

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMATICA
ESCUELA DE MATEMATICA
MAESTRIA EN ESTADISTICA.



Universidad de El Salvador
Hacia la libertad por la cultura

MEMORIA FINAL DE TESIS:
**ANÁLISIS DE VARIABLES MACROECONÓMICAS DE EL SALVADOR: PERIODO
1995 AL SEGUNDO TRIMESTRE DE 2009.**

PRESENTADO POR:
LICDA. INGRID CAROLINA MARTINEZ BARAHONA
CARNET: MB02004

PARA OPTAR AL TITULO DE:
MAESTRA EN ESTADISTICA

ASESOR EXTERNO
DR. JAVIER CASTRO CANTALEJO
UNIVERSIDAD COMPLUTENSE DE MADRID

ASESOR INTERNO
DR. JOSE NERYS FUNES

CIUDAD UNIVERSITARIA, DICIEMBRE DE 2009

UNIVERSIDAD DE EL SALVADOR

MSC. RUFINO ANTONIO QUEZADA SÁNCHEZ
RECTOR

LIC. DOUGLAS VLADIMIR ALFARO CHÁVEZ
SECRETARIO GENERAL

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

DR. RAFAEL ANTONIO GÓMEZ ESCOTO
DECANO

MSC. MARIA TRINIDAD TRIGUEROS DE CASTRO
SECRETARIA

ESCUELA DE MATEMÁTICA

ING. CARLOS MAURICIO CANJURA LINARES
DIRECTOR

AGRADECIMIENTOS

A MIS PADRES Y HERMANOS POR TODO EL APOYO QUE ME BRINDARON PARA PODER ALCANZAR ESTA NUEVA META EN LA VIDA.

A MIS ASESORES POR TODO EL TIEMPO DEDICADO A LA REVISIÓN Y CORRECCIÓN DEL TRABAJO.

A TODOS GRACIAS.

Índice

1. INTRODUCCIÓN.....	6
2. PLANTAMIENTO DEL PROBLEMA.....	11
3. MARCO TEÓRICO.....	16
3.1. Análisis Univariante. Modelo ARIMA.....	16
3.1.1. Descripción de series temporales.....	16
3.1.2. Funciones de autocorrelación.....	17
3.1.3. Procesos Autorregresivos.....	20
3.1.4. Procesos de Media Móvil.....	28
3.1.5. Procesos ARMA(p,q).....	33
3.1.6. Series No Estacionarias. Procesos ARIMA(p,d,q).....	34
3.1.7. Modelos Estacionales. Procesos ARIMA(p,d,q)(ps, ds,qs)s.....	34
3.2. Análisis Multivariante. Regresión Dinámica.....	36
3.2.1. Funciones de correlaciones cruzadas.....	36
3.2.2. Modelo de Función de Transferencia.....	37
3.3. Comparación de series temporales. Índice cualitativo de similitud.....	39
3.3.1. Normalización.....	40
3.3.2. Subsecuencia Común Máxima (LCS).....	41
3.3.3. Índice Cualitativo de Similitud (QSI).....	42
4. ANÁLISIS DESCRIPTIVOS DE LOS DATOS.....	46
4.1. Análisis descriptivo para las variables del sector Servicios.....	46
4.2. Análisis descriptivo para las variables del sector Comercio.....	48
4.3. Análisis descriptivo para las variables del sector Construcción.....	49
4.4. Análisis descriptivo para las variables del sector Industria.....	51
4.5. Análisis descriptivo para las variables Globales.....	53
5. MODELADO DE LOS DATOS GLOBALES.....	55
5.1. Modelado mediante modelos ARIMA.....	55
5.1.1. Modelado de la serie <i>Empleo</i>	55
5.1.2. Modelado de la serie <i>Ventas</i>	60
5.1.3. Modelo para la serie <i>Inversiones</i>	62
5.1.4. Modelo para la serie <i>Importaciones</i>	64
5.2. Modelado mediante regresión dinámica.....	65
5.2.1. Correlación cruzada.....	65
5.2.2. Modelo para la serie <i>Empleo</i>	69
5.2.3. Modelo para la serie <i>Ventas</i>	89
5.2.4. Modelo para la serie <i>Inversiones</i>	91
5.2.5. Modelo para la serie <i>Importaciones</i>	93
5.3. Comparación de los modelos ARIMA y los modelos de regresión dinámica... ..	95
5.3.1. Serie <i>Empleo</i>	95
5.3.2. Serie <i>Ventas</i>	97
5.3.3. Serie <i>Inversiones</i>	98
5.3.4. Serie <i>Importaciones</i>	100
6. COMPARACIÓN DE DATOS.....	102
6.1. Comparación por Sectores.....	103
6.1.1. Sector Servicios.....	104
6.1.2. Sector Comercio.....	108
6.1.3. Sector Construcción.....	111
6.1.4. Sector Industria.....	115
6.1.5. Datos Globales.....	118

6.2. Comparación por Variables.....	122
6.2.1. Variable <i>Ventas</i>	122
6.2.2. Variable <i>Empleo</i>	128
6.2.3. Variable <i>Inversiones</i>	133
6.2.4. Variable <i>Importaciones</i>	139
7. CONCLUSIONES.....	145
8. BIBLIOGRAFÍA.....	152

1. INTRODUCCIÓN

El estudio que presentamos, se basa en el análisis de series temporales. Las series temporales tienen por objeto analizar la evolución de una variable a través del tiempo.

La diferencia esencial entre las series temporales y los análisis no temporales (Estadística descriptiva, Diseño de experimentos o Regresión) es que en los análisis previos no importaba el orden en que estaban tomadas las observaciones y éste se podía variar sin problemas. En series temporales el orden es muy importante y variarlo supone cambiar la información contenida en la serie.

