200.45248	218.52938
nov-05	211,08	209.28847	1.79	199.45938	219.11756
dic-05	211,87	209.89007	1.98	199.31914	220.46099
ene-06	216,87	211.26215	5.61	199.98854	222.53576
feb-06	217,35	210.76395	6.59	198.81991	222.70799
mar-06	217,07	211.84861	5.22	199.26123	224.43598
abr-06	217,28	212.26348	5.02	199.05590	225.47105
may-06	218,16	212.34181	5.82	198.53404	226.14957
jun-06	219,5	213.08271	6.42	198.69227	227.47314
jul-06	220,06	214.50114	5.56	199.54350	229.45879
ago-06	221,79	214.33749	7.45	198.82640	229.84858
sep-06	221,69	215.81504	5.87	199.76286	231.86722
oct-06	222,46	216.40851	6.05	199.82637	232.99066
nov-06	222,13	216.27148	5.86	198.61046	233.93249
dic-06	222,51	216.83381	5.68	198.14279	235.52483
ene-07	222,11	218.46907	3.64	198.78925	238.14889
feb-07	222,73	217.91831	4.81	197.28495	238.55166
mar-07	223,85	219.26116	4.59	197.70486	240.81745
abr-07	224,03	219.41741	4.61	196.96499	241.86982
may-07	224,37	219.50798	4.86	196.18317	242.83279
jun-07	224,73	220.42399	4.31	196.24794	244.60003
jul-07	224,72	221.51234	3.21	196.50406	246.52063
ago-07	224,62	221.32604	3.29	195.50268	247.14940
sep-07	224,76	222.19766	2.56	195.57481	248.82051
oct-07	225,41	222.96125	2.45	195.55313	250.36937

Observando el gráfico de los pronósticos junto con los datos reales, notamos que nuestro modelo si se ajusta adecuadamente a los datos. Por lo tanto aceptamos el modelo  $I(1) \times ARI(1.1)_{12}$ , como el que mejor se ajusta a los datos.

## ANALISIS DE INTERVENCION

Como señalamos en el marco teórico, las dos características de los modelos de intervención que debemos ser capaces de identificar son: el período de comienzo de la intervención y la forma general del impacto. Conocemos que la intervención se dio en  $t_0=90$ , una intuición de la forma del impacto de la intervención es: comienzo brusco y duración permanente. En consecuencia el modelo de intervención que debemos especificar, considerando adicionalmente un período de desfase temporal para el efecto ( $b=1$ ), es:

$$y_t = \omega L S_t^{t_0=90}$$

Siendo  $S_t$  la variable escalón definido como:

$$S_t = \begin{cases} 1 & \text{para } t \geq t_0 = 90 \\ 0 & \text{para } t < t_0 = 90 \end{cases}$$

O lo que es igual:

$$y_t = \omega S_t^{t_0=91}$$

Ya identificado la Estimación del modelo ARIMA generador de la serie sin Análisis de intervención procedemos a obtener los residuos de un modelo en donde se añade el modelo de la intervención, esto es:

$$y_t = \omega S_t^{t_0=91} + \frac{u_t}{(1-\Phi_1 L)(1-L)^2(1-L)}$$

Los residuos que se obtienen a partir de este modelo, son los siguientes:

$$\tilde{N}_t = y_t - \omega S_t^{t_0=91}$$

Donde: 
$$\tilde{N}_t = \frac{u_t}{(1-\Phi_1 L)(1-L)^2(1-L)}$$

Al introducir este modelo en el programa STATISTICA 7.0, para la estimación de parámetros, y análisis de los residuos, se obtienen los resultados siguientes:

**Tabla 7.1.2.8: Estimación de los parámetros:**

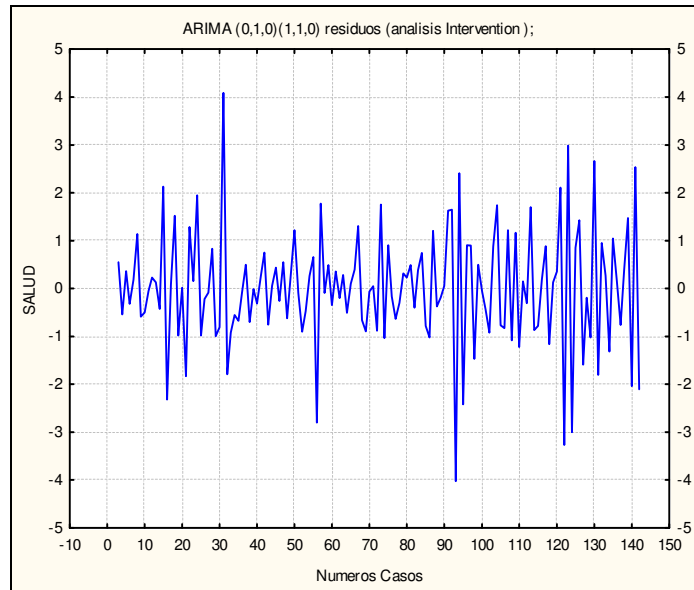
Input: SALUD								
Transformaciones: D(1),D(1) (ARIMA intervencion)								
Modelo:(0,1,0)(1,1,0) estacional lag: 12 MS Residual= 1.5121								
Paramet.	Param.	Std.Err.	t( 138)	p	Superior 95% Conf	Inferior 95% Conf	N caso Interv.	Tipo Interv.
Ps(1)	0.025151	0.091070	0.276173	0.782828	-0.154922	0.205225		
Omega(1)	4.789049	0.873007	5.485695	0.000000	3.062849	6.515248	91	Abr/Perm

El parámetro estacional AR no se incluye al modelo, resulta ser no significativo, el valor de t es menor que dos, en nuestra estimación se observa que no existe parte ARIMA, nosotros conocemos que este es un modelo estimado y observamos la FASR y FAPR para conocer el verdadero modelo.

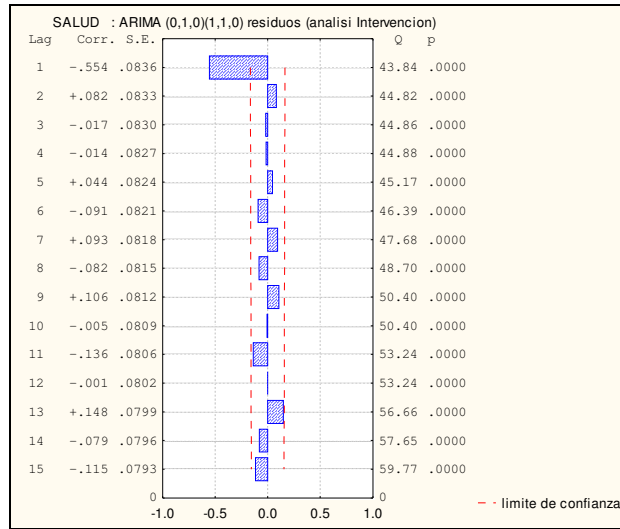
El modelo con sus parámetros estimados es el siguiente

$$\sqrt{y_t} = 0.4.789S_t^{t_0=91} + \frac{u_t}{(1-L^2)(1-L)} \text{ con } \Phi = 0$$

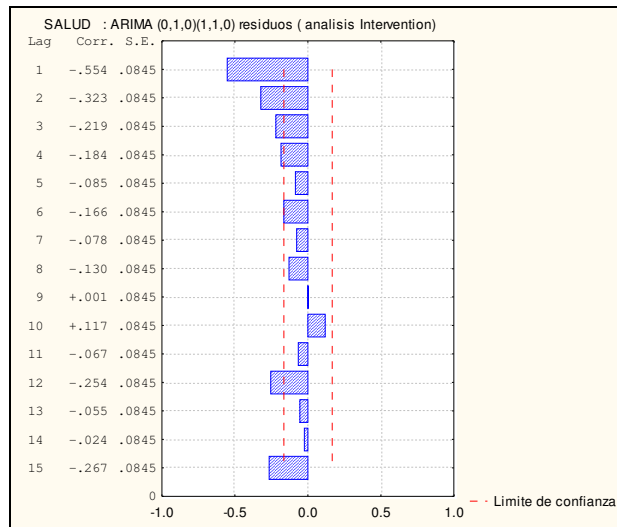
**Gráfico7.1.2.17: Residuos del modelo**



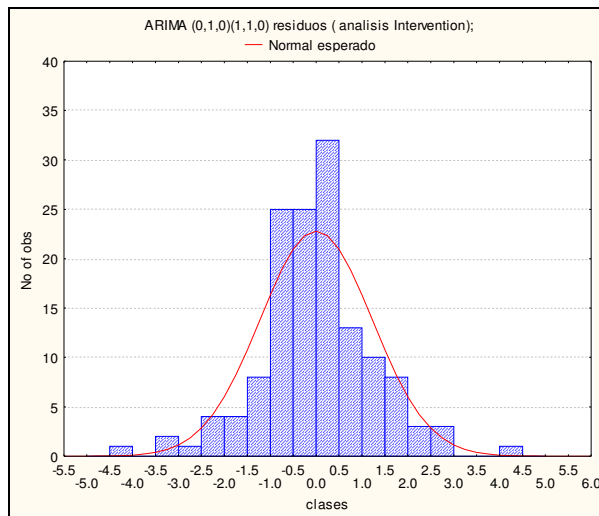
**Gráfico 7.1.2.18: FASR**



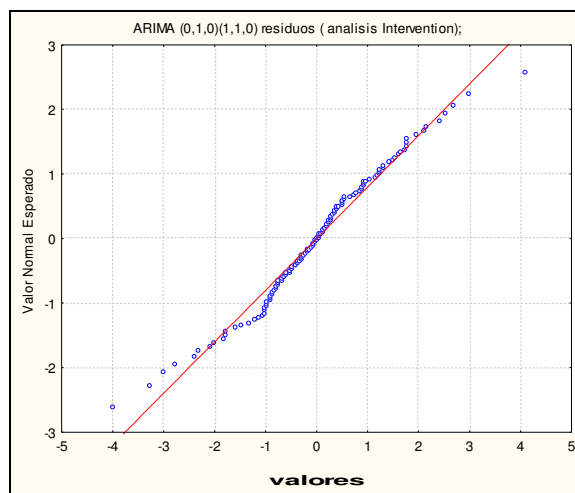
**Gráfico 7.1.2.19: FAPR**



**Gráfico 7.1.2.20: Histograma Residual:**



**Gráfico 7.1.2.21: Probabilidad Normal de los Residuos:**



Este modelo ha generado residuos, que no se comportan como un ruido blando. En efecto, el correlograma muestral residual y correlograma parcial muestral residual, nos llevan a identificar nuevos modelos ARIMA, que se analizan para identificar un nuevo mejor Modelo, que pueda ser el que mejor se ajuste a los datos tomando en cuenta la intervención.

Los modelos identificados a partir de los correlogramas de los residuos son:

*MA (1), AR (1), ARMA (1,1), AR (2), ARMA (2,1).*

Tomando en cuenta la transformación que se le hizo a los datos para lograr estacionariedad en varianza, y las diferenciaciones en la parte regular y estacional, para lograr la estacionariedad en media, los modelos serian:

$IMA(1,1) \times I(1)_{12}$ ,  $ARI(1,1) \times I(1)_{12}$ ,  $ARIMA(1,1,1) \times I(1)_{12}$ ,  $ARI(2,1) \times I(1)_{12}$ ,  $ARIMA(2,1,1) \times I(1)_{12}$ .

Matemáticamente e incluyendo el modelo de la intervención los modelos se escriben como:

$$1- y_t = \omega S_t^{t_0=91} + \frac{(1-\theta_1 L)}{(1-L^{12})(1-L)} u_t$$

$$2- y_t = \omega S_t^{t_0=91} + \frac{u_t}{(1-\phi_1 L)(1-L^{12})(1-L)}$$

$$3- y_t = \omega S_t^{t_0=91} + \frac{u_t}{(1-\phi_1 L - \phi_2 L^2)(1-L^{12})(1-L)}$$

$$4- y_t = \omega S_t^{t_0=91} + \frac{(1-\theta_1 L)}{(1-\phi_1 L)(1-L^{12})(1-L)} u_t$$

$$5- y_t = \omega S_t^{t_0=91} + \frac{(1-\theta_1 L)}{(1-\phi_1 L - \phi_2 L^2)(1-L^{12})(1-L)} u_t$$

Con el STATISTICA 7.0, se obtiene la estimación de parámetros de los modelos identificados, mediante el método de predicción de Mínimos Cuadrados Incondicional, con valores iniciales automáticos para la estimación, los resultados son los siguientes:

$$1- y_t = 0.9252 S_t^{t_0=91} + \frac{(1-6.2283L)}{(1-L^{12})(1-L)} u_t$$

$$2- y_t = 6.3693 S_t^{t_0=91} + \frac{u_t}{(1+0.5941L)(1-L^{12})(1-L)}$$

$$3- y_t = 6.263 S_t^{t_0=91} + \frac{u_t}{(1+0.7753L+0.322L^2)(1-L^{12})(1-L)}$$

$$4- y_t = 6.474 S_t^{t_0=91} + \frac{(1-0.8943L)}{(1-L^{12})(1-L)} u_t, \text{ el parámetro autoregresivo no es significativo, y se}$$

Convierte en el modelo 1.

$$5- y_t = 6.5224 S_t^{t_0=91} + \frac{(1-0.9536L)}{(1-L^{12})(1-L)} u_t, \text{ los parámetros de la parte autoregresiva no son}$$

significativo, y se convierte en el modelo 1.

Ver corridas en anexos

Al analizar la estimación de los parámetros, el modelo 4 y 5, se descartan y no pasan a la siguiente etapa. Ya que no son útiles para las predicciones necesarias.

**Tabla 7.1.2.9: VERIFICACIÓN DE COEFICIENTES PARA ESTACIONARIEDAD E INVERTIBILIDAD**

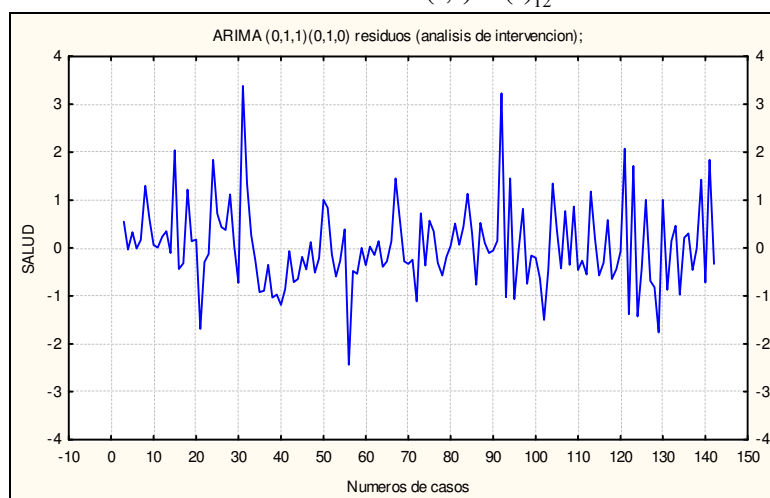
MODELO	ESTACIONARIEDAD	INVERTIBILIDAD
$IMA(1,1) \times I(1)_{12}$	El modelo es un MA y una de las características de estos modelos es que siempre es Estacionario	Debe cumplirse que $ \theta_1  < 1$ , es decir, $ 0.9252  < 1$ , el modelo es invertible.
$ARI(1,1) \times I(1)_{12}$	Debe cumplirse que $ \phi_1  < 1$ , es decir, $ -0.5941  < 1$ , el modelo es estacionario.	El modelo es un AR y una de las características de estos modelos es que siempre es invertible
$ARI(2,1) \times I(1)_{12}$	Debe cumplirse que $\phi_1 + \phi_2 < 1$ , $\phi_1 - \phi_2 < 1$ y $ \phi_2  < 1$ , es decir, $ (-0.7753) + (-0.322)  < 1$ , $ (-0.322) - (-0.7753)  < 1$ , Y $ -0.7753  < 1$ . El modelo es estacionario.	El modelo es un AR y una de las características de estos modelos es que siempre es invertible

Los tres modelos cumplen la condición de invertibilidad y estacionariedad, aunque el último apenas cumple la condición de estacionariedad.

Sometiendo los modelos al análisis de los residuos, las condiciones de media cero y varianza constante, tiene un mayor realce para el modelo  $IMA(1,1) \times I(1)_{12}$  de igual manera la condición de incorrelación.