Las series temporales son aplicadas en varias ciencias, algunos ejemplos son:

- En Ciencias Sociales se tiene uno de los primeros estudios de las series temporales que data del año 1662 cuando J. Graunt hace una interpretación de los movimientos demográficos y de la conducta social, a partir de las cantidades brutas de nacimientos y muertes, por causas, en Londres. La fuente de datos utilizada fue las tablas semanales de compañía de Sacristanes Parroquiales. Estas tablas contenían la cantidad de muertes por causas, y la cantidad de nacimientos ocurrida en cada parroquia londinense, durante el periodo de 1604 a 1661.
- En Meteorología podemos ver varios ejemplos de series temporales, como la cantidad de lluvia, que es una variable medida mensualmente, o la temperatura, que también es medida mensualmente.
- En Cardiología específicamente en Hemodinámica (que se encarga del estudio anatómico y funcional del corazón mediante la introducción de catéteres finos a través de las arterias de la ingle o del brazo) la monitorización de cada paciente incluye la observación continua de una gran cantidad de variables fisiológicas: diferentes presiones sanguíneas, pulso, temperatura, etc. a lo largo del tiempo. Estas series temporales proporcionan información crucial sobre el estado del paciente crítico.
- En Astronomía tenemos una de las serie más antiguas: la serie anual de las manchas solares. Las manchas solares son regiones oscuras de dimensiones muy variadas compuestas por un núcleo central llamado sombra, rodeado de amplias fajas llamadas penumbras, que aparecen a veces formando agrupaciones que pueden llegar a alcanzar varios centenares de miles de

kilómetros de longitud, siendo visibles en tal caso a simple vista desde la tierra. Se tiene registro de este tipo de actividad solar desde el siglo XVIII, y lo que se mide es el número medio de manchas anuales.

- En biología tenemos la serie de diversidad de especie (la diversidad es la indicadora de un conjunto de aspectos del funcionamiento de un ecosistema).
- En economía tenemos multitud de series, hoy en día es sin duda una de las aplicaciones más importantes de las series temporales. La mayoría de las variables utilizadas en economía como por ejemplo: el Índice del Volumen de la Actividad Económica (IVAE), el Saldo del Sector Público No Financiero (SPNF), el Producto Interno Bruto (PIB), las exportaciones, las importaciones, la balanza comercial, la factura petrolera, las remesas familiares etc. Son variables medidas en forma mensual, trimestral o anual. Claramente todas las variables anteriores pueden ser estudiadas con la ayuda de las series temporales.

Para el análisis de series temporales es imprescindible tener acceso a un paquete estadístico capaz de realizar los cálculos necesarios entre los cuales podemos mencionar: el Statgraphics, Minitab, SPSS, S-plus y R. El Statgraphics proporciona las herramientas básicas para la identificación y estimación de un modelo univariante para una serie temporal, pero no permite realizar la estimación máximo verosímil exacta, ni el tratamiento de datos atípicos. Tampoco permite la regresión dinámica ni el análisis multivariante. El Minitab y el SPSS son muy simples de usar, estos programas permiten construir la regresión dinámica, pero no incluyen tratamiento de datos atípicos. El S-plus es un programa más completo, que incluye muchas posibilidades de análisis sofisticados y que puede programarse. El programa R es muy similar al S-plus y puede descargarse gratuitamente en la web. Por su facilidad de manejo y su suficiencia para el tipo de estudios que se desean realizar, utilizaremos el SPSS para obtener los resultados del estudio.

En el estudio, se aplicarán algunas de las técnicas de series temporales para ver el comportamiento de las siguientes variables de la economía en El Salvador: Empleo, Inversiones, Importaciones y Ventas del periodo 1995 al segundo trimestre de 2009. Dicha información fue brindada por la Fundación Salvadoreña para el Desarrollo Social (FUSADES), organización privada, sin fines de lucro, dedicada a la investigación y a la

realización de estudios para promover políticas públicas orientadas al desarrollo sostenible de El Salvador.

Ya que se hará un estudio de las variables macroeconómicas de El Salvador en el periodo 1995-2009 es necesario tener un panorama de lo sucedido en la economía en ese periodo.

Tras los acuerdos de paz de 1992 la economía de El Salvador ha estado creciendo a un ritmo moderado pero constante gracias a las iniciativas de mercado libre y la privatización de muchos servicios antes controlados por el Estado (las telecomunicaciones y el fondo de pensiones entre otros). Este crecimiento, sin embargo, mantiene muchas de las desigualdades y aún no ha alcanzado los niveles de la década de 1970. El quinto más rico de la población controla el 45% de la ganancia del país, mientras que el quinto más pobre recibió solamente 5,6%.

La economía de El Salvador se ha visto afectada por los desastres naturales. En Noviembre de 1998 el paso del huracán Mitch dejó en El Salvador 240 muertos, miles de damnificados y \$262 millones en pérdidas económicas. El sector más golpeado fue la agricultura, ya que las zonas productoras de granos básicos quedaron totalmente devastadas.

Bajo la administración del ex presidente Francisco Flores (1999-2004), con el respaldo del Congreso dominado por la derecha, El Salvador se dolarizó el 1 de enero de 2001, al entrar en vigencia la llamada Ley de Integración Monetaria (LIM), que fijó el tipo de cambio en 8,75 colones por un dólar. La ley además permitió las transacciones en otras monedas fuertes como el yen, la libra esterlina y el euro. A pesar de que la ley permitía el bimonetarismo, los bancos del sistema comenzaron a retener los colones, por lo que en el presente sólo circula el dólar. Con la dolarización, El Salvador se quedó sin política monetaria y el Banco Central de Reserva pasó a depender de la Reserva Federal de Estados Unidos, a la que ahora le paga para que le imprima la moneda. Con la entrada del dólar la tasa activa (intereses a los préstamos) bajaron en los bancos hasta un 8%, pero a pesar de eso la tan ansiada reactivación no se presentó en virtud de que no hay demanda en el mercado por la escasa capacidad de compra de la población. Con el dólar, el ahorro no se vio estimulado, por cuanto devenga intereses de entre el 0,25% y

el 1,75% dependiendo del tipo de cuenta y la cantidad de efectivo que posea en el sistema financiero.

También en el 2001 El Salvador sufre un duro golpe a la economía causado por el peor desastre natural en la historia. Dos terremotos, el primero el 13 de enero y el segundo exactamente un mes después, el 13 de febrero. Dejando grandes daños: 1,500 fallecidos, 193 soterrados, 125 desaparecidos, 5,565 heridos, 1,364,160 damnificados, 68,777 evacuaciones, 39,000 desempleos y 24,000 pescadores artesanales afectados. Las cifras de los daños materiales son grandes, se calcula que: 277953 viviendas fueron destruidas incluyendo 688 soterradas, 32000 micros y pequeñas empresas destruidas, 1385 escuelas dañadas y 190 destruidas por completo, 94 hospitales dañados y aproximadamente la cuarta parte de carreteras pavimentadas fueron destruidas. El impacto económico causado por los terremotos fue grande, además obligó al estado a destinar miles de millones de dólares para la reconstrucción de casi todo un país.

Luego de los terremotos la economía se fue estabilizando poco a poco. En junio de 2002 entra en vigencia el tratado de libre comercio entre El Salvador y Chile, en dicho tratado se plantean:

- Programas de desgravación arancelaria: Donde cada país estableció las preferencias arancelarias otorgadas a la otra parte y el respectivo esquema de desgravación para cada producto negociado.
- Reglas de origen y comercio transfronterizo de servicios

Pero este no es el único tratado de libre comercio. El 1 de marzo de 2006, El Salvador se convirtió en el primer país en poner en vigencia el tratado de Centroamérica y República Dominicana con Estados Unidos, conocido como CAFTA-DR, por sus siglas en inglés. Desde su vigencia más de 80 empresas se han establecido en el país con una inversión superior a 409 millones de dólares. Además, se generaron 12.000 empleos directos y 26.000 indirectos, sobre todo en la agroindustria, el sector textil y en servicios médicos. El Salvador tiene además acuerdos de libre comercio con México y Panamá. También tiene tratados con Canadá, Colombia, Taiwán, actualmente se negocia con la Unión Europea.

Pero llegamos a un punto importante que afecta a la economía actual no solo de El Salvador sino de todo el mundo. La actual crisis económica originada en los Estados Unidos en el año 2008 debido, entre otros factores, a los elevados precios del petróleo y materias primas, ha desatado una de las peores crisis crediticia e hipotecaria. En enero de 2008, el precio del petróleo superó los US\$100/barril por primera vez en su historia, y alcanzó los US\$147/barril en julio debido a fenómenos especulativos de alta volatilidad. En septiembre de ese mismo año, los problemas se agravaron con la bancarrota de diversas entidades financieras estadounidenses relacionadas con el mercado de las hipotecas inmobiliarias, como el banco de inversión Lehman Brothers, las compañías hipotecarias Fannie Mae y Freddie Mac y la aseguradora AIG.

El Salvador comenzó a sufrir los efectos de la crisis mundial en el tercer trimestre del 2008 cuando los índices de ventas, importaciones, inversiones y empleo comenzaron a crecer negativamente en forma estrepitosa. Los cuatro sectores: Comercio, Servicio, Industria y Construcción se han visto gravemente afectados por la crisis, pero el más vulnerable ha sido el sector de la construcción que en el primer trimestre del 2009 ha presentado un decrecimiento importante.

La crisis ha generado un impacto social en El Salvador debido principalmente a dos factores. La reducción de las remesas familiares (que es un aspecto muy importante ya que se estima que uno de cada cuatro hogares salvadoreños recibe remesas de familiares en Estados Unidos) y la pérdida de empleo. Esta pérdida de empleo se ve reflejado en el número de cotizaciones en el Instituto Salvadoreño de Seguro Social (ISSS). De septiembre de 2008 a febrero de 2009 el número de cotizaciones cayó en 24,797. El desempleo afecta más a los sectores de bajos ingresos, por lo que el desempleo toca la puerta de los hogares más pobres de El Salvador.

Por todos los sucesos que hemos analizado en este periodo se hace necesario hacer un estudio objetivo de las variables macroeconómicas de El Salvador que permita ver la crisis de una forma objetiva, eliminando en cierto grado los posibles sesgos que se pueden producir al tratar con opiniones.

2. PLANTAMIENTO DEL PROBLEMA.

En esta sección analizaremos los distintos pasos que se han seguido para la modelización matemática.

Como se dijo en la introducción, los datos originales provienen de la organización FUSADES. La organización recolecta la información a través del sistema de encuestas de coyuntura económica, de la que se desprende la encuesta dinámica empresarial. Dicha encuesta se desarrolla en forma trimestral desde 1987. Y va orientada hacia los cuatro sectores económicos del país: industria, construcción, comercio y servicios. Además se realiza en función del tamaño de las empresas: grande, mediana, pequeña y microempresa.

En esta encuesta los datos recogidos son:

1. Sector Comercio.
 - 1.1. Ventas
 - 1.2. Inventarios de productos terminados
 - 1.3. Importaciones totales
 - 1.4. Mora de clientes
 - 1.5. Empleados Fijos
 - 1.6. Inversiones
 - 1.7. Demanda por los productos/servicios
 - 1.8. Pedidos de productos para vender
2. Sector Construcción:
 - 2.1. Actividad global de construcción
 - 2.2. Importaciones totales
 - 2.3. Mora de los clientes
 - 2.4. Empleados fijos
 - 2.5. Inversiones
 - 2.6. Trabajos de construcción
 - 2.7. Duración de la actividad asegurada.

3. Industria.
 - 3.1. Producción
 - 3.2. Inventario de materias primas
 - 3.3. Entrega de proveedores locales
 - 3.4. Atrasos en entregas a clientes
 - 3.5. Ventas
 - 3.6. Exportaciones totales
 - 3.7. Inventario de productos terminados
 - 3.8. Importaciones totales
 - 3.9. Mora de los clientes
 - 3.10. Empleados fijos
 - 3.11. Inversiones
4. Servicio
 - 4.1. Ventas
 - 4.2. Importaciones totales
 - 4.3. Mora de los clientes
 - 4.4. Empleados fijos
 - 4.5. Inversiones
5. Global
 - 5.1. Ventas
 - 5.2. Exportaciones totales
 - 5.3. Inventario de productos terminados
 - 5.4. Importaciones totales
 - 5.5. Mora de los clientes
 - 5.6. Empleados fijos
 - 5.7. Inversiones

De las variables anteriores hemos seleccionado para el estudio cuatro, el Empleo, las Importaciones, las Inversiones y las Ventas. Hemos seleccionado esas variables por ser las más representativas que estaban presentes en los cuatro sectores, y además presentaban mejor calidad de datos. Con estas cuatro variables es posible tener un panorama general de la economía del país. Estas cuatro variables las estudiaremos de una forma global (datos agrupados para los distintos sectores) y posteriormente compararemos su comportamiento en los distintos sectores.

La base de datos la podemos dividir en dos:

1. Del primer trimestre de 1987 hasta el cuarto trimestre de 1993. En esta parte de la serie hay valores de las distintas variables en distintos momentos de tiempo que no fueron tomados, datos perdidos o missing. Además hay que recordar que esos datos fueron tomados en el periodo de guerra y primeros años de post-guerra que vivió El Salvador.
2. Desde el primer trimestre de 1994 hasta el segundo trimestre del 2009. En esa base solo hay datos perdidos en el año 1994. Aunque los datos fueron recogidos con una encuesta con modificaciones.