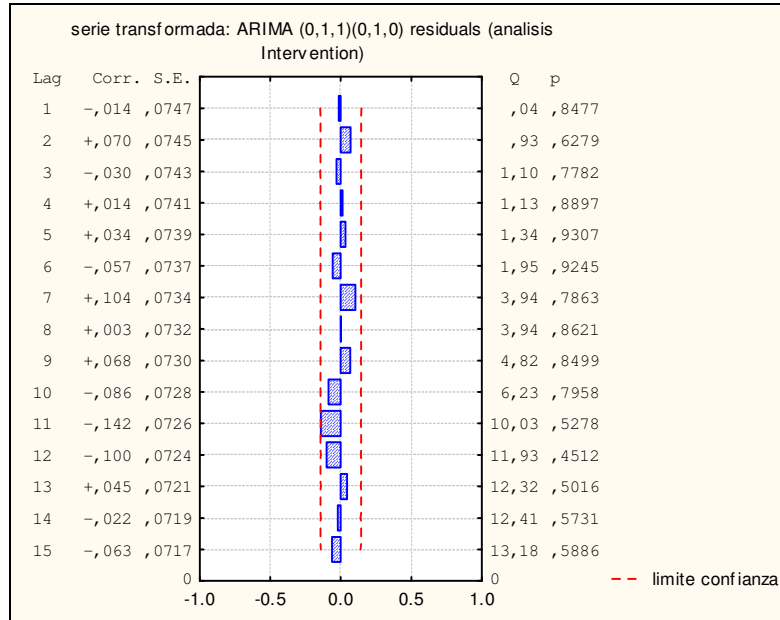
A continuación se verifica observando los gráficos de secuencia y de las autocorrelaciones de los residuos de dicho modelo.

**Gráfico 7.1.2.22: Residuos con el modelo  $IMA(1,1) \times I(1)_{12}$**

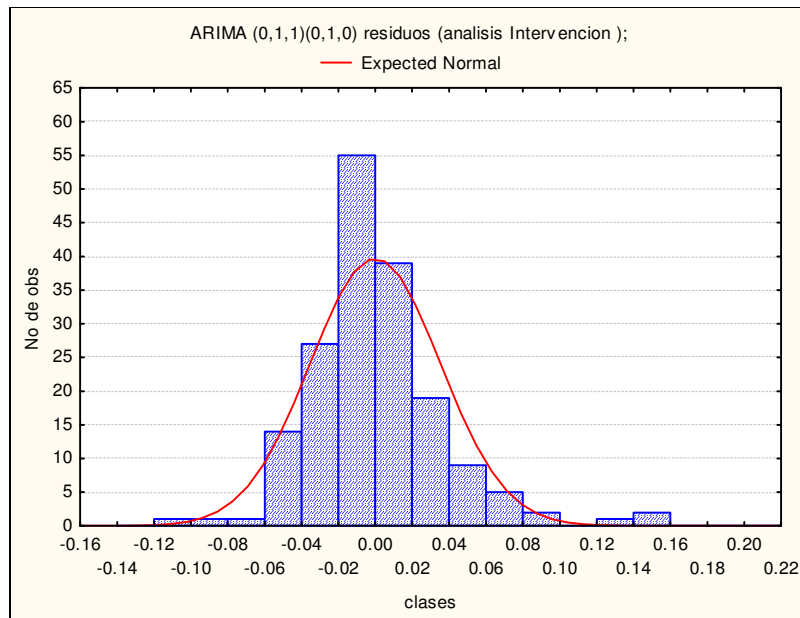


De este gráfico podemos decir que los residuos cumplen con la condición de ser constantes en el tiempo.

**Gráfico 7.1.2.23: FASR:**

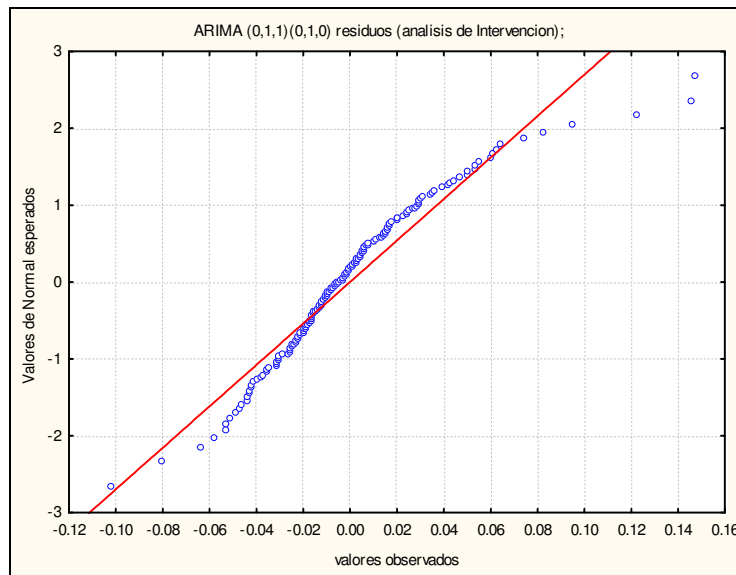


**Gráfico 7.1.2.24: Histograma Residual:**





**Grafico 7.1.2.25: Probabilidad Normal de los Residuos:**



Al observar el gráfico 7.1.2.22, notamos que los residuos de este modelo tienen un comportamiento constante a lo largo del tiempo, además en gráfico 7.1.2.23, más del 95% de los retardos no son significativos, aquí se cumple la incorrelación de los residuos. Con el gráfico 7.1.2.24 y 7.1.2.25, verificamos que los residuos tienen un comportamiento normal, y así certificamos que los residuos de este modelo se comportan como un ruido blanco.

En consecuencia, aceptando este proceso como el generador de la serie el modelo se puede escribir, obviando el símbolo de estimación en  $N_t$  como:

$$N_t = \frac{(1 - \theta_1 L)}{(1 - L^2)(1 - L)} (1 - \theta_1 L) u_t$$

El modelo ARIMA con intervención finalmente especificado es:

$$y_t = \omega S_t^{t_0=91} + \frac{(1 - \theta_1 L)}{(1 - L^2)(1 - L)} u_t$$

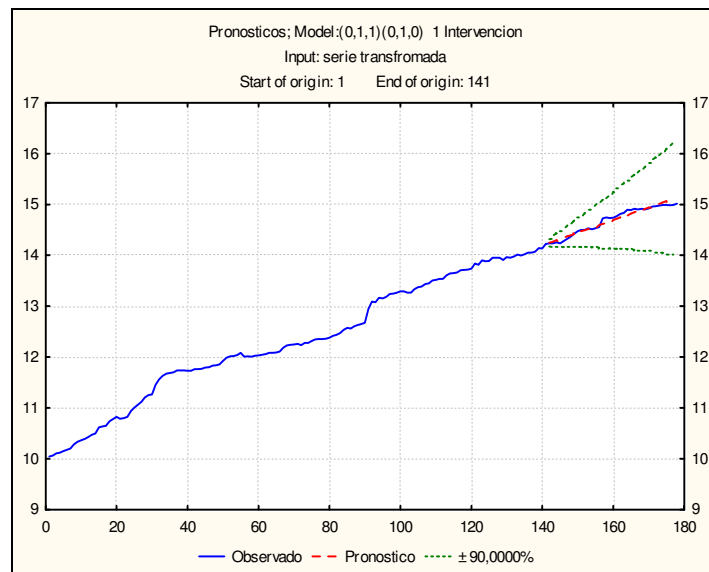
La estimación de este modelo, por el método de máxima verosimilitud exacta es:

$$y_t = 6.2283 S_t^{t_0=91} + \frac{(1 - 0.9252L)}{(1 - L^2)(1 - L)} u_t$$

En la etapa de La diagnosis del modelo, que se inicio con el análisis de los residuos, se puede concluir que los parámetros son significativos individualmente (adoptando los niveles de significancia y estadístico  $t$ , estándares habituales en el enfoque de verificación), también se verifico la incorrelación observando el correlograma muestral residual que no tiene ningún coeficiente que sobrepase los limites de confianza.

Para poder concluir definitivamente que el modelo  $IMA(1,1) \times I(1)_{12}$  resulta adecuado nos queda por analizar la capacidad predictiva del mismo.

**Gráfico 7.1.2.26: Superpuesto de los Datos Pronosticados y los Datos Originales:**



Al observar este gráfico, nos damos cuenta que los valores pronosticados tienen mucha similitud con el comportamiento de los datos originales.

**Tabla 7.1.2.10: Valores de los pronósticos**

Fecha	Datos reales	pronósticos	error	Limite inferior	Limite superior
nov-04	202.56	203.2405	0.6805	201.7774	204.7037
dic-04	203.26	203.9011	0.6411	201.7532	206.0490
ene-05	202.7	204.5616	1.8616	201.8337	207.2896
feb-05	203.93	205.2222	1.2922	201.9588	208.4856
mar-05	205.25	205.8827	0.6327	202.1063	209.6591
abr-05	206.33	206.5433	0.2133	202.2654	210.8211
may-05	208.17	207.2038	-0.9662	202.4299	211.9777
jun-05	209.45	207.8644	-1.5856	202.5960	213.1327
jul-05	210.28	208.5249	-1.7551	202.7612	214.2886
ago-05	210.24	209.1855	-1.0545	202.9239	215.4470
sep-05	211.12	209.8460	-1.2740	203.0827	216.6093
oct-05	210.49	210.5065	0.0165	203.2370	217.7761
nov-05	211.08	211.1671	0.0871	203.3859	218.9482
dic-05	211.87	211.8276	-0.0424	203.5292	220.1261
ene-06	216.87	212.4882	-4.3818	203.6663	221.3100
feb-06	217.35	213.1487	-4.2013	203.7971	222.5003
mar-06	217.07	213.8093	-3.2607	203.9214	223.6971
abr-06	217.28	214.4698	-2.8102	204.0390	224.9006
may-06	218.16	215.1304	-3.0296	204.1499	226.1108
jun-06	219.5	215.7909	-3.7091	204.2539	227.3279
jul-06	220.06	216.4515	-3.6085	204.3511	228.5518
ago-06	221.79	217.1120	-4.6780	204.4413	229.7827
sep-06	221.69	217.7725	-3.9175	204.5246	231.0204
oct-06	222.46	218.4331	-4.0269	204.6011	232.2651
nov-06	222.13	219.0936	-3.0364	204.6706	233.5167
dic-06	222.51	219.7542	-2.7558	204.7332	234.7751
ene-07	222.11	220.4147	-1.6953	204.7890	236.0405
feb-07	222.73	221.0753	-1.6547	204.8380	237.3126
mar-07	223.85	221.7358	-2.1142	204.8801	238.5915
abr-07	224.03	222.3964	-1.6336	204.9155	239.8772
may-07	224.37	223.0569	-1.3131	204.9442	241.1696
jun-07	224.73	223.7175	-1.0125	204.9662	242.4687
jul-07	224.72	224.3780	-0.3420	204.9816	243.7744
ago-07	224.62	225.0385	0.4185	204.9904	245.0866
sep-07	224.76	225.6991	0.9391	204.9927	246.4055
oct-07	225.41	226.3596	0.9496	204.9885	247.7308

Numéricamente podemos indicar que estos últimos pronósticos son mejores que los obtenidos con el análisis ARIMA sin intervención, los errores de los pronósticos con el ANALISIS de intervención es menor que los errores del análisis sin intervención. Permitiendo así testificar, las mejorías al incorporar el análisis de intervención, en una serie.

## 7.2. Aplicación de ST con Outliers

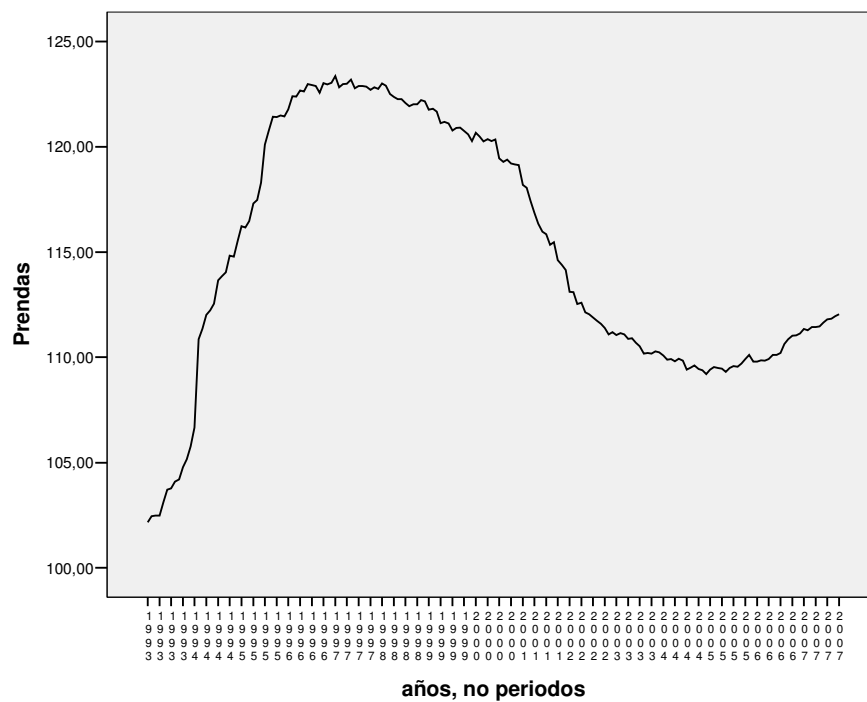
### 7.2.1. Serie Temporal del IPC en Prendas de Vestir y Calzado

#### ANÁLISIS DESCRIPTIVO

La Serie Temporal del Índices de Precios (IPC) al Consumidor en prendas de vestir y calzado, consta de 178 datos con periodicidad mensual desde enero de 1993 a octubre de 2007 (ver serie en anexos), para efectos de comparación con los pronósticos se dejaron 36 datos, tomados a partir de noviembre de 2004 a octubre de 2007.

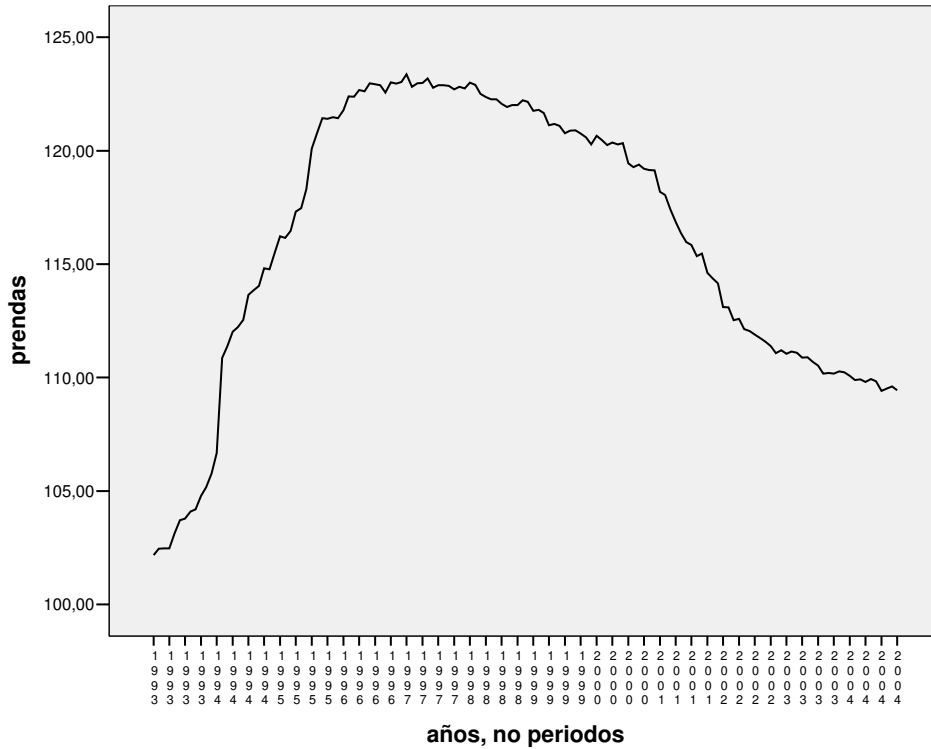
Veamos el gráfico de secuencia de la serie original:

**Gráfico 7.2.1.1: Serie prendas de vestir y calzado**



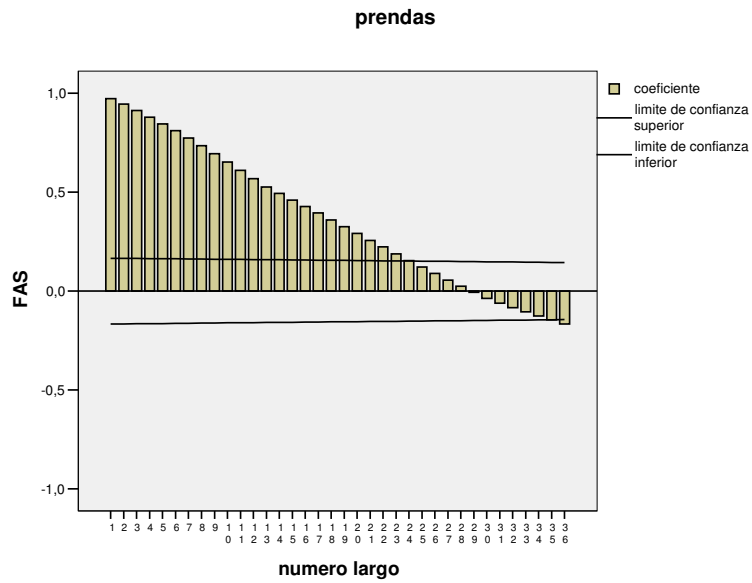
Observemos el gráfico de la serie quitando los 36 datos, que se tomaran para comparar pronósticos.

**Gráfico 7.2.1.2: Serie prendas de vestir y calzado menos 20%**

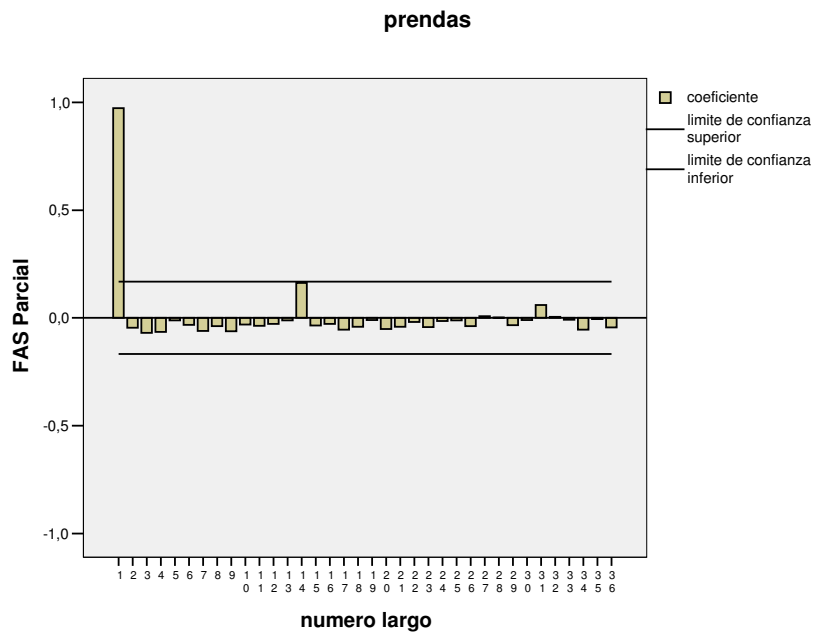


En el análisis ARIMA, es necesario determinar si la serie de interés presenta estacionariedad y homoscedasticidad, antes de identificar los posibles modelos y solo se puede saber, interpretando las gráficas de las funciones de autocorrelación (FAS Y FAP) que se muestran a continuación:

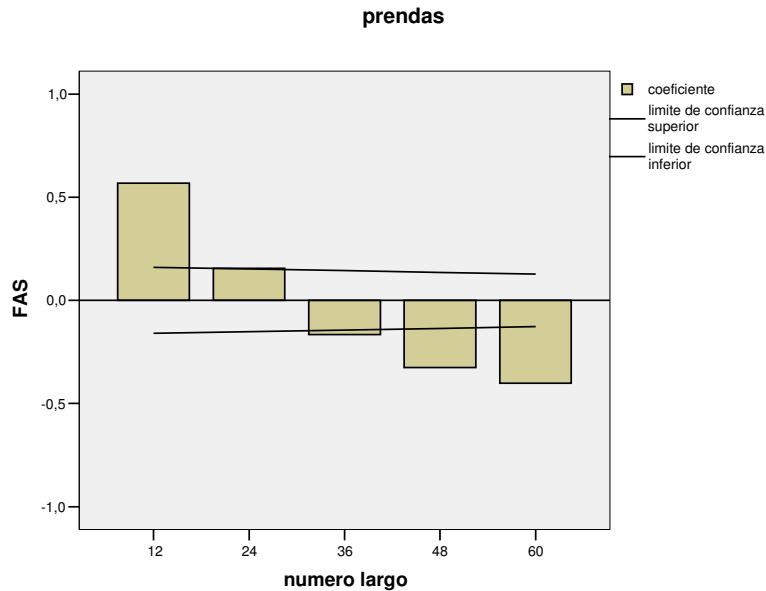
**Gráfico 7.2.1.3: Autocorrelaciones Simples Regulares (FAS)**



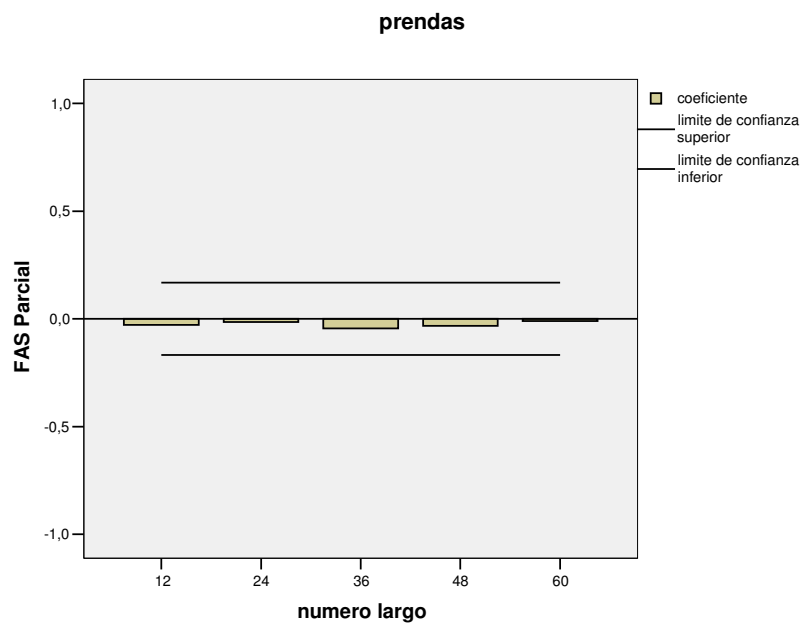
**Gráfico 7.2.1.4: Autocorrelaciones Parciales Regulares (FAP)**



**Gráfico 7.2.1.5: Autocorrelaciones Simples Estacionales (FAS)**



**Gráfico 7.2.1.6: Autocorrelaciones Parciales Estacionales (FAP)**



Podemos observar que no se tiene estacionariedad en media tanto en la parte regular como en la estacional; por que no se mantiene el supuesto de que los datos varían alrededor de una única media; esto se confirma observando detenidamente las fluctuaciones que presenta la serie en el gráfico de secuencia, y el comportamiento de las funciones de Autocorrelación (tanto en la parte regular y estacional), razón que nos lleva a diferenciar regular y

estacionalmente. Aunque la componente estacional decae rápidamente, permitiendo probar si con solo diferenciar regularmente se logra la estacionariedad en media en la parte estacional (ver gráficos 7.2.1.5 y 7.2.1.6). Por el otro lado la estacionariedad en varianza si se tiene, los datos no están muy dispersos; aunque aseguraremos esta aseveración analizando el test de levene posteriormente.

En la verificación de la estacionariedad en varianza con el test de levene, se concluye que no se tendrá que transformar la serie; ya que la significancia del test, es uno para los tres tipos de transformación de igual forma para la sin transformación, indicándolos que se puede continuar con el análisis sin hacer transformación a la serie.

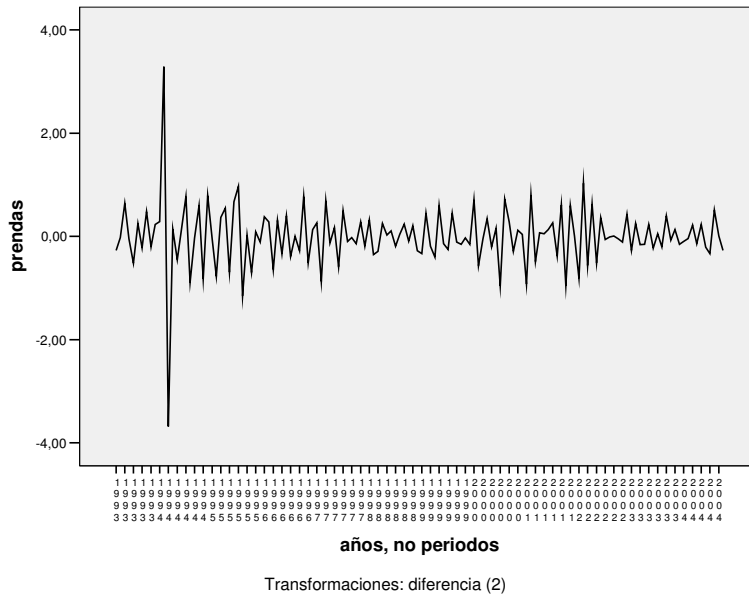
**Tabla 7.2.1.1: Test de levene y significancia**

Transformación	Test	Significancia
Sin	0.055	1.000
Log. Natural	0.059	1.000
Raíz cuadrada inversa	0.067	1.000
Raíz cuadrada	0.059	1.000

Ahora bien para lograr la estacionariedad en media se transforma la serie con dos diferenciaciones regulares; al principio dijimos que la serie presentaba componente estacional pero que se diferenciaría primero la parte regular por el decaimiento rápido que se ve. En los gráficos 7.2.1.8 y 7.2.1.9, se comprueba que, una diferenciación regular es suficiente para darse cuenta que no es necesario hacer diferencias estacionales; ya que los retardos decaen rápido en la parte estacional perdiendo la significatividad de los mismos. Observemos el gráfico de secuencia y las autocorrelaciones siguientes donde la serie esta libre de heteroscedasticidad:

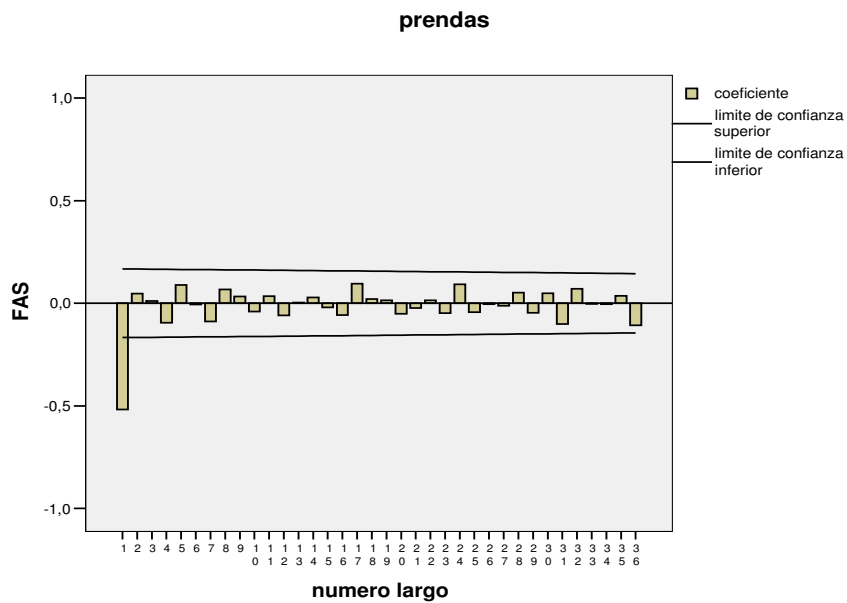


**Gráfico7.2.1.7: Serie con dos diferenciaciones Regulares**

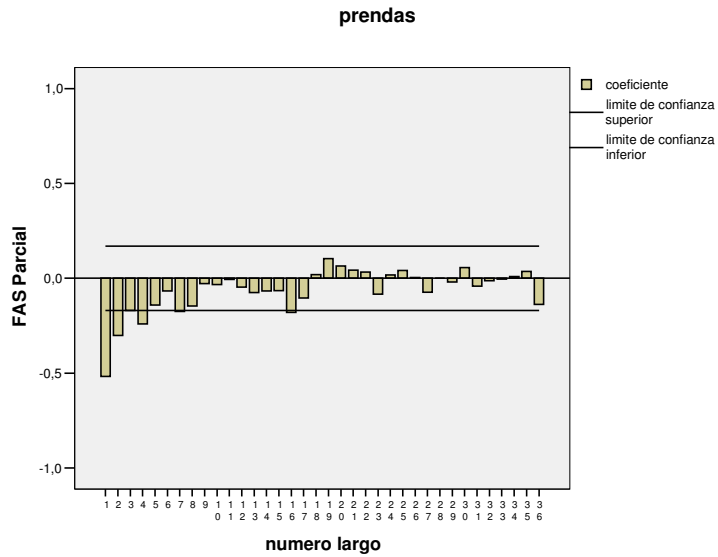


**Gráfico7.2.1.8: Autocorrelaciones Regulares con dos Diferenciaciones Regulares**

FAS

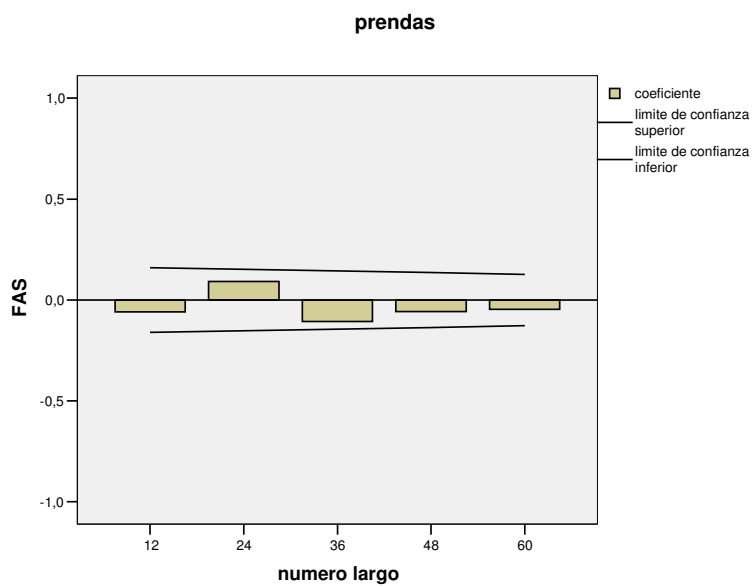


FAP

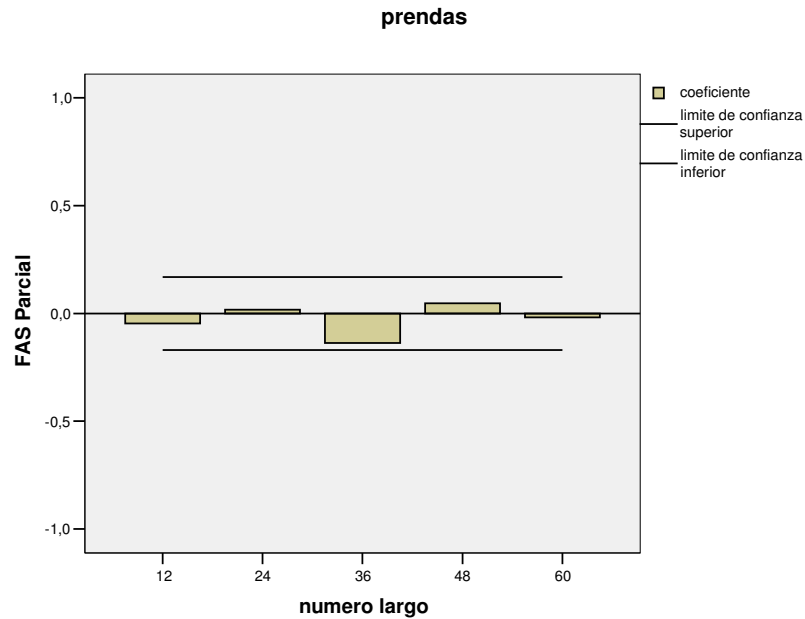


**Gráfico 7.2.1.9: Autocorrelaciones Estacionales con dos Diferenciaciones Regulares**

FAS



FAP



Con estos gráficos se comprueba lo que ya se decía que es suficiente diferenciar dos veces regularmente para lograr la estacionariedad en media. Ahora que se cumple la homoscedasticidad, se plantean los posibles modelos adecuados.

Del gráfico 7.2.1.8, los modelos identificados son:

*Modelo 1: IMA(2,1)*

$$(1-L)^2 y_t = (1-\theta_1 L) u_t$$

*Modelo 2: ARI(1,2)*

$$(1-\phi_1 L)(1-L)^2 y_t = u_t$$

*Modelo 3: ARI(2,2)*

$$(1-\phi_1 L - \phi_2 L^2)(1-L)^2 y_t = u_t$$

Ya planteados los modelos, se estiman los parámetros y se verifican los coeficientes para ver si cumplen la estacionariedad e invertibilidad, siendo significativos los coeficientes de los parámetros de cada modelo si, se cumple que el **t-valor** en valor absoluto es mayor que dos.