En un principio se quería realizar una serie temporal desde 1987 hasta la actualidad, pero eso nos obligaría a:

- Estimar diversos valores perdidos, lo que siempre desvirtúa los posibles resultados.
- Suponer que el cambio de encuesta no altera los datos y por lo tanto son comparables. Lo que no podemos garantizar dado que no disponemos del cuestionario de los primeros años.
- Suponer que el modelo de Ventas, Importaciones, Inversiones y Empleo en los años de guerra y primeros años de post-guerra es similar al actual

Por todo ello y con el fin de lograr los resultados más limpios posibles, se ha decidido trabajar con las series desde el año de 1995 hasta el segundo trimestre del 2009. Cada serie consta de 58 datos que es un valor bajo para predecir, pero suficiente para analizar las posibles relaciones entre las variables.

En conclusión, las variables utilizadas en el estudio son:

- *Empleo: Porcentaje de variación con respecto al número de empleados fijos en El Salvador en el trimestre anterior.*
- *Inversiones: Porcentaje de variación con respecto a la inversión global en El Salvador en el trimestre anterior.*
- *Importaciones: Porcentaje de variación con respecto a las importaciones globales en El Salvador en el trimestre anterior.*

- *Ventas: Porcentaje de variación con respecto a las ventas globales en El Salvador en el trimestre anterior.*
- *EmpleoServicios: Porcentaje de variación con respecto al número de empleados fijos en el sector Servicios en el trimestre anterior.*
- *InversionesServicios: Porcentaje de variación con respecto a la inversión en el sector Servicios en el trimestre anterior.*
- *ImportacionesServicios: Porcentaje de variación con respecto a las importaciones en el sector Servicios en el trimestre anterior.*
- *VentasServicios: Porcentaje de variación con respecto a las ventas en el sector Servicios en el trimestre anterior*
- *EmpleoIndustria: Porcentaje de variación con respecto al número de empleados fijos en el sector Industria en el trimestre anterior.*
- *InversionesIndustria: Porcentaje de variación con respecto a la inversión en el sector Industria en el trimestre anterior.*
- *ImportacionesIndustria: Porcentaje de variación con respecto a las importaciones en el sector Industria en el trimestre anterior.*
- *VentasIndustria: Porcentaje de variación con respecto a las ventas en el sector Industria en el trimestre anterior.*
- *EmpleoComercio: Porcentaje de variación con respecto al número de empleados fijos en el sector Comercio en el trimestre anterior.*
- *InversionesComercio: Porcentaje de variación con respecto a la inversión en el sector Comercio en el trimestre anterior.*
- *ImportacionesComercio: Porcentaje de variación con respecto a las importaciones en el sector Comercio en el trimestre anterior.*
- *VentasComercio: Porcentaje de variación con respecto a las ventas en el sector Comercio en el trimestre anterior.*
- *EmpleoConstrucción: Porcentaje de variación con respecto al número de empleados fijos en el sector Construcción en el trimestre anterior.*
- *InversionesConstrucción: Porcentaje de variación con respecto a la Inversión en el sector Construcción en el trimestre anterior.*
- *ImportacionesConstrucción: Porcentaje de variación con respecto a las importaciones en el sector Construcción en el trimestre anterior.*

- *Ventas Construcción: Porcentaje de variación con respecto a las ventas en el sector Construcción en el trimestre anterior.*

Con dichas variables se realizarán los siguientes análisis:

1. Un análisis descriptivo para cada variable. Media, varianza, gráfico etc. con el objetivo de ver el comportamiento que presenta la serie.
2. Se ajustará un modelo para cada variable global. Se buscará para cada variable global el modelo ARIMA que mejor la describa. Con ello estaremos identificando de que retardos depende cada variable.
3. Se realizará un análisis de correlaciones cruzadas entre las variables globales. Con ello se analizará que tan relacionadas están esas variables dos a dos.
4. Se calcularán distintas regresiones dinámicas: Con la regresión dinámica buscaremos una representación de la relación existente entre dos o más variables. Por ejemplo la inversión que se realiza en el período actual influirá en la producción de éste y en la de años posteriores.

Los modelos que se plantean para el estudio son los siguientes:

- *Empleo* en función de *Inversiones, Importaciones, Ventas* y retardos de la misma variable *Empleo*.
 - *Inversiones* en función de *Empleo, Importaciones, Ventas* y retardos de la misma variable *Inversiones*.
 - *Importaciones* en función de *Empleo, Inversiones, Ventas* y retardos de la misma variable *Importaciones*.
 - *Ventas* en función de *Empleo, Inversiones, Importaciones* y retardos de la misma variable *Ventas*.
5. Por último se realizará una comparación de todas las series temporales. Para ello se utilizará el índice cualitativo de similitud, QSI, que se calcula con dos variables y mide como de relacionadas están esas dos variables. Las comparaciones se realizarán por sectores y por tipo de variables.

3. MARCO TEÓRICO

En esta sección estudiaremos distintos aspectos teóricos de las series temporales. Una serie temporal es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares, cada día, cada mes, cada trimestre, cada año etc. Los aspectos descritos en esta sección, serán los utilizados con posterioridad para el estudio de los datos.

3.1. Análisis Univariante. Modelo ARIMA.

Antes de comenzar con la explicación del modelo ARIMA, debemos definir distintos conceptos de una serie temporal.

3.1.1. Descripción de series temporales.

Las series suelen representarse mediante un gráfico que muestra su evolución con el tiempo. Cuando se representa una serie se suele prestar atención a una serie de características que dan lugar a los distintos tipos de series.