## ESTIMACIÓN DE PARÁMETROS

**Tabla 7.2.1.2: Modelo 1**

<b>Parametros Estimados</b>					
		Estimaciones	Error Std.	t	Approx Sig
no-estacional	MA1	,888	,043	20,426	,000
constante		-,004	,005	-,949	,344

**Tabla 7.2.1.3: Modelo 2**

<b>Parametros Estimados</b>					
		Estimaciones	Error Std	t	Approx Sig
no-estacional	AR1	-,516	,073	-7,091	,000
constante		-,002	,029		

**Tabla 7.2.1.4: Modelo 3**

<b>Parametros Estimados</b>					
		Estimaciones	Error Std	t	Approx Sig
no-estacional	AR1	-,673	,081	-8,269	,000
	AR2	-,300	,081	-3,691	,000
constante		-,001	,021	-,065	,948

Con estos resultados podemos decir que en ningún modelo se incorporara la constante por que su t-valor es inferior a dos. Veamos ahora la Estacionariedad e Invertibilidad de los parámetros (tabla 7.2.1.5).

**Tabla 7.2.1.5: Verificación de la Estacionariedad e Invertibilidad**

MODELO	ESTACIONARIEDAD	INVERTIBILIDAD
<p style="text-align: center;"><b>1</b> <i>IMA(2,1)</i></p>	<p>El modelo es un MA y una de las características de estos modelos es que siempre es Estacionario</p>	<p>Debe cumplirse que <math> \theta_1  &lt; 1</math>, es decir; <math> 0.888  &lt; 1</math>, el modelo es invertible.</p>
<p style="text-align: center;"><b>2</b> <i>ARI(1,2)</i></p>	<p>Debe cumplirse que <math> \phi_1  &lt; 1</math>, es decir; <math> -0.516  &lt; 1</math>, el modelo es estacionario.</p>	<p>El modelo es un AR una de las características de estos modelos es que siempre es Invertible</p>
<p style="text-align: center;"><b>3</b> <i>ARI(2,2)</i></p>	<p>Debe cumplirse que <math>\phi_1 + \phi_2 &lt; 1</math>, <math>\phi_2 - \phi_1 &lt; 1</math> y <math> \phi_2  &lt; 1</math>; dado que <math>\phi_1 = -0.673</math> y <math>\phi_2 = -0.300</math> el modelo es estacionario.</p>	<p>El modelo es un AR una de las características de estos modelos es que siempre es Invertible</p>

Según esta verificación todos los modelos siguen en el análisis, ninguno falla este criterio aunque la suma de las raíces del modelo 3, están muy cerca de uno.

En el análisis de los residuos se verifica que todos los modelos tengan media cero, varianza constante y que estén incorrelados. En este caso todos tienen media cero, la varianza constante, se cumple más para el modelo 1: *IMA(2,1)* de igual manera la condición de incorrelación. Para reafirmar observemos el grafico de secuencia y la FAS de los residuos de los tres modelos.

## Gráficos de Secuencias de los Residuos

Gráfico 7.2.1.10: Residuos del Modelo 1

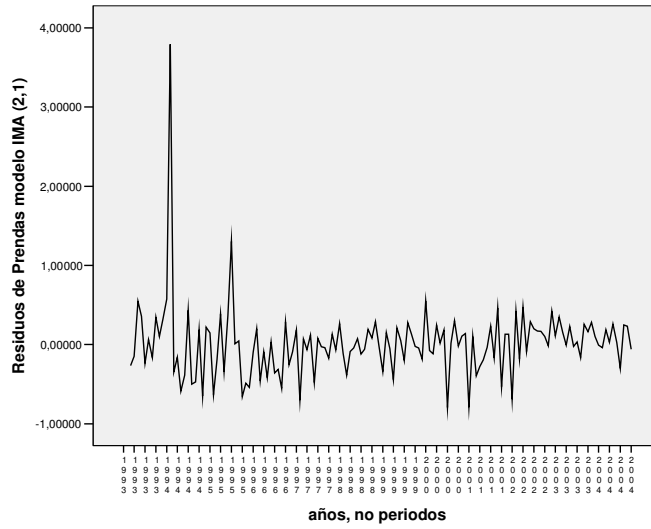
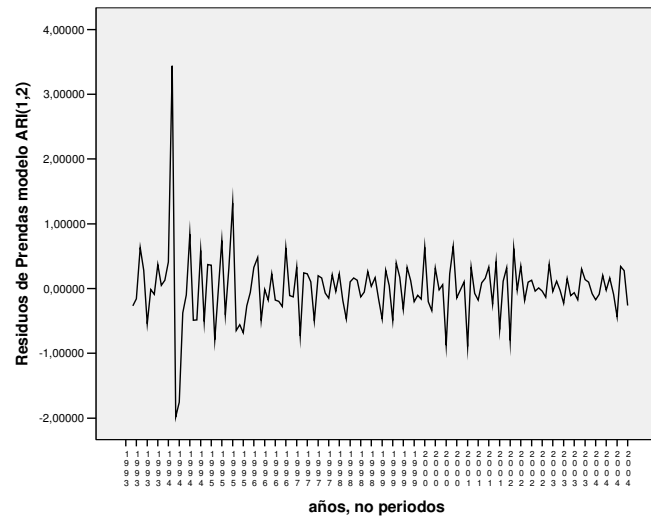
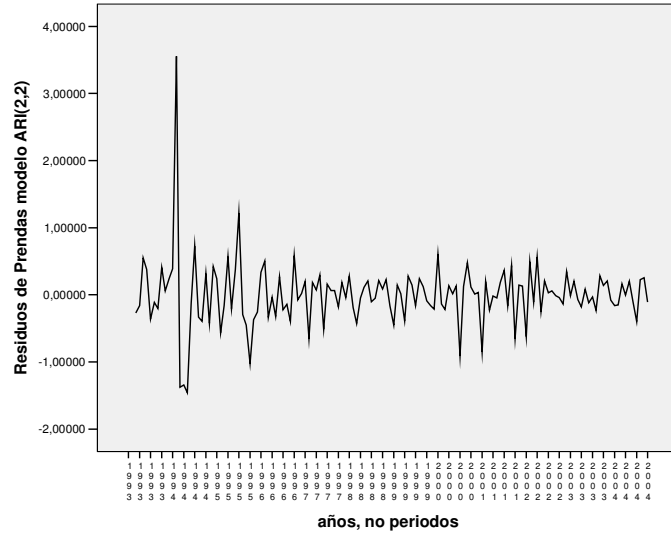


Gráfico 7.2.1.11: Residuos del Modelo 2

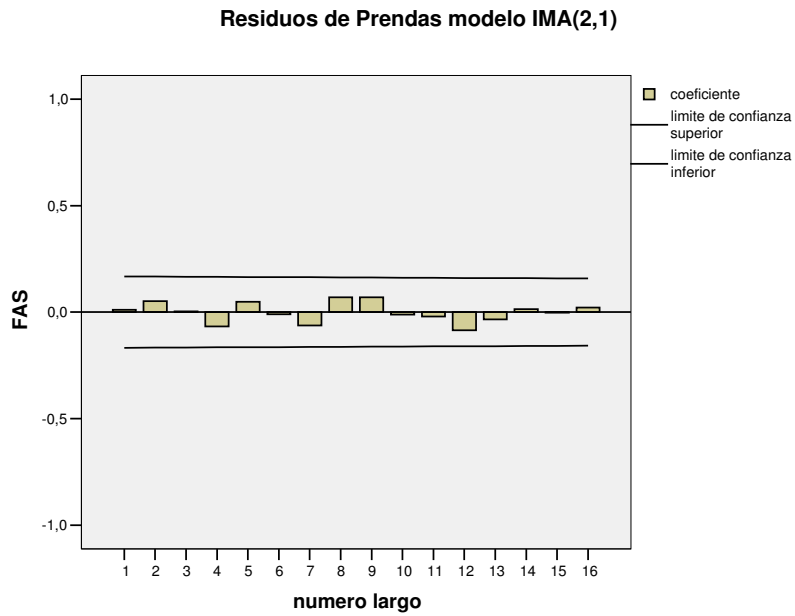


**Gráfico 7.2.1.12: Residuos del Modelo 3**

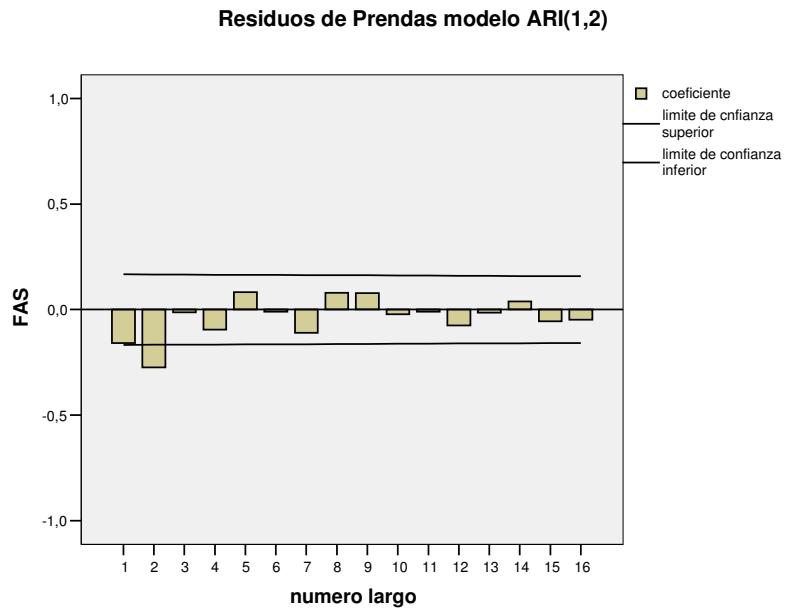


**Funciones de Autocorrelación FAS de los residuos**

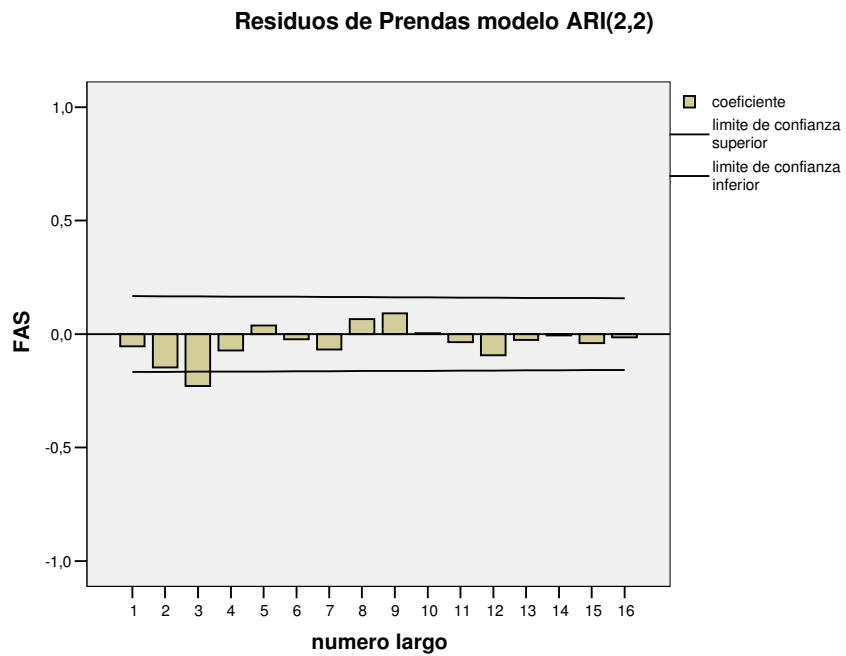
**Gráfico 7.2.1.13: FASR del Modelo1**



**Gráfico 7.2.1.14: FASR del Modelo 2**



**Gráfico 7.2.1.15: FASR del Modelo 3**





Podemos ver que el modelo 1, cumple mejor las condiciones de incorrelación y por lo tanto lo de media cero y varianza constante (en el gráfico de secuencia de los residuos se puede ver). Un segundo modelo es el modelo 3; para hacer una buena elección analicemos los modelos con los estadísticos AIC (El Criterio Akaike) y BIC (El criterio de información Bayesiana); así como también la calidad de las predicciones.

Con los estadísticos AIC y BIC se elegirá el modelo, que menor valor de estos tenga. Ver tabla 7.2.1.6.

**Tabla 7.2.1.6: Criterio AIC y BIC**

MODELO	AIC	BIC
1	205.270	210.998
2	213.376	219.095
3	206.922	215.502

El modelo que cumple con esta condición, es el mismo que se ha identificado en el análisis de residuo. Una vez más podemos ver que el modelo 3, difiere muy poco del modelo 1. Para mayor seguridad en la elección y tener buena calidad de las predicciones analicemos, la **estabilidad de los parámetros** para el modelo 1 y 3.

Un contraste estadístico que resulta útil para comprobar si las estimaciones de los coeficientes, obtenidos a partir del conjunto de datos, son significativamente diferentes, es el siguiente. Denominemos por  $\hat{\beta}_i^{(1)}$  y  $\hat{\beta}_i^{(2)}$  las estimaciones del coeficiente correspondientes  $\beta_i$ , respectivamente, a la primera y segunda mitad de la muestra. Entonces,

y dado que la desviación típica de  $\hat{\beta}_i^{(1)} - \hat{\beta}_i^{(2)}$  será igual a:

$$\left[ \text{var}\left(\hat{\beta}_i^{(1)}\right) + \text{var}\left(\hat{\beta}_i^{(2)}\right) \right]^{1/2}$$

Puede plantearse el siguiente estadístico:

$$\frac{\hat{\beta}_i^{(1)} - \hat{\beta}_i^{(2)}}{\left[ \text{var}\left(\hat{\beta}_i^{(1)}\right) + \text{var}\left(\hat{\beta}_i^{(2)}\right) \right]^{1/2}}$$

Que se distribuirá asintóticamente como una N (0,1). Así pues, fijado el nivel de significancia  $\varepsilon = 0.05$ , si el valor el valor del estadístico es inferior a 1.96 ( $\cong 2$ ) concluiremos que puede aceptarse la hipótesis de estabilidad.

El contraste propuesto, en realidad, solo es aplicable en el caso en que el modelo contenga un único coeficiente, o bien si todos los coeficientes tienen correlaciones prácticamente nulas. Dado que esta no es la situación más habitual, deberá pensarse en aplicar algún otro contraste. Como una aproximación puede utilizarse a este respecto el test de Chow (1960). La hipótesis nula a contrastar será la estabilidad de los coeficientes, es decir:

$$H_0 = \beta^{(1)} = \beta^{(2)} = \beta$$

Siendo  $\beta$  el vector columna de  $k$  parámetros del modelo, y denotando por (1) y (2) los correspondientes vectores de parámetros para cada una de las mitades de la muestra. Bajo esta hipótesis nula, el estadístico:

$$F = \frac{\left[ \frac{\sum_{t=1}^T \hat{a}_t^2 - \left( \sum_{t=1}^{T_1} \hat{a}_{1t}^2 + \sum_{t=1}^{T_2} \hat{a}_{2t}^2 \right)}{k} \right]}{\frac{\left( \sum_{t=1}^{T_1} \hat{a}_{1t}^2 + \sum_{t=1}^{T_2} \hat{a}_{2t}^2 \right)}{(T - 2k)}}$$

Se distribuye como un F de Snedecor con  $k$  y  $T-2k$  grados de libertad. Donde:

$k$ : es el número de parámetros del modelo.

$T_1$ : es el primer subconjunto de datos.

$T_2$ : es el segundo subconjunto de datos.

$T = T_1 + T_2$

$\hat{u}_t$ : residuo  $t$ -ésimo del modelo utilizado en la estimación todo el periodo muestral ( $T$ ).

$\hat{u}_{1t}$ : residuo  $t$ -ésimo del modelo utilizado en la estimación los  $T_1$  primeros datos.

$\hat{u}_{2t}$ : residuo  $t$ -ésimo del modelo utilizado en la estimación los  $T_2$  últimos datos.

La regla de actuación, fijado un nivel de significancia  $\varepsilon = 0.05$ , será:

- ✓ Si  $F < F_\varepsilon(k, T - 2k)$ , aceptamos la hipótesis nula de estabilidad.
- ✓ Si  $F > F_\varepsilon(k, T - 2k)$ , rechazamos la hipótesis nula de estabilidad, y, en consecuencia, el modelo estimado no resulta adecuado para predecir.

Ahora bien para el caso particular utilizaremos el estadístico que se distribuye como una normal para el modelo ARIMA (0,2,1); porque, solo tiene un parámetro y para contrastar la misma hipótesis con el modelo ARIMA (2,2,0), utilizaremos el estadístico F. Contrastando la hipótesis nula de estabilidad para el modelo 1, obtenemos:

$$\hat{\beta}_1^{(1)} = 1.00, \hat{\beta}_1^{(2)} = 0.882, \hat{\beta}_1^{(1)} - \hat{\beta}_1^{(2)} = 0.118$$

El valor del estadístico es:

$$\frac{\hat{\beta}_i^{(1)} - \hat{\beta}_i^{(2)}}{\left[ \text{var}\left(\hat{\beta}_i^{(1)}\right) + \text{var}\left(\hat{\beta}_i^{(2)}\right) \right]^{1/2}} = \frac{0.118}{(0.000196 + 0)^{1/2}} = 8.428$$

Como el valor del estadístico (8.428), es superior a 1.96 ( $\cong 2$ ) concluiremos que no puede aceptarse la hipótesis nula, razón, por la cual este modelo no puede seleccionarse como adecuado.

Para el modelo 3, los resultados son los siguientes:

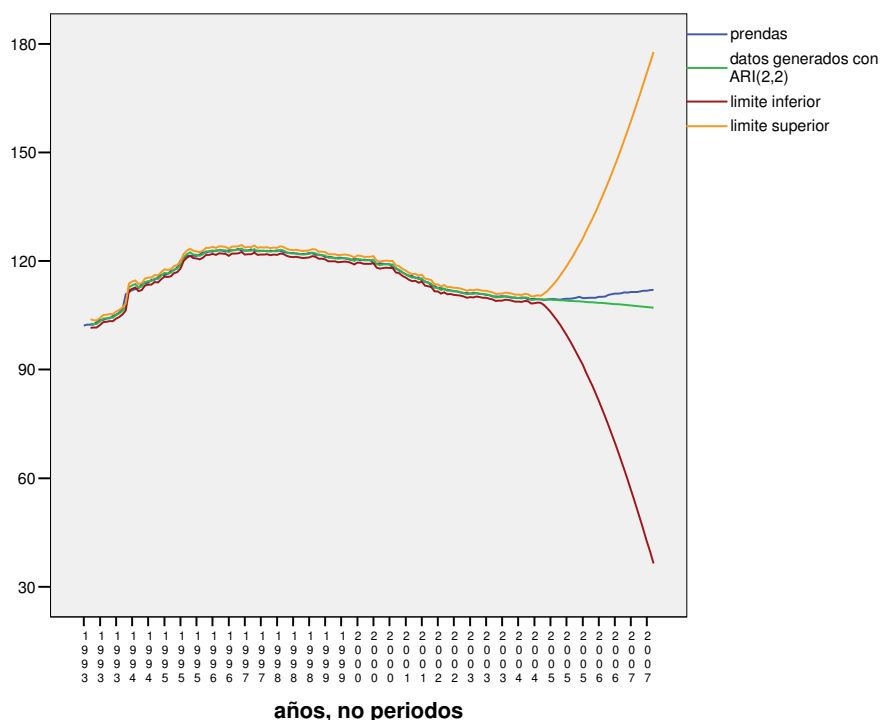
$$F = \frac{\left[ \frac{33.24974116 - (27.317273 + 5.23372996)}{2} \right]}{\frac{27.317273 + 5.23372996}{138}} = \frac{0.34936908}{0.235876833} = 1.48115$$

Por lo tanto aceptamos la hipótesis de estabilidad ya que  $F = 1.48115 < F_{0.05}(2, 138) = 3.00$ , y por consiguiente aceptamos este modelo como el modelo que se adecua a los datos, por consiguiente obtenemos los pronósticos de la tabla 7.2.1.7.

**Tabla 7.2.1.7: Pronósticos**

fecha	Serie	pronósticos modelo 3	limite inferior modelo 3	limite superior modelo 3
Nov-04	109.38	109.44591	108.46862	110.4232
Dic-04	109.21	109.41178	107.78215	111.0414
Ene-05	109.42	109.34906	106.93418	111.76393
Feb-05	109.54	109.31485	105.9379	112.69181
Mar-05	109.49	109.26731	104.85945	113.67517
Abr-05	109.46	109.21745	103.67902	114.75588
May-05	109.31	109.17042	102.40808	115.93276
Jun-05	109.5	109.11944	101.05443	117.18446
Jul-05	109.59	109.06754	99.61892	118.51616
Ago-05	109.55	109.01471	98.10601	119.92341
Sep-05	109.71	108.96005	96.51838	121.40173
Oct-05	109.91	108.90417	94.8581	122.95023
Nov-05	110.11	108.84692	93.12741	124.56643
Dic-05	109.79	108.78823	91.32804	126.24841
Ene-06	109.79	108.72818	89.46157	127.99479
Feb-06	109.86	108.66675	87.52944	129.80405
Mar-06	109.85	108.60392	85.53292	131.67492
Abr-06	109.91	108.53971	83.47315	133.60627
May-06	110.12	108.47412	81.3512	135.59704
Jun-06	110.11	108.40714	79.16801	137.64626
Jul-06	110.21	108.33877	76.92446	139.75308
Ago-06	110.63	108.26902	74.62137	141.91667
Sep-06	110.86	108.19788	72.25946	144.1363
Oct-06	111.03	108.12536	69.83945	146.41127
Nov-06	111.04	108.05145	67.36195	148.74095
Dic-06	111.13	107.97616	64.82758	151.12474
Ene-07	111.34	107.89948	62.23688	153.56208
Feb-07	111.28	107.82141	59.59038	156.05245
Mar-07	111.43	107.74196	56.88856	158.59537
Abr-07	111.43	107.66113	54.13187	161.19038
May-07	111.47	107.57891	51.32076	163.83705
Jun-07	111.65	107.4953	48.45562	166.53498
Jul-07	111.81	107.41031	45.53684	169.28377
Ago-07	111.84	107.32393	42.56478	172.08308
Sep-07	111.96	107.23617	39.53977	174.93256
Oct-07	112.04	107.14702	36.46214	177.83189

**Gráfico7.2.1.16: Serie de los Pronósticos y Datos Originales**



Podemos observar que la calidad de los pronósticos, es buena ya que los datos originales y los pronosticados caen dentro de los intervalos de confianza; permitiéndonos decir que el modelo se adecua a los datos. La expresión matemática del modelo es:

$$(1 + 0.673L + 0.300L^2)(1 - L)^2 y_t = u_t$$

Sin embargo, al efectuar la diagnosis del modelo y, en concreto, al analizar el correlograma muestral residual, observamos que aparecían elevados valores de algunos coeficientes de autocorrelación muestral residual, así como de autocorrelación parcial muestral residual, por lo que apuntamos la posibilidad de que existan ciertos atípicos para comprobarlo escudriñamos la serie con el **Procedimiento de detección de outliers y estimación conjunta de los parámetros y los efectos de los outliers**, desarrollado por Chen y Liu y disponible en el programa TRAMO<sup>8</sup>.

Antes de empezar a obtener resultados es necesario decir que este software permite detectar los outliers en sus localizaciones, estimar conjuntamente los parámetros y los efectos de los outliers, mostrando el modelo final y la serie perturbada con los efectos de los datos atípicos junto con su análisis de los residuos en una sola vuelta. Esto dejando que el modelo ARIMA inicial sea Automático (por defecto toma un  $ARIMA(0,1,1) * ARIMA(0,1,1)_{12}$ ), de igual forma las transformaciones.

<sup>8</sup> Revisar algoritmo de TRAMO en <http://www.spgg.pap.meh.es>

La desventaja de dar un modelo inicial radica en que los residuos no cumplan los criterios establecidos (por ejemplo: la incorrelación, la normalidad), viéndose en el caso de tomar la nueva serie resultante perturbada por los efectos de los outliers que a detectado, y volver a calcular un modelo ARIMA. Este proceso es tedioso; pararía hasta que se encuentre un modelo que permita que los residuos de la nueva serie pasen los criterios de normalidad e incorrelación.

No precisamente esta situación sucederá cada vez que el usuario introduzca un modelo inicial, puede ser, que de una vez, los residuos sobrepasen las condiciones; es decir el modelo detecta todos los outliers de una vez y hace buena estimación de los parámetros.

Incorporar este análisis a la serie es importante, por cuanto Chen, Liu y Hudak (1990), dicen que un tratamiento adecuado de los outliers permite mejorar:

1. **La compresión de la serie objeto de análisis**, dado que la detección de outliers puede suministrar información respecto a la ocurrencia de determinados factores externos que afectan a la serie, señalando tanto el momento en que se producen como la manera en que actúan sobre la misma. Esto permite no solo comprender mejor la evolución de la serie, sino también descubrir observaciones espúreas (por ejemplo, errores en la transcripción de los datos) que pueden “oscurecer” la propia modelización de una serie temporal.
2. **La modelización y estimación**, en tanto que sucesos externos desconocidos pueden alterar la estructura de los estadísticos utilizados habitualmente en la etapa de identificación; así como, aun cuando se efectuó una adecuada identificación del proceso que subyace a la serie temporal, las estimaciones de los parámetros.
3. **El análisis de intervención**, dado que si utilizamos un modelo de intervención debemos tener seguridad de que los efectos de las intervenciones que especifiquemos no están contaminados por los efectos de observaciones atípicas.
4. **La calidad de las predicciones**, debe tenerse en cuenta que dependiendo del momento y la naturaleza del suceso externo que origina el outliers, este puede afectar seriamente la calidad de las predicciones que se obtienen desde el modelo identificado. Ajustando la presencia de los outliers podemos, en consecuencia, mejorar la calidad de las predicciones que se obtengan.

Buscaremos detectar los datos atípicos (outliers) con el modelo inicial  $(1 + 0.673L + 0.300L^2)(1 - L)^2 y_t = u_t$ , por ser el que se ajusta mejor a la serie prendas de vestir y calzado.

Fijando el valor de la constante positiva  $c = 3.3$  (valor automático para el programa). Cabe decir que  $c$  constituye el punto crítico para aceptar o rechazar la hipótesis nula de ausencia de outliers, se recomienda a partir de determinados experimentos de simulación que debe tomar valores comprendidos entre 3 y 4.5. Lo habitual es tomar el valor 3.

Sin hacer transformación logarítmica y corrección a la media, obtenemos los atípicos que se recogen en la tabla 7.2.1.8.

**Tabla 7.2.1.8: Detección de outliers y estimación conjunta de los parámetros**

Parámetro/Período	Estimación	T	Tipo
$\phi_1$	0.6178	12049.82	-
$\phi_2$	0.169	3460.6	-
$\sigma_u^2$	0.0277	-	-
26	-3.5802	-13.69	LS
15	-3.41	-8.24	IO
16	-1.9469	-5.51	IO
31	1.4196	6.28	LS
17	-1.5674	-4.53	IO
38	-2.4166	-5.57	IO
91	-0.8671	-4.16	LS
33	0.66576	4.7	AO
85	1.0504	5.03	LS
50	-0.7973	-4.36	TC
18	-0.63871	-4.21	AO

El modelo especificado con los once outliers detectados es:

$$\begin{aligned}
 (1 + \phi_1 L + \phi_2 L^2)(1 - L)^2 y_t = & \frac{w_1}{(1 - L)} I_{1t}^{t_0=26} + w_2 I_{2t}^{t_0=15} + w_3 I_{3t}^{t_0=16} + \frac{w_4}{(1 - L)} I_{4t}^{t_0=31} + \\
 & w_5 I_{5t}^{t_0=17} + w_6 I_{6t}^{t_0=38} + \frac{w_7}{(1 - L)} I_{7t}^{t_0=91} + w_8 I_{8t}^{t_0=33} + \frac{w_9}{(1 - L)} I_{9t}^{t_0=85} + \frac{w_{10}}{(1 - 0.7L)} I_{10t}^{t_0=50} + \\
 & w_{11} I_{11t}^{t_0=18} + u_t
 \end{aligned}$$

donde:

$$\begin{aligned}
 I_{1t}^{t_0=26} &= \begin{cases} 1 & \text{para } t \geq 26 \\ 0 & \text{para } t < 26 \end{cases} & I_{2t}^{t_0=15} &= \begin{cases} 1 & \text{para } t = 15 \\ 0 & \text{para } t \neq 15 \end{cases} \\
 I_{3t}^{t_0=16} &= \begin{cases} 1 & \text{para } t = 16 \\ 0 & \text{para } t \neq 16 \end{cases} & I_{4t}^{t_0=31} &= \begin{cases} 1 & \text{para } t \geq 31 \\ 0 & \text{para } t < 31 \end{cases} \\
 I_{5t}^{t_0=17} &= \begin{cases} 1 & \text{para } t = 17 \\ 0 & \text{para } t \neq 17 \end{cases} & I_{6t}^{t_0=38} &= \begin{cases} 1 & \text{para } t = 38 \\ 0 & \text{para } t \neq 38 \end{cases} \\
 I_{7t}^{t_0=91} &= \begin{cases} 1 & \text{para } t \geq 91 \\ 0 & \text{para } t < 91 \end{cases} & I_{8t}^{t_0=33} &= \begin{cases} 1 & \text{para } t = 33 \\ 0 & \text{para } t \neq 33 \end{cases} \\
 I_{9t}^{t_0=85} &= \begin{cases} 1 & \text{para } t \geq 85 \\ 0 & \text{para } t < 85 \end{cases} & I_{10t}^{t_0=50} &= \begin{cases} 1 & \text{para } t = 50 \\ 0 & \text{para } t \neq 50 \end{cases} \\
 I_{11t}^{t_0=18} &= \begin{cases} 1 & \text{para } t = 18 \\ 0 & \text{para } t \neq 18 \end{cases}
 \end{aligned}$$

Y los *outliers* corresponden a las siguientes fechas: la observación 26 es febrero de 1995, la 15 es marzo de 1994, la 16 es abril de 1994, la 31 es julio de 1995, la 17 es mayo de 1994, la 38 es febrero de 1996, la 91 es julio de 2000, la 33 es septiembre de 1995, la 85 es enero de 2000, la 50 es febrero de 1997 y la 18 es junio de 1994. Los resultados obtenidos al estimar por máxima verosimilitud exacta el modelo anterior, concentrando fuera de la verosimilitud los parámetros de la regresión, son los siguientes:

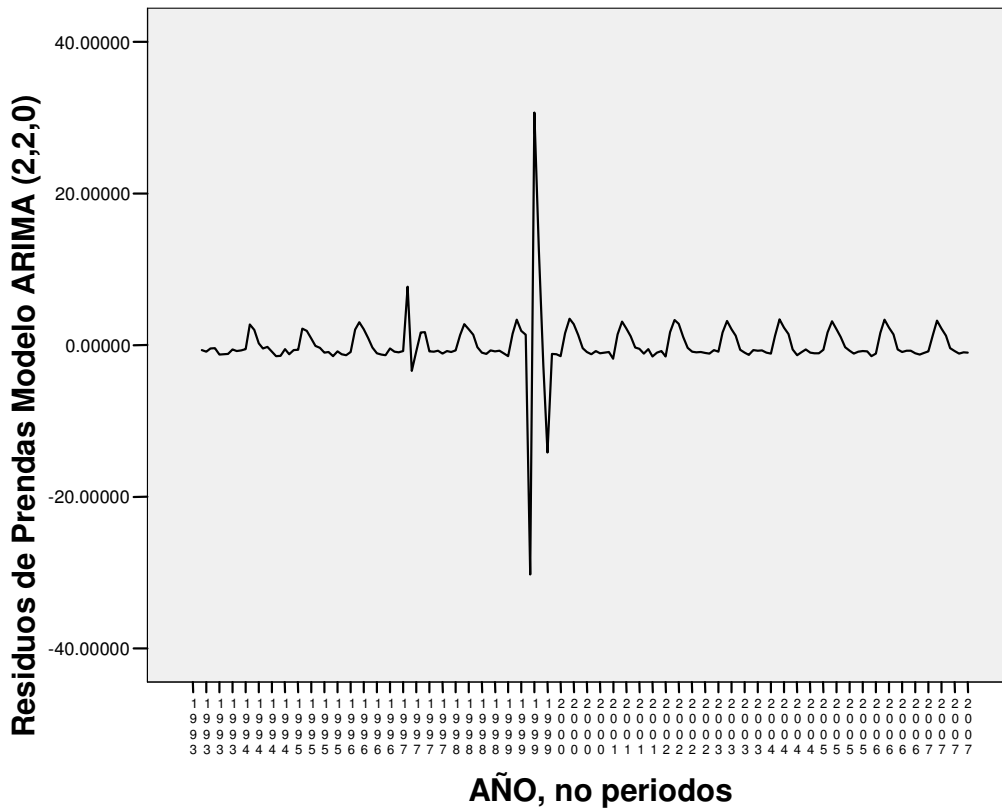
El período muestral analizado es el comprendido entre enero de 1993 y octubre de 2007. El error estándar estimado de los residuos es: 0.3462, siendo inferior al de otros modelos alternativos, que se han considerado en el proceso de identificación (por ejemplo el modelo identificado sin outliers su error estándar es: 2.197). La media estimada de la serie de residuos es cero (el valor del estadístico  $t = 0.6653$  es muy inferior al valor crítico 2), el test de normalidad no permite rechazar la hipótesis nula de normalidad de los mismos (0.2288 frente a un valor crítico al 95 por ciento de 5.99).

El estadístico de *Ljung — Box* de orden 24 es 41.63 no permite aceptar la hipótesis nula de que las correlaciones son cero, por ser superior al valor crítico ( $X_{0.05}^2(22) = 33.924$ ).