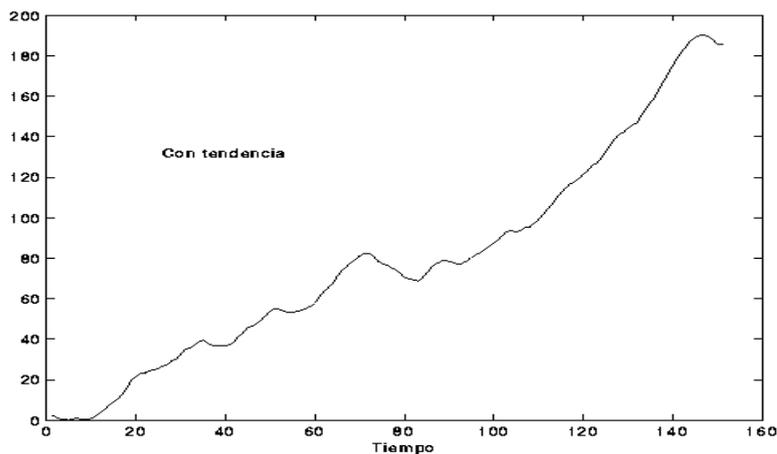


Figura 1: Serie con tendencia

Podemos clasificar las series temporales en los siguientes tipos:

- *Serie Estacionaria*: Cuando los valores oscilan alrededor de un nivel constante.
- *Serie No Estacionaria*: Cuando los valores no oscilan alrededor de un nivel constante. También se las conoce como series con tendencia
- *Serie Estacional*: Cuando tiene un comportamiento superpuesto que se repite a lo largo del tiempo.
- *Serie No Estacional*: Cuando no tiene un comportamiento superpuesto que se repite a lo largo del tiempo.
- *Serie homocedástica*: si su variabilidad se mantiene constante a lo largo de la serie.
- *Serie heterocedástica*: Cuando la variabilidad de la serie aumenta o disminuye a lo largo del tiempo.

Hasta ahora hemos descrito el tipo de serie. Sin embargo, cuando se quiere analizar la serie, es necesario identificar la estructura que la genera, es decir, cómo influyen las observaciones del pasado en las observaciones del futuro. Para identificar esta dependencia, utilizamos dos fuentes de información, la Función de Autocorrelación Simple (FAS) y la Función de Autocorrelación Parcial (FAP).

3.1.2. Funciones de autocorrelación.

Función de Autocorrelación Simple (FAS)

La función de autocorrelación simple en una serie, proporciona la estructura de dependencia lineal de la misma. En otras palabras la FAS proporciona información sobre cómo una observación influye en las siguientes.

Si denominamos z_t a la serie temporal, los valores que se observan van a ser: $z_1, z_2, z_3, \dots, z_t, z_{t+1}$, donde z_1 representa el primer valor de la serie, z_2 el segundo, y z_t será el valor actual de la serie.

Si denominamos, z_1 influye sobre z_2 , como $z_1 \rightarrow z_2$, la función de autocorrelación simple tiene por objetivo estudiar cómo se influyen las diversas observaciones, $z_1 \rightarrow z_2 \rightarrow \dots \rightarrow z_{t-1} \rightarrow z_t \rightarrow z_{t+1} \rightarrow \dots$

Más concretamente, la idea de la función de autocorrelación es proporcionar el coeficiente de correlación entre las observaciones separadas un número determinado de periodos. Así la FAS, va a ser una sucesión de números, $\rho_1, \rho_2, \rho_3, \dots, \rho_k, \dots$. Estos números representan cómo influye una observación sobre la siguiente (ρ_1), sobre la segunda posterior (ρ_2) o sobre la k periodos posterior (ρ_k).

Los coeficientes de la FAS, $\rho_1, \rho_2, \rho_3, \dots, \rho_k, \dots$ están acotados entre $[-1, 1]$. Cuando un ρ_i vale cero, quiere decir que no existe efecto entre una observación y las i posiciones posterior. Si ρ_i es próximo a 1 indica que hay mucha relación entre una observación y la i posición posterior, y que esa relación es positiva. Si ρ_i es próximo a -1 indica que hay mucha relación entre una observación y la i posición posterior, y que esa relación es negativa.

La FAS tiene un problema, y es que si por ejemplo ρ_1 es distinto de cero, entonces $z_1 \rightarrow z_2 \rightarrow \dots \rightarrow z_{t-1} \rightarrow z_t \rightarrow z_{t+1} \rightarrow \dots$. Es decir, existe una cadena de influencia separada por un retardo. Pero si $z_1 \rightarrow z_2$ y $z_2 \rightarrow z_3$ entonces $z_1 \rightarrow z_3$. Por tanto, la FAS, en general, si ρ_1 es distinto de cero, encontrará que $\rho_2, \rho_3, \dots, \rho_k, \dots$ etc., serán distintos de cero. Para resolver este problema se construye la Función de Autocorrelación Parcial.

Función de Autocorrelación Parcial (FAP)

La función de autocorrelación parcial proporciona la relación directa que existe entre observaciones separadas por k retardos. Esta es una información muy valiosa sobre la estructura de la serie, ya que elimina el problema que presentaba la función de autocorrelación simple de que si $z_1 \rightarrow z_2$ y $z_2 \rightarrow z_3$, entonces $z_1 \rightarrow z_3$. En la función de autocorrelación simple, el primer palo (relación entre z_1 y z_2 ó z_2 y z_3) será significativo,

y el segundo también, ya que si $z_1 \rightarrow z_2$ y $z_2 \rightarrow z_3$, entonces $z_1 \rightarrow z_3$. En la FAP, esto no ocurre, el primer palo será significativo, y el segundo no lo será.

Una vez definidas las funciones de autocorrelación, pasaremos a explicar como se logra que una serie sea estacionaria y homocedastica cuando no lo es en su origen. Esto será necesario dado que los modelos ARIMA son válidos solo para series estacionarias.

Transformaciones para conseguir estacionariedad y homocedasticidad.

Cuando una serie es no estacionaria, es preciso transformarla. El proceso más habitual para lograr que la serie sea estacionaria, es muy sencillo. Para eliminar la tendencia se toman una o varias diferencias en la serie. Una serie se diferencia restando a cada observación la observación anterior:

$$w_t = z_t - z_{t-1}$$

Es decir, la serie w_t se obtiene restando de cada observación de z_t la observación anterior. Evidentemente, la serie diferenciada w tiene una observación menos que la serie original z , ya que la primera observación se pierde. Si el gráfico de una serie muestra tendencia, se diferencia la serie y se comprueba si ha perdido la tendencia. En caso de que no la haya perdido se diferencia una segunda vez. Es muy raro necesitar más de una o dos diferencias para eliminar la tendencia de una serie. Si no se elimina con este procedimiento, debemos eliminarla con procesos más complejos.

Esta transformación se puede explicar mejor en función del operador de retardo y del de diferencia.

El operador de retardo, denotado por L , retrasa una posición la observación. Así por ejemplo $Lz_t = z_{t-1}$. Es decir, L aplicada a z_t retrasa ésta y la convierte en z_{t-1} . Aplicando el operador dos veces obtenemos, $L^2 z_t = L(Lz_t) = Lz_{t-1} = z_{t-2}$. Y aplicándolo k veces, obtenemos $L^k z_t = z_{t-k}$.

El operador de diferencia, denotado por ∇ , es equivalente a $1-L$, es decir, que tomar una diferencia equivale a multiplicar z_t por $1-L$, $\nabla z_t = w_t = (1-L)z_t = z_t - z_{t-1}$.

Cuando escribimos ∇z_t se indica que se ha tomado una diferencia de orden uno a la serie. Si se toman dos diferencias de orden uno, se representa por $\nabla^2 z_t$.

Para lograr que una serie heterocedástica, se transforme en una serie homocedástica, es necesario realizar algún tipo de transformación en la variable original. La transformación más habitual es mediante la función logaritmo.

En caso de una serie no sea homocedástica, ni tenga estacionariedad, en primer lugar, se debe eliminar la heterocedasticidad mediante una transformación, y con la variable ya transformada, eliminar la tendencia para lograr una serie estacionaria, si fuera necesario.

Una vez asumido que se han realizado las transformaciones necesarias para que nuestra serie sea estacionaria y homocedástica, pasaremos a definir los modelos más sencillos que integrarán el modelo o proceso ARIMA.

3.1.3. Procesos Autorregresivos.

Los procesos autorregresivos forman una familia de procesos tales que una observación depende de las observaciones anteriores. Se denominan procesos AR y se caracterizan por su orden.

Proceso Autorregresivo de primer orden AR(1).

El proceso autorregresivo de primer orden es el más sencillo de la familia de procesos autorregresivos. Se dice que un proceso o serie es autorregresivo de primer orden si sigue la siguiente ecuación:

$$z_t = \phi z_{t-1} + a_t + c = \phi L z_t + a_t \Leftrightarrow (1 - \phi L) z_t = a_t + c \quad (1)$$

Donde c y $-1 < \phi < 1$ son constantes a determinar y a_t es un proceso de ruido blanco con varianza σ^2 .

Si una serie sigue un proceso AR(1), cada observación se construye a partir de la anterior, más la perturbación aleatoria a_t . Por ejemplo, si $\phi = 0.8$ y $z_5 = 50$, podemos decir que $z_6 = 0.8 * 50 + a_t = 40 + a_t$. El papel de a_t es precisamente permitir que z_6

valga “algo” en torno a 40, y no exactamente 40. Si no existiera el término a_t , la serie sería determinista, y conocido un valor de z ya conoceríamos todos los demás. Como en la práctica esto no ocurre, es necesario introducir un término aleatorio que permita la introducción de errores.

La condición $-1 < \phi < 1$ es necesaria para que el proceso sea estacionario. Para comprobarlo, supongamos que el proceso comienza con $z_0 = h$, siendo h un valor cualquiera fijo. El valor siguiente será $z_1 = c + \phi h + a_1$ y sustituyendo sucesivamente, tenemos:

$$\begin{aligned} z_1 &= c + \phi h + a_1 \\ z_2 &= c(1 + \phi) + \phi^2 h + \phi a_1 + a_2 \\ z_3 &= c(1 + \phi + \phi^2) + \phi^3 h + \phi^2 a_1 + \phi a_2 + a_3 \\ &\dots \\ z_n &= c \sum_{i=0}^{t-1} \phi^i + \phi^t h + \sum_{i=0}^{t-1} \phi^i a_{t-i} \end{aligned}$$

Si calculamos la esperanza de z_t , como $E[a_t] = 0$,

$$E[z_t] = c \sum_{i=0}^{t-1} \phi^i + \phi^t h.$$

Para que sea estacionario es necesario que esta función no dependa de t . La media será constante si ambos sumandos lo son, lo que requiere que al aumentar t el primer término converja a una constante y el segundo se anule. Ambas condiciones se verifican

si $|\phi| < 1$, ya que entonces $\sum_{i=0}^{t-1} \phi^i$ es la suma de una progresión geométrica indefinida de

razón ϕ y converge a $\frac{c}{1-\phi}$, y el término ϕ^t converge a cero, con lo que la suma lo

hace a la constante $\frac{c}{1-\phi}$, que será la media del proceso. Con esta condición, después de

un periodo transitorio inicial, cuando $t \rightarrow \infty$, todas las variables z_t tendrán la misma

esperanza, $\mu = \frac{c}{1-\phi}$, independientemente de las condiciones iniciales.

Observemos también que en este proceso la innovación a_t está incorrelada con los valores previos del proceso, z_{t-k} para k positivo. En efecto el valor de z_{t-k} depende de los valores de las innovaciones hasta ese instante, a_1, \dots, a_{t-k} , pero no de sus valores

futuros. Como la innovación es un proceso de ruido blanco, sus valores futuros están incorrelados con los pasados y por tanto, con los valores previos, z_{t-k} .

Para la total descripción del modelo AR(1), es necesario conocer la esperanza, la varianza y las Funciones de autocorrelación.

Para el cálculo de la **esperanza**, debemos tomar las esperanzas en la ecuación (1)

$z_t = \phi z_{t-1} + a_t + c$, supuesto que $|\phi| < 1$, de manera que $E[z_t] = E[z_{t-1}] = \mu$, se obtiene

que: $\mu = c + \phi\mu \Rightarrow \mu = \frac{c}{1-\phi}$, sustituyendo c por $\mu(1-\phi)$, el proceso puede escribirse

en desviaciones a su media, $z_t - \mu = \phi(z_{t-1} - \mu) + a_t$ y llamando $\tilde{z} = z_t - \mu$,

$$\tilde{z}_t = \phi\tilde{z}_{t-1} + a_t \quad (2)$$

Que es una expresión alternativa para el proceso AR(1).