Veamos también gráficamente el comportamiento de estos residuos:



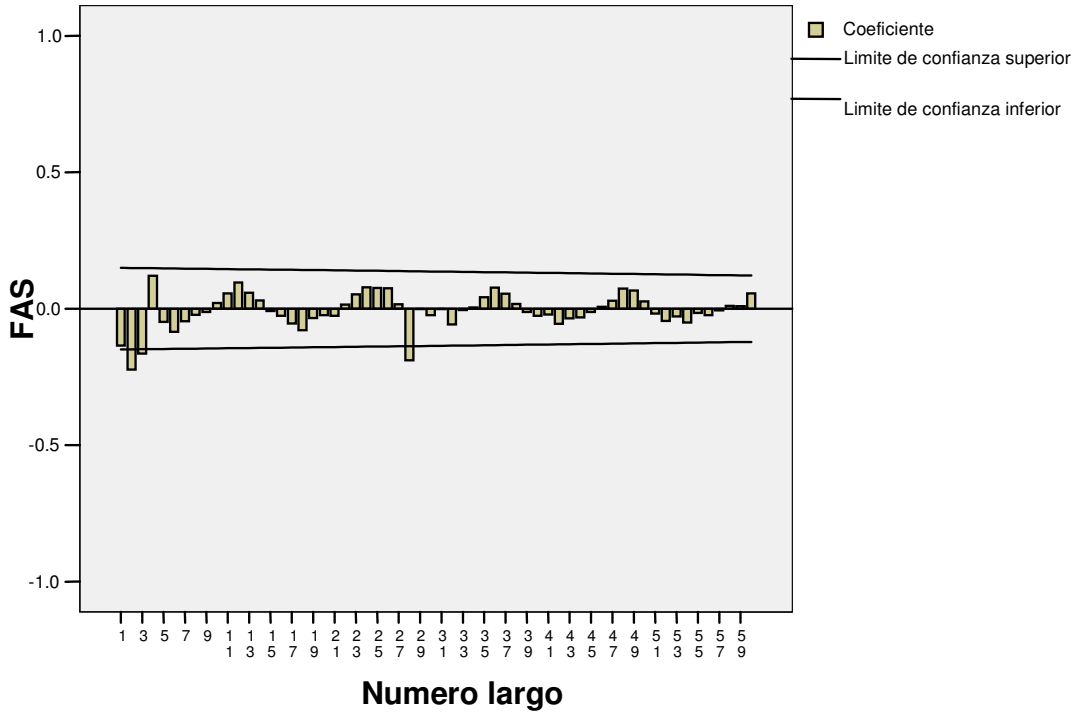
**Gráfico 7.2.1.17: Residuos del modelo estimado con 11 outliers**



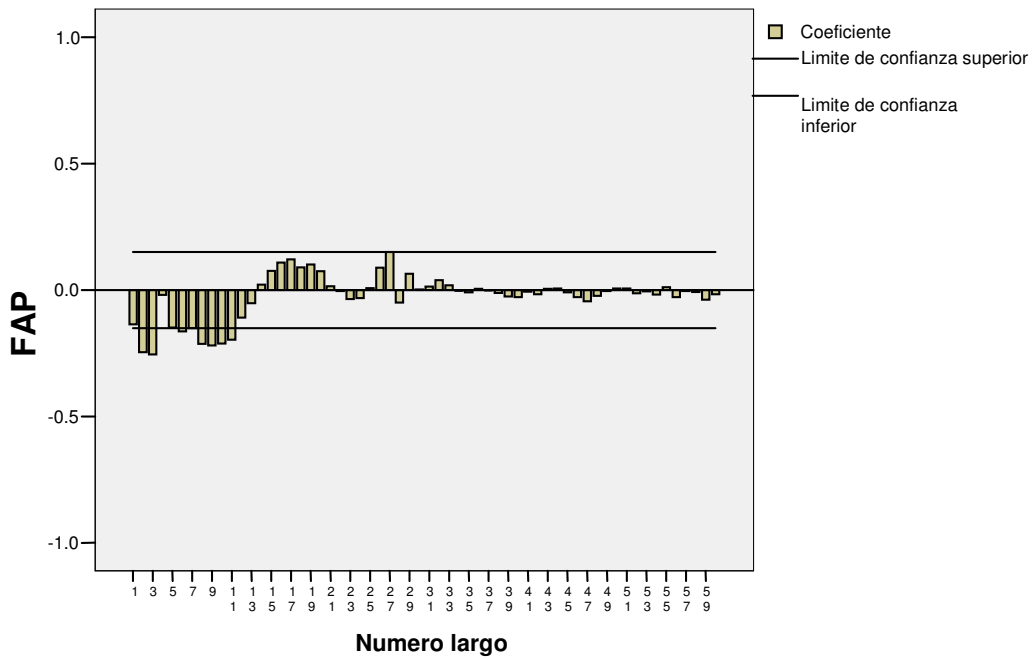
A partir de este gráfico no se observa presencia de heteroscedasticidad. Por otra parte, la serie de residuos parece oscilar en torno en un valor medio próximo a cero, por lo que podríamos aceptar que la media de dicho residuo es cero, esto confirma lo del test de normalidad mencionado ( $0.6653 < 2$ ).

Finalmente veamos la representación de los gráficos 7.2.1.18 y 7.2.1.19, que muestran las autocorrelaciones FASMR (función de autocorrelación simple muestral residual) Y FAPMR (función de autocorrelación parcial muestral residual), en los que se observa que la condición de incorrelación no se cumple; indicando que se tendría que repetir el proceso, empezando por encontrar un modelo ARIMA a la serie perturbada por los 11 outliers que se han detectado; este nuevo modelo será el que se tome como inicial para hacer la nueva detección y estimación de outliers. Como ya se dijo es un trabajo engorroso; pero es el algoritmo que TRAMO sigue ocultamente al seleccionar el ARIMA automáticamente.

**Gráfico 7.2.1.18: FASMR del modelo estimado con 11 outliers**



**Gráfico 7.2.1.19: FAPMR del modelo estimado con 11 outliers**



Tomando en cuenta esto, dejamos que el programa trabaje automáticamente y nos de el modelo que cumpla con todos los criterios requeridos.

El modelo ARIMA estimado final, que encuentra el programa es un:

$ARIMA(1,2,2) = (1 - \phi_1 L)(1 - L)^2 y_t = (1 - \theta_1 L - \theta_2 L^2) + u_t$ , por medio del cual se obtiene la detección de los outliers y estimación conjunta de los parámetros y efectos de dichos atípicos que se recogen en la tabla 7.2.1.9.

**Tabla 7.2.1.9: Detección de outliers y estimación conjunta de los parámetros automáticamente**

<i>Parámetro/Período</i>	<i>Estimación</i>	<i>t</i>	<i>tipo</i>
$\phi_1$	-0.53794	-3.33	-
$\theta_1$	-1.6209	-12.43	-
$\theta_2$	0.71602	6.31	-
$\sigma_u^2$	0.0177	-	-
14	3.9839	18.66	LS
31	1.7103	8.03	LS
91	-0.82725	-3.54	IO
97	-0.74517	-3.19	IO
109	-0.63017	-3.13	TC

Permitiendo formularse el modelo con los cinco outliers detectados de la forma siguiente:

$$(1 - \phi_1 L)(1 - L)^2 y_t = \frac{w_1}{(1 - L)} I_{1t}^{t_0=14} + \frac{w_2}{(1 - L)} I_{2t}^{t_0=31} + w_3 (1 - \theta_1 L - \theta_2 L^2) I_{3t}^{t_0=91} \\ + w_4 (1 - \theta_1 L - \theta_2 L^2) I_{4t}^{t_0=97} + \frac{w_5}{(1 - 0.7L)} I_{5t}^{t_0=109} + (1 - \theta_1 L - \theta_2 L^2) u_t$$

donde:

$y_t$ : representa la variable prendas de vestir y calzado

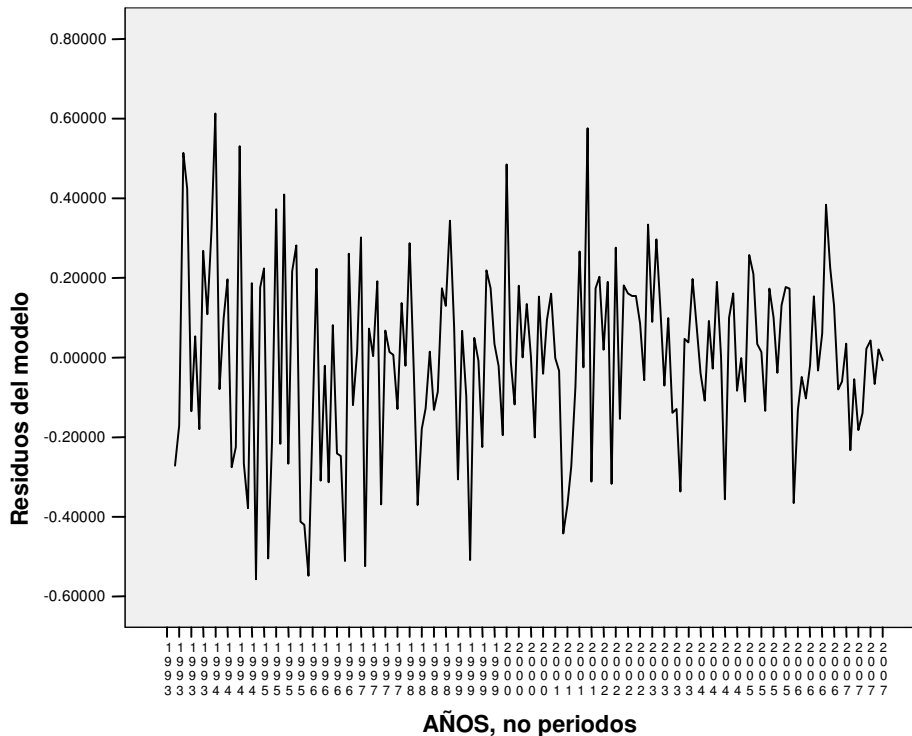
$$I_{1t}^{t_0=14} = \begin{cases} 1 & \text{para } t \geq 14 \\ 0 & \text{para } t < 14 \end{cases} \quad I_{2t}^{t_0=31} = \begin{cases} 1 & \text{para } t \geq 31 \\ 0 & \text{para } t < 31 \end{cases}$$

$$I_{3t}^{t_0=91} = \begin{cases} 1 & \text{para } t = 91 \\ 0 & \text{para } t \neq 91 \end{cases} \quad I_{4t}^{t_0=97} = \begin{cases} 1 & \text{para } t = 97 \\ 0 & \text{para } t \neq 97 \end{cases}$$

$$I_{5t}^{t_0=109} = \begin{cases} 1 & \text{para } t = 109 \\ 0 & \text{para } t \neq 109 \end{cases}$$

Los *outliers* corresponden a las siguientes fechas: la observación 14 es febrero de 1994, la 31 es julio de 1995, la 91 es julio de 2000, la 97 es enero de 2001 y la 109 es enero de 2002. El análisis de los residuos del modelo estimado lo empezaremos obteniendo sus representaciones gráficas. Tal como se exponen a continuación:

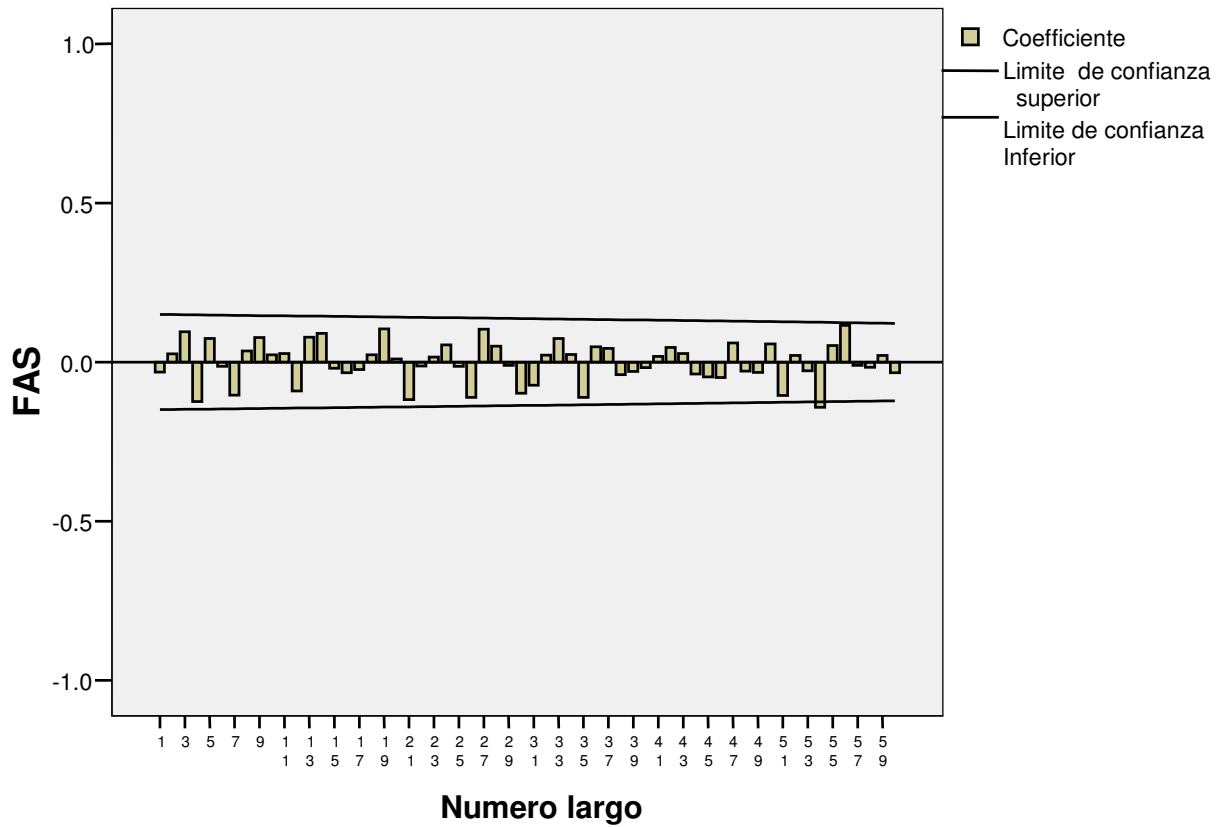
**Gráfico 7.2.1.20: Residuos del modelo estimado con cinco outliers**



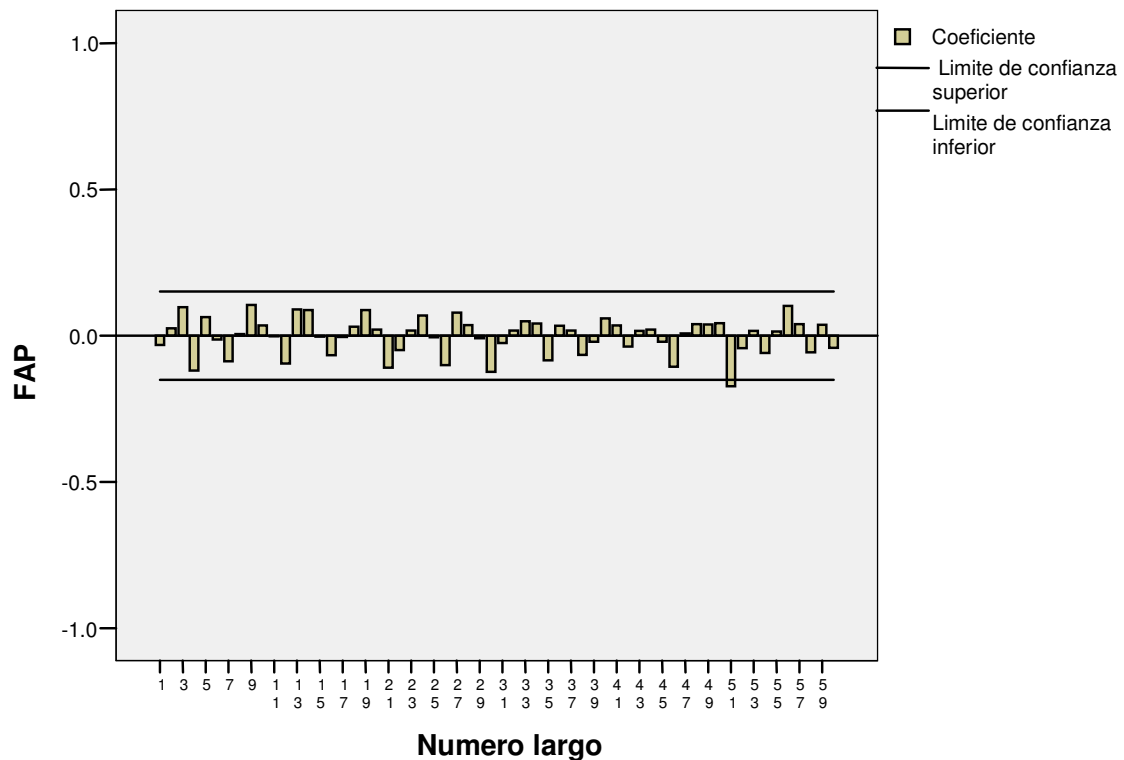
Obsérvese que los residuos no presentan heteroscedasticidad según el gráfico 7.2.1.20 y, por otra parte parecen oscilar en torno a un valor medio próximo a cero, lo que nos lleva a aceptar que la media de los residuos es cero y específicamente así es, el valor del estadístico  $t = -0.1228$  es muy inferior al valor crítico 2.

El test de normalidad no permite rechazar la hipótesis nula de normalidad de los residuos (1.647 frente a un valor crítico al 95 por ciento de 5.99), el estadístico  $Q^*$  de Ljung-Box de orden 24 es 21.27 permite aceptar la hipótesis de incorrelación por ser inferior al valor crítico ( $X^2_{0.05}(21) = 32.671$ ). Además se puede ver mejor en los gráficos de las autocorrelaciones siguientes:

**Gráfico 7.2.1.21: FASMR del modelo estimado con 5 outliers**



**Gráfico 7.2.1.22: FAPMR del modelo estimado con 5 outliers**



Vemos que las autocorrelaciones de los residuos si se mantienen incorreladas.

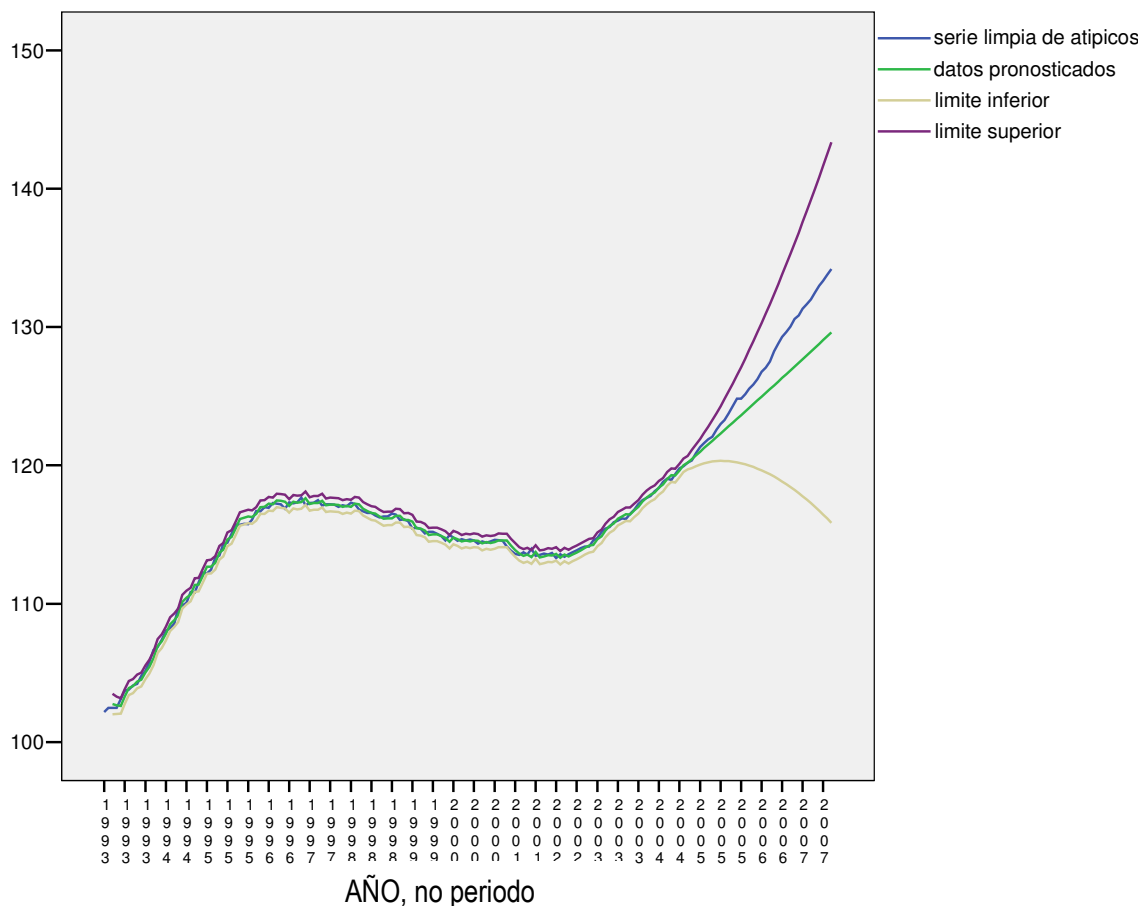
El error estándar estimado de los residuos es: 0.233, siendo inferior al de otros modelos alternativos que se han considerado en el proceso de identificación (por ejemplo el modelo identificado con 11 outliers su error estándar es 0.3462). Permitiendo concluir que el modelo que se ha calculado es bueno, ya que los residuos pasan todos los criterios aquí analizados.

Para finalizar, comparemos los pronósticos que se generan con el modelo y los datos de la serie perturbada, por los efectos de los outliers que se detectaron para los cuales se desconoce la causa externa del suceso de ocurrencia.

**Tabla 7.2.1.10: Comparación de pronósticos y datos originales**

fecha	serie	pronósticos	limite inferior	limite superior	error
Nov-04	120.197	120.19685	119.70595	120.68775	0.00015
Dic-04	120.351	120.45953	119.80868	121.11037	-0.10853
Ene-05	120.885	120.72183	119.93442	121.50925	0.16317
Feb-05	121.329	120.98418	120.05147	121.91689	0.34482
Mar-05	121.602	121.24677	120.14957	122.34396	0.35523
Abr-05	121.896	121.50971	120.22575	122.79367	0.38629
May-05	122.07	121.77306	120.28005	123.26607	0.29694
Jun-05	122.584	122.03685	120.31369	123.76	0.54715
Jul-05	122.998	122.30108	120.32815	124.27401	0.69692
Ago-05	123.282	122.56577	120.32485	124.8067	0.71623
Sep-05	123.766	122.83093	120.30501	125.35684	0.93507
Oct-05	124.29	123.09654	120.26967	125.92341	1.19346
Nov-05	124.814	123.36262	120.21966	126.50558	1.45138
Dic-05	124.818	123.62916	120.15568	127.10264	1.18884
Ene-06	125.141	123.89616	120.0783	127.71403	1.24484
Feb-06	125.535	124.16363	119.98802	128.33925	1.37137
Mar-06	125.849	124.43156	119.88524	128.97789	1.41744
Abr-06	126.233	124.69996	119.77031	129.62961	1.53304
May-06	126.767	124.96882	119.64354	130.2941	1.79818
Jun-06	127.081	125.23814	119.5052	130.97109	1.84286
Jul-06	127.505	125.50793	119.35553	131.66034	1.99707
Ago-06	128.249	125.77819	119.19472	132.36165	2.47081
Sep-06	128.803	126.0489	119.02298	133.07483	2.7541
Oct-06	129.297	126.32008	118.84047	133.7997	2.97692
Nov-06	129.63	126.59173	118.64734	134.53611	3.03827
Dic-06	130.044	126.86383	118.44374	135.28393	3.18017
Ene-07	130.578	127.13641	118.22979	136.04302	3.44159
Feb-07	130.842	127.40944	118.00562	136.81327	3.43256
Mar-07	131.316	127.68294	117.77133	137.59456	3.63306
Abr-07	131.64	127.95691	117.52703	138.38679	3.68309
May-07	132.004	128.23134	117.27281	139.18987	3.77266
Jun-07	132.508	128.50623	117.00876	140.0037	4.00177
Jul-07	132.992	128.78159	116.73496	140.82822	4.21041
Ago-07	133.345	129.05741	116.45148	141.66333	4.28759
Sep-07	133.789	129.33369	116.15841	142.50897	4.45531
Oct-07	134.193	129.61044	115.85581	143.36507	4.58256

**Gráfico 7.2.1.23: Pronósticos y datos reales**



Si comparamos el modelo ARIMA (1, 2,2) con cinco outliers seleccionado, con el modelo ARIMA (1, 2,2) sin outliers, se comprueba que el tratamiento de los atípicos mejora sustancialmente la capacidad predictiva del modelo.

Es de importancia decir que el modelo que sea ha detectado con outliers se adecua a los datos de la serie que disponíamos de entrada. Aunque podría preguntarse por que dicho modelo, no se planteo en el análisis ARIMA; la explicación que cabe es que las autocorrelaciones eran afectadas por los outliers y no permitían identificarlo.

Por lo tanto el modelo final que se ajusta adecuadamente a la serie *Índices de Precios al Consumidor en la Variable Prendas de Vestir y Calzado*, matemáticamente se expresa así:



$$(1+0.5379L)(1-L)^2 y_t = \frac{3.9839}{(1-L)} I_{1t}^{t_0=14} + \frac{1.7103}{(1-L)} I_{2t}^{t_0=31} - 0.82725(1+1.6209L-0.71605L^2) I_{3t}^{t_0=91} - 0.7451(1+1.6209L-0.71605L^2) I_{4t}^{t_0=97} + \frac{(-0.6301)}{(1-0.7L)} I_{5t}^{t_0=109} + (1+1.6209L-0.71605L^2) \mu_t$$

donde:

$y_t$ : representa la serie prendas de vestir y calzado

$$I_{1t}^{t_0=14} = \begin{cases} 1 & \text{para } t \geq 14 \\ 0 & \text{para } t < 14 \end{cases} \quad I_{2t}^{t_0=31} = \begin{cases} 1 & \text{para } t \geq 31 \\ 0 & \text{para } t < 31 \end{cases}$$

$$I_{3t}^{t_0=91} = \begin{cases} 1 & \text{para } t = 91 \\ 0 & \text{para } t \neq 91 \end{cases} \quad I_{4t}^{t_0=97} = \begin{cases} 1 & \text{para } t = 97 \\ 0 & \text{para } t \neq 97 \end{cases}$$

$$I_{5t}^{t_0=109} = \begin{cases} 1 & \text{para } t = 109 \\ 0 & \text{para } t \neq 109 \end{cases}$$

Este modelo es el que permitirá que se conciban buenas predicciones, con respecto al Índice de Precios al Consumidor en las Prendas de Vestir y Calzado.

## 8. CONCLUSIONES

- ❖ Al introducir modelos de intervención a nuestro análisis, llegamos a obtener las mejores predicciones ya que se toma en cuenta todo lo ocurrido en la trayectoria pasada de la serie, muchas veces esta trayectoria es intervenida con algún suceso externo, que distorsiona las autocorrelaciones de los residuos y no nos permiten aceptar un buen Modelo generador de la serie.
  
- ❖ Aplicar el análisis de datos atípicos, a una serie que presenta observaciones espurias y que se desconoce la causa de los factores externos que las han creado, es ventajoso, por que al obviar la presencia de dichos datos, puede ocasionar que:
  - ✓ El modelo no se especifique adecuadamente ó
  - ✓ Que incluso si el modelo ARIMA esta adecuadamente especificado; los datos atípicos pueden producir sesgos importantes en los estimadores de los parámetros.

Mientras que al incluirlos en el análisis nos, permite ajustar un modelo a la serie perturbada por los efectos de los outliers que se detecten y además facilita la revisión a los datos originales; es decir que sabiendo la localización de un dato atípico se puede comprobar si fue una mala digitación o si en realidad en ese período se dio un suceso; en caso de que se compruebe esto último, podríamos hacer un nuevo análisis pero con intervención. En resumen se dice que este análisis permite de forma fácil limpiar una serie de tiempo, y por lo tanto mejorar la calidad de las predicciones, en comparación con el modelo ARIMA convencional.

## 9. RECOMENDACIONES

- ❖ Sería conveniente, que siempre que se analice una serie se aplique el análisis de intervención u outliers, aunque se considere que se ha especificado correctamente un modelo ARIMA; ya que hay series que tienen datos atípicos no aislados, obstaculizando esto la calidad de las predicciones.
- ❖ A pesar de que la calidad de las predicciones de un modelo con intervenciones u outliers, mejora mucho en comparación a las de un modelo ARIMA ordinario; se espera que se han mas eficaces, al expandirlos como una función de transferencia, en vista de que se pueden incorporar otras variables (series)

## 10. REFERENCIAS BIBLIOGRÁFICAS

A continuación se muestra la documentación y paginas Web revisadas para realizar este trabajo de tesis:

- ✓ Aznar, Antonio y Fransisco Jabier Trivez; Métodos de Predicción en Economía (II) (1993). Editorial Ariel, S. A. Barcelona.
- ✓ Bono. R, Universidad de Barcelona. Diseño de Series Temporales Interrumpidas en la evaluación de programas, en La Universidad de Barcelona. (Archivo en formato PDF)  
[http://www.ugr.es/~cmetodo/pdf/simposio/simposio\\_arnau.pdf](http://www.ugr.es/~cmetodo/pdf/simposio/simposio_arnau.pdf)
- ✓ Espasa, Antoni y José Ramón Cancelo; Métodos Cuantitativos para el Análisis de la Coyuntura Económica (1993), Alianza editorial, Madrid.
- ✓ Galiano, San Miguel Pedro (2003), Universidad III de Madrid, facultad de ciencias sociales y jurídicas.  
[http://www.cibernetia.com/tesis\\_es/MATEMATICAS/ESTADISTICA/SERIES\\_TEMPORALES/1](http://www.cibernetia.com/tesis_es/MATEMATICAS/ESTADISTICA/SERIES_TEMPORALES/1)
- ✓ Meliá, J.L. (1995). Un proceso de intervención para reducir los accidentes laborales [An intervention process to reduce work-related accidents]. *Revista de Psicología del Trabajo y de las Organizaciones*, 11 (32), 51-65.  
<http://www.uv.es/~meliajl/Segur/A65Intervhtml/A65.htm>
- ✓ Outlier detection and plus/minus 3-sigma charts.  
<http://www.autobox.com/outlier.html>
- ✓ Pankratz, Alan; Forecasting with Dynamic Regression Model (1991), New York.
- ✓ Universidad Carlos III de Madrid.  
[http://www.est.uc3m.es/esp/nueva\\_docencia/getafe/ciencias\\_estadisticas/series\\_temporales/practicas/notaseviews6.pdf](http://www.est.uc3m.es/esp/nueva_docencia/getafe/ciencias_estadisticas/series_temporales/practicas/notaseviews6.pdf)
- ✓ Víctor Gómez, David Taguas; Detección y corrección automática de outliers con tramo: una aplicación al ipcde bienes industriales no energéticos (General de Presupuestos y Gastos del Ministerio de Economía y Hacienda de España.) Noviembre de 1995.  
<http://www.sgpg.pap.meh.es>
- ✓ W. S., William; Time Series Analysis Univariate and Multivariate Methods (1990), Ciudad de la Secoya, California.