La **varianza** del proceso se obtiene elevando al cuadrado la expresión (2) y tomando esperanzas, con lo que tenemos:

$$\begin{aligned} E[\tilde{z}_t^2] &= E[(\phi\tilde{z}_{t-1} + a_t)^2] \\ &= E[\phi^2\tilde{z}_{t-1}^2 + 2\phi\tilde{z}_{t-1}a_t + a_t^2] \\ &= \phi^2 E[\tilde{z}_{t-1}^2] + 2\phi E[\tilde{z}_{t-1}a_t] + E[a_t^2] \end{aligned}$$

Llamaremos σ_z^2 a la varianza del proceso estacionario. El segundo término de esta última expresión es cero, ya que como \tilde{z}_{t-1} y a_t son independientes la esperanza de su producto es el producto de las esperanzas, y ambas variables tienen esperanza nula. El tercero es la varianza de la innovación, σ^2 , y concluimos que, $\sigma_z^2 = \phi^2\sigma_z^2 + \sigma^2$, de donde obtenemos que la varianza del proceso es:

$$\sigma_z^2 = \frac{\sigma^2}{1-\phi^2} \quad (3)$$

La **covarianza** entre las observaciones separadas k periodos es :

$$\gamma_k = E[(z_{t-k} - \mu)(z_t - \mu)] = E[\tilde{z}_{t-k}(\phi\tilde{z}_{t-1} + a_t)]$$

Y como $E[\tilde{z}_{t-k}a_t] = 0$, ya que las innovaciones están incorreladas con los valores pasados de la serie:

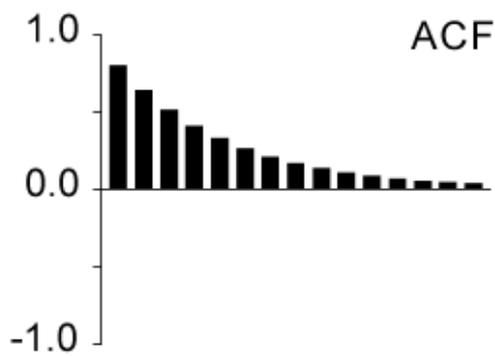
$$\gamma_k = \phi\gamma_{k-1} \quad k = 1, 2, \dots \quad (4)$$

Donde $\gamma_0 = \sigma_z^2$. Esta ecuación muestra que, como $|\phi| < 1$, la dependencia entre observaciones se amortigua al aumentar el retardo. En particular, utilizando (3):

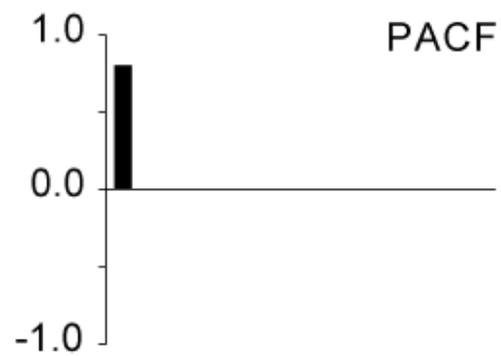
$$\gamma_1 = \frac{\phi\sigma^2}{1-\phi^2}$$

Las **funciones de autocorrelación**, FAS y FAP, (en inglés ACF y PACF) en un modelo AR(1) presentan una estructura diferente en función del signo del parámetro. En los siguientes gráficos se muestran dos ejemplos del comportamiento en función del signo.

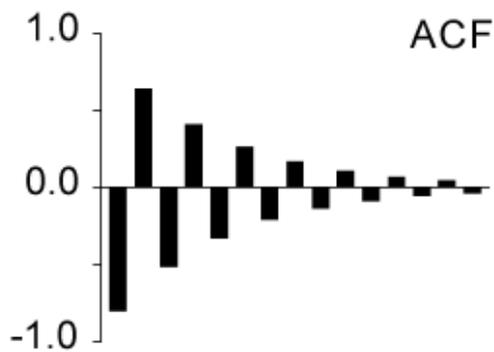
FAS $\phi > 0$



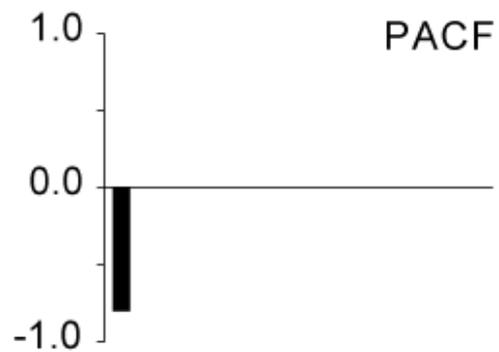
FAP $\phi > 0$



FAS $\phi < 0$



FAP $\phi < 0$



Proceso Autorregresivo de segundo orden AR(2).

La dependencia entre los valores presentes y los pasados que establece un proceso AR(1) puede generalizarse permitiendo que z_t dependa linealmente no sólo de z_{t-1} , sino también de z_{t-2} . Se obtiene entonces el proceso autorregresivo de segundo orden, o proceso AR(2) que tiene por ecuación:

$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} = \phi_1 L z_t + \phi_2 L^2 z_t \Leftrightarrow (1 - \phi_1 L - \phi_2 L^2) z_t = a_t + c \quad (5)$$

Donde c , ϕ_1 y ϕ_2 son constantes a determinar y a_t es un proceso de ruido blanco con varianza σ^2 .

Vamos a obtener qué condiciones tienen que verificar los parámetros del proceso para que sea estacionario. Tomando esperanzas en (5) e imponiendo que la media sea constante, resulta:

$$\mu = c + \phi_1 \mu + \phi_2 \mu,$$

que implica que la **media** para el proceso AR(2) es:

$$\mu = \frac{c}{1 - \phi_1 - \phi_2}$$

Y la condición para que el proceso tenga media finita es que: $1 - \phi_1 - \phi_2 \neq 0$.

Sustituyendo c por $\mu(1 - \phi_1 - \phi_2)$ y llamando $\tilde{z} = z_t - \mu$ al proceso de desviaciones a su media, el proceso AR(2) es:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \Leftrightarrow (1 - \phi_1 L - \phi_2 L^2) \tilde{z}_t = a_t \quad (6)$$

El operador $(1 - \phi_1 L - \phi_2 L^2)$ puede siempre expresarse como $(1 - G_1 L)(1 - G_2 L)$, donde G_1^{-1} y G_2^{-1} son las raíces de la ecuación

La **varianza** del proceso AR(2) la calculamos de la siguiente forma:

$$\begin{aligned} E[\tilde{z}_t^2] &= E[(\phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t)^2] \\ &= E[\phi_1^2 \tilde{z}_{t-1}^2 + \phi_2^2 \tilde{z}_{t-2}^2 + 2\phi_1 \phi_2 \tilde{z}_{t-1} \tilde{z}_{t-2} + 2\phi_1 \tilde{z}_{t-1} a_t + 2\phi_2 \tilde{z}_{t-2} a_t + a_t^2] \\ &= \phi_1^2 E[\tilde{z}_{t-1}^2] + \phi_2^2 E[\tilde{z}_{t-2}^2] + 2\phi_1 \phi_2 E[\tilde{z}_{t-1} \tilde{z}_{t-2}] + 2\phi_1 E[\tilde{z}_{t-1} a_t] + 2\phi_2 E[\tilde{z}_{t-2} a_t] + E[a_t^2] \end{aligned}$$