## ANEXOS

### Serie del IVAE, en Comidas, Restaurantes y Hoteles

fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-90	105,01	ene-91	105,22	ene-92	120,87	ene-93	133,72	ene-94	143,68
feb-90	95,58	feb-91	96,35	feb-92	115,23	feb-93	129,65	feb-94	143,18
mar-90	103,79	mar-91	100,44	mar-92	126,41	mar-93	142,75	mar-94	161,62
abr-90	93,18	abr-91	100,07	abr-92	124,58	abr-93	135,04	abr-94	146,58
may-90	109,69	may-91	116,24	may-92	139,11	may-93	147,71	may-94	167,38
jun-90	99,71	jun-91	107,43	jun-92	131,10	jun-93	144,70	jun-94	160,23
jul-90	96,47	jul-91	105,34	jul-92	130,79	jul-93	147,57	jul-94	162,88
ago-90	101,71	ago-91	109,08	ago-92	136,32	ago-93	136,29	ago-94	159,37
sep-90	101,58	sep-91	112,78	sep-92	110,70	sep-93	138,76	sep-94	158,07
oct-90	111,10	oct-91	118,36	oct-92	130,83	oct-93	151,91	oct-94	169,87
nov-90	105,84	nov-91	121,21	nov-92	136,61	nov-93	161,37	nov-94	182,51
dic-90	132,48	dic-91	153,17	dic-92	165,09	dic-93	193,51	dic-94	226,14

fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-95	162,02	ene-96	162,69	ene-97	172,54	ene-98	171,58	ene-99	176,94
feb-95	161,95	feb-96	162,62	feb-97	174,89	feb-98	168,55	feb-99	170,07
mar-95	189,34	mar-96	182,84	mar-97	193,19	mar-98	185,41	mar-99	187,50
abr-95	171,06	abr-96	169,50	abr-97	187,53	abr-98	174,24	abr-99	180,20
may-95	196,09	may-96	186,31	may-97	188,91	may-98	192,19	may-99	198,89
jun-95	194,96	jun-96	176,03	jun-97	175,79	jun-98	180,30	jun-99	187,84
jul-95	167,27	jul-96	164,06	jul-97	166,74	jul-98	170,37	jul-99	176,87
ago-95	166,36	ago-96	161,73	ago-97	166,90	ago-98	169,80	ago-99	174,08
sep-95	169,67	sep-96	160,73	sep-97	167,16	sep-98	171,74	sep-99	172,06
oct-95	179,09	oct-96	186,54	oct-97	178,17	oct-98	181,35	oct-99	179,50
nov-95	192,34	nov-96	204,66	nov-97	192,15	nov-98	194,00	nov-99	196,00
dic-95	224,76	dic-96	235,87	dic-97	230,16	dic-98	236,58	dic-99	236,39

fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-00	181,06	ene-01	172,02	ene-02	180,42	ene-03	184,08	ene-04	187,90
feb-00	172,94	feb-01	174,69	feb-02	175,95	feb-03	178,94	feb-04	169,42
mar-00	186,54	mar-01	187,86	mar-02	188,93	mar-03	191,50	mar-04	192,16
abr-00	176,92	abr-01	176,13	abr-02	180,51	abr-03	185,72	abr-04	186,86
may-00	198,52	may-01	198,32	may-02	199,50	may-03	206,33	may-04	208,11
jun-00	182,67	jun-01	186,23	jun-02	188,85	jun-03	195,67	jun-04	197,40
jul-00	174,99	jul-01	178,45	jul-02	185,26	jul-03	191,78	jul-04	193,16
ago-00	172,71	ago-01	175,24	ago-02	183,17	ago-03	186,55	ago-04	186,99
sep-00	170,36	sep-01	173,55	sep-02	179,26	sep-03	183,86	sep-04	184,41
oct-00	180,15	oct-01	184,03	oct-02	187,39	oct-03	192,74	oct-04	194,19
nov-00	194,06	nov-01	195,22	nov-02	198,96	nov-03	202,99	nov-04	207,60
dic-00	228,83	dic-01	235,32	dic-02	235,47	dic-03	241,29	dic-04	244,00

fecha	dato	fecha	dato	fecha	Dato
ene-05	188,78	ene-06	198,64	ene-07	195,85
feb-05	172,99	feb-06	191,86	feb-07	194,86
mar-05	202,61	mar-06	218,38		
abr-05	195,66	abr-06	195,89		
may-05	212,49	may-06	237,63		
jun-05	200,21	jun-06	202,36		
jul-05	197,14	jul-06	208,41		
ago-05	190,55	ago-06	193,29		
sep-05	183,94	sep-06	189,74		
oct-05	193,11	oct-06	195,03		
nov-05	210,62	nov-06	202,55		
dic-05	247,73	dic-06	260,37		

### Estimación de los parámetros para la serie Comidas, Restaurante y Hoteles

#### Parametros Estimados

		Estimados	Error Std	t	Sig Aprox
Sin	MA1	,266	,090	2,970	,004
estacionalidad	MA2	,370	,090	4,104	,000
Constante		,052	,154	,335	,739

#### Parametros Estimados

		Estimado	Error Std	t	Sig Aprox
Sin	AR1	-,124	,093	-1,336	,184
estacionalidsaa	AR2	-,337	,093	-3,614	,000
Constante		,091	,289	,314	,754

#### Parametros Estimados

		Estimado	Error Std	t	Sig Aprox
Sin	MA1	,369	,089	4,145	,000
estacionalidad	MA2	,431	,094	4,569	,000
Estacional	Estacional MA1	,493	,097	5,092	,000
Constante		,035	,046	,759	,449

**Parametros Estimados**

		Estimado	Error Std	t	Sig Approx
sin	AR1	-,135	,093	-1,457	,148
estacionalidad	AR2	-,364	,094	-3,891	,000
Estacional	Estacional AR1	-,398	,092	-4,326	,000
Constante		,082	,195	,418	,677

**Parametros Estimado**

		Estimado	Error Std	t	Sig Approx
Sin estacional	AR1	,559	,139	4,032	,000
	AR2	-,800	,113	-7,069	,000
	MA1	,725	,176	4,120	,000
	MA2	-,620	,163	-3,796	,000
Constante		,101	,298	,338	,736

**Parametros Estimado**

		Estimado	Error Std	t	Sig Approx
Sin estacionalidadl	AR1	,374	,796	,470	,639
	AR2	,043	,497	,086	,932
	MA1	,712	,783	,909	,365
	MA2	,272	,792	,344	,732
Estacional	estacional	,231	,230	1,002	,319
	estacinoal	,673	,205	3,280	,001
Constante		,029	,017	1,689	,094

**Base de datos del IPC en Salud.**

fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-93	100,91	ene-94	109,67	ene-95	121,14	ene-96	137,69	ene-97	140,61
feb-93	101,23	feb-94	110,15	feb-95	122,4	feb-96	137,76	feb-97	142,22
mar-93	102,1	mar-94	112,77	mar-95	123,63	mar-96	137,82	mar-97	143,74
abr-93	102,43	abr-94	113,06	abr-95	125,63	abr-96	137,59	abr-97	144,35
may-93	103,12	may-94	113,44	may-95	126,64	may-96	137,6	may-97	144,49
jun-93	103,49	jun-94	115,33	jun-95	126,88	jun-96	138,34	jun-97	144,91
jul-93	104,04	jul-94	116,24	jul-95	131,18	jul-96	138,43	jul-97	145,97
ago-93	105,73	ago-94	117,19	ago-95	133,69	ago-96	138,53	ago-97	144,23
sep-93	106,83	sep-94	116,29	sep-95	135,23	sep-96	139,04	sep-97	144,27
oct-93	107,43	oct-94	116,66	oct-95	136,25	oct-96	139,28	oct-97	144,21
nov-93	107,98	nov-94	117,18	nov-95	136,6	nov-96	140,05	nov-97	144,65
dic-93	108,76	dic-94	119,65	dic-95	136,91	dic-96	140,2	dic-97	144,73
fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-98	145,17	ene-99	150,69	ene-00	158,05	ene-01	175,36	ene-02	182,47
feb-98	145,44	feb-99	150,76	feb-00	157,87	feb-01	175,44	feb-02	182,75
mar-98	145,99	mar-99	151,74	mar-00	158,92	mar-01	176,05	mar-02	183,19
abr-98	146,01	abr-99	152,54	abr-00	159,59	abr-01	176,6	abr-02	183,32
may-98	146,12	may-99	152,71	may-00	160,06	may-01	176,71	may-02	185,14
jun-98	146,63	jun-99	152,59	jun-00	160,58	jun-01	175,9	jun-02	186,07
jul-98	148,46	jul-99	152,82	jul-00	167,52	jul-01	176,01	jul-02	186,24
ago-98	149,56	ago-99	153,26	ago-00	171,32	ago-01	177,9	ago-02	186,64
sep-98	149,81	sep-99	154,17	sep-00	171,11	sep-01	178,92	sep-02	187,9
oct-98	149,99	oct-99	154,68	oct-00	173,3	oct-01	179,17	oct-02	187,98
nov-98	150,23	nov-99	155,57	nov-00	173,08	nov-01	180,58	nov-02	188,22
dic-98	149,58	dic-99	157,18	dic-00	173,78	dic-01	180,93	dic-02	188,79
fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-03	191,5	ene-04	196,37	ene-05	202,7	ene-06	216,87	ene-07	222,11
feb-03	190,91	feb-04	195,96	feb-05	203,93	feb-06	217,35	feb-07	222,73
mar-03	193,31	mar-04	196,67	mar-05	205,25	mar-06	217,07	mar-07	223,85
abr-03	192,7	abr-04	197,48	abr-05	206,33	abr-06	217,28	abr-07	224,03
may-03	192,98	may-04	197,55	may-05	208,17	may-06	218,16	may-07	224,37
jun-03	194,66	jun-04	198,04	jun-05	209,45	jun-06	219,5	jun-07	224,73
jul-03	194,73	jul-04	199,96	jul-05	210,28	jul-06	220,06	jul-07	224,72
ago-03	194,61	ago-04	199,84	ago-05	210,24	ago-06	221,79	ago-07	224,62
sep-03	193,49	sep-04	202,23	sep-05	211,12	sep-06	221,69	sep-07	224,76
oct-03	195	oct-04	202,58	oct-05	210,49	oct-06	222,46	oct-07	225,41
nov-03	194,71	nov-04	202,56	nov-05	211,08	nov-06	222,13		
dic-03	195,38	dic-04	203,26	dic-05	211,87	dic-06	222,51		



## Estimación de los Parámetros para los modelos propuestos de la serie Salud

**Parametro estimado**

	Estimado	Error Std	t	Sig Approx
Sin estacionalidad MA1	-.167	.088	-1.887	.061
Constante	-.012	.154	-.075	.940

**Parametro Estimado**

	Estimado	Error Std	t	Sig Approx
estacional estacional MA1	.999	30.954	.032	.974
Constante	-.017	.025	-.682	.497

**Parametro Estimado**

	Estimado	Error Std	t	Sig Approx
Sin estacionalidad AR1	.182	.087	2.077	.040
Constante	-.011	.161	-.067	.946

Melard's algorithm was used for estimation.

**Parametro Estimado**

	Estimates	Error Std	t	Sig Approx
Estacional estacional AR1	-.550	.075	-7.327	.000
Constante	-.019	.074	-.263	.793

**Parameter Estimates**

	Estimado	Error Std	t	Sig Approx
Sin estacionalidad MA1	-.135	.082	-1.636	.104
estacional estacionalMA1	.945	.347	2.719	.007
Constante	-.017	.028	-.603	.547

Melard's algorithm was used for estimation.

**Parametro Estimado**

	Estimado	Error Std	t	Sig Approx
Sin estacionalidad AR1	.176	.088	1.994	.048
Estacional estacional AR1	-.554	.078	-7.110	.000
Constante	-.019	.088	-.217	.829

**Parametro Estimado**

	Estimado	Error Std	t	Sig Aprox
Sin estacionalidad AR1	.567	.390	1.455	.148
MA1	.438	.430	1.019	.310
Seasonal Seasonal AR1	-.085	.110	-.773	.441
Lags Seasonal MA1	.994	4.075	.244	.808
Constant	-.016	.029	-.561	.576

**Estimación de los parámetros de los modelos con intervención de la serie Salud.**

Input: SALUD Transformations: D(1),D(1) (Interrupted ARIMA) Model:(0,1,1)(0,1,0) MS Residual= .78069								
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 138)	p	Lower 95%	Upper 95%	Interv. Case No.	Interv. Type
q(1)	0.925294	0.045969	20.12852	0.000000	0.834399	1.016190		
Omega(1)	6.228351	0.874375	7.12320	0.000000	4.499446	7.957256	91	Abr/Perm

Input: SALUD Transformations: D(1),D(1) (Interrupted ARIMA) Model:(1,1,0)(0,1,0) MS Residual= 1.0132								
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 138)	p	Lower 95% Conf	Upper 95% Conf	Interv. Case No.	Interv. Type
p(1)	-0.594152	0.071118	-8.35444	0.000000	-0.734774	-0.453530		
Omega(1)	6.369385	0.835898	7.61981	0.000000	4.716561	8.022209	91	Abr/Perm

Input: SALUD Transformations: D(1),D(1) (Interrupted ARIMA) Model:(2,1,0)(0,1,0) MS Residual= .91814								
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 137)	p	Lower 95% Conf	Upper 95% Conf	Interv. Case No.	Interv. Type
p(1)	-0.775349	0.083305	-9.30737	0.000000	-0.940079	-0.610620		
p(2)	-0.322196	0.082727	-3.89467	0.000153	-0.485784	-0.158608		
Omega(1)	6.263010	0.859213	7.28924	0.000000	4.563975	7.962045	91	Abr/Perm

Input: SALUD Transformations: D(1),D(1) (Interrupted ARIMA) Model:(1,1,1)(0,1,0) MS Residual= .78099								
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 137)	p	Lower 95% Conf	Upper 95%	Interv. Case No.	Interv. Type
p(1)	-0.103414	0.106022	-0.97540	0.331080	-0.313064	0.106237		
q(1)	0.894359	0.061799	14.47203	0.000000	0.772155	1.016562		
Omega(1)	6.474065	0.902382	7.17442	0.000000	4.689667	8.258464	91	Abr/Perm

Input: SALUD								
Transformations: D(1),D(1) (Interrupted ARIMA)								
Model:(2,1,1)(0,1,0) MS Residual= .77730								
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 136)	p	Lower 95% Conf	Upper 95%	Interv. Case No.	Interv. Type
p(1)	-0.035999	0.123839	-0.29069	0.771729	-0.280898	0.208900		
p(2)	0.145780	0.120069	1.21414	0.226800	-0.091664	0.383224		
q(1)	0.953600	0.075973	12.55191	0.000000	0.803360	1.103841		
Omega(1)	6.522428	0.896717	7.27367	0.000000	4.749115	8.295741	91	Abr/Perm

**Base de datos del IPC en prendas de vestir y calzado.**

fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-93	102,17	ene-94	106,67	ene-95	116,23	ene-96	121,78	ene-97	123,36
feb-93	102,46	feb-94	110,86	feb-95	116,16	feb-96	122,4	feb-97	122,82
mar-93	102,48	mar-94	111,37	mar-95	116,46	mar-96	122,38	mar-97	122,97
abr-93	102,48	abr-94	112,02	abr-95	117,31	abr-96	122,67	abr-97	122,99
may-93	103,12	may-94	112,22	may-95	117,47	may-96	122,62	may-97	123,18
jun-93	103,71	jun-94	112,55	jun-95	118,3	jun-96	122,97	jun-97	122,78
jul-93	103,78	jul-94	113,65	jul-95	120,1	jul-96	122,93	jul-97	122,88
ago-93	104,1	ago-94	113,86	ago-95	120,75	ago-96	122,89	ago-97	122,88
sep-93	104,2	sep-94	114,04	sep-95	121,43	sep-96	122,57	sep-97	122,86
oct-93	104,78	oct-94	114,82	oct-95	121,41	oct-96	123,02	oct-97	122,7
nov-93	105,16	nov-94	114,78	nov-95	121,48	nov-96	122,96	nov-97	122,82
dic-93	105,77	dic-94	115,53	dic-95	121,44	dic-96	123,03	dic-97	122,75
fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-98	123	ene-99	121,75	ene-00	120,66	ene-01	118,19	ene-02	113,11
feb-98	122,9	feb-99	121,8	feb-00	120,48	feb-01	118,05	feb-02	113,1
mar-98	122,51	mar-99	121,66	mar-00	120,25	mar-01	117,42	mar-02	112,53
abr-98	122,37	abr-99	121,12	abr-00	120,36	abr-01	116,86	abr-02	112,59
may-98	122,26	may-99	121,18	may-00	120,27	may-01	116,35	may-02	112,14
jun-98	122,26	jun-99	121,1	jun-00	120,34	jun-01	115,97	jun-02	112,05
jul-98	122,07	jul-99	120,77	jul-00	119,45	jul-01	115,85	jul-02	111,9
ago-98	121,92	ago-99	120,89	ago-00	119,28	ago-01	115,35	ago-02	111,74
sep-98	122,01	sep-99	120,9	sep-00	119,39	sep-01	115,46	sep-02	111,59
oct-98	122,01	oct-99	120,76	oct-00	119,21	oct-01	114,62	oct-02	111,39
nov-98	122,22	nov-99	120,59	nov-00	119,15	nov-01	114,38	nov-02	111,08
dic-98	122,15	dic-99	120,27	dic-00	119,13	dic-01	114,15	dic-02	111,2
fecha	dato	fecha	dato	fecha	dato	fecha	dato	fecha	dato
ene-03	111,05	ene-04	110,08	ene-05	109,42	ene-06	109,79	ene-07	111,34
feb-03	111,15	feb-04	109,89	feb-05	109,54	feb-06	109,86	feb-07	111,28
mar-03	111,09	mar-04	109,92	mar-05	109,49	mar-06	109,85	mar-06	111,43
abr-03	110,88	abr-04	109,81	abr-05	109,46	abr-06	109,91	abr-07	111,43
may-03	110,9	may-04	109,93	may-05	109,31	may-06	110,12	may-07	111,47
jun-03	110,69	jun-04	109,84	jun-05	109,5	jun-06	110,11	jun-07	111,65
jul-03	110,53	jul-04	109,42	jul-05	109,59	jul-06	110,21	jul-07	111,81
ago-03	110,17	ago-04	109,51	ago-05	109,55	ago-06	110,63	ago-07	111,84
sep-03	110,21	sep-04	109,61	sep-05	109,71	sep-06	110,86	sep-07	111,96
oct-03	110,18	oct-04	109,44	oct-05	109,91	oct-06	111,03	oct-07	112,04
nov-03	110,28	nov-04	109,38	nov-05	110,11	nov-06	111,04		
dic-03	110,23	dic-04	109,21	dic-05	109,79	dic-06	111,13		