Y como $E[\tilde{z}_{t-k} a_t] = 0$, tenemos que:

$$\gamma_0 = \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + 2\phi_1 \phi_2 \gamma_1 + \sigma^2 \quad (7)$$

Para calcular las **covarianzas**, multipliquemos la ecuación (6) por \tilde{z}_{t-k} y tomemos esperanzas:

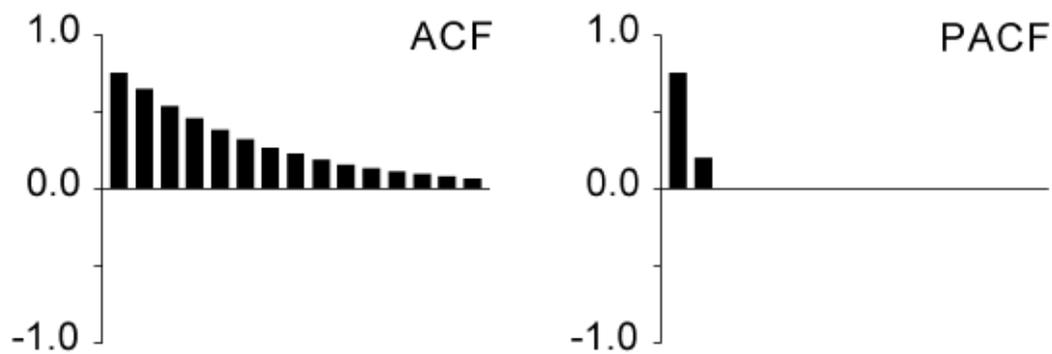
$$\begin{aligned} E[\tilde{z}_t \tilde{z}_{t-k}] &= E[(\phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t) \tilde{z}_{t-k}] \\ &= E[\phi_1 \tilde{z}_{t-1} \tilde{z}_{t-k} + \phi_2 \tilde{z}_{t-2} \tilde{z}_{t-k} + a_t \tilde{z}_{t-k}] \\ &= \phi_1 E[\tilde{z}_{t-1} \tilde{z}_{t-k}] + \phi_2 E[\tilde{z}_{t-2} \tilde{z}_{t-k}] + E[a_t \tilde{z}_{t-k}] \end{aligned}$$

Por lo que la expresión general de las autocovarianzas es:

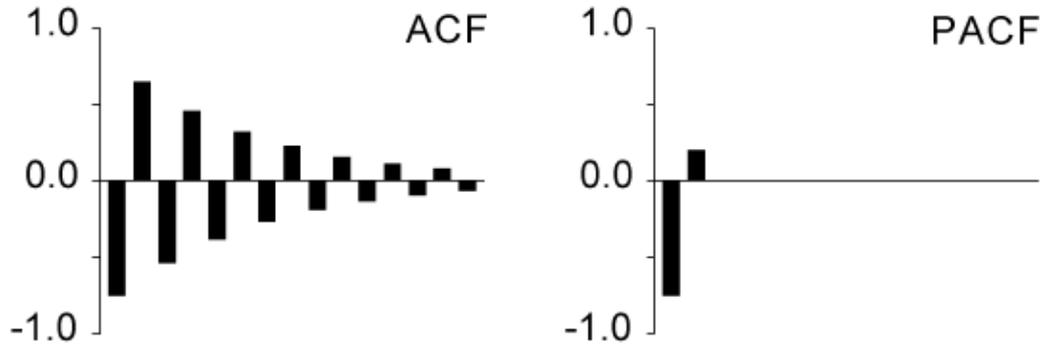
$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \quad k \geq 1 \quad (8)$$

Las **funciones de autocorrelación (FAS y FAP)** en un modelo AR(2) presentan una estructura diferente en función de las raíces del denominador polinomio característico del proceso $1 - \phi_1 L - \phi_2 L^2 = 0$. En los siguientes gráficos se muestran ejemplos del comportamiento en función de estas funciones.

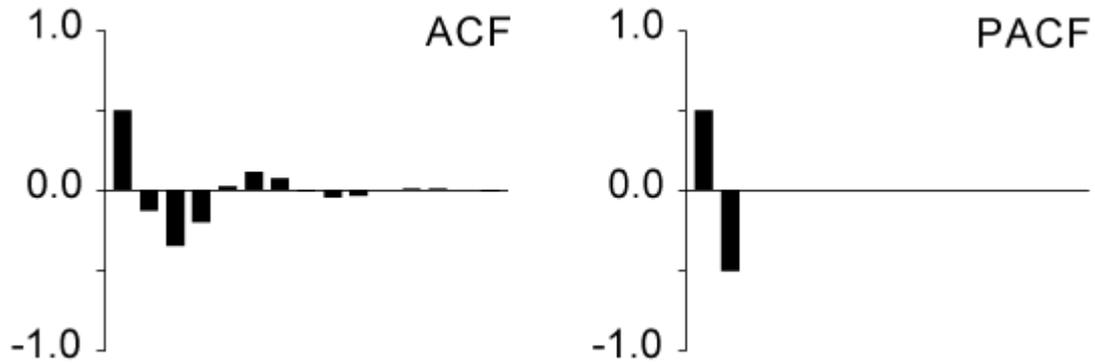
- Dos raíces reales positivas. En este caso la FAS es la superposición de dos exponenciales decrecientes como las mostradas en las figuras. La FAP tendrá los dos primeros palos.



- Raíces reales negativas. La FAS será la superposición de dos exponenciales decrecientes con los palos de signos alternados como las que se mostraron en las figuras. La FAP tendrá dos palos.



- Raíces imaginarias. En este caso la FAS de la serie tiene un aspecto cíclico. La FAP tendrá dos únicos palos



Proceso autorregresivo general AR(p).

Diremos que una serie temporal z_t estacionaria sigue un proceso autorregresivo de orden p si:

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + c + a_t \quad (9)$$

Donde a_t es un proceso de ruido blanco.

La **media** la calculamos tomando esperanzas de (9) y sabiendo que $E[z_t] = E[z_{t-1}] = \dots = E[z_{t-p}] = \mu$, obtenemos que:

$$\begin{aligned} E[z_t] &= E[\phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + c + a_t] \\ E[z_t] &= \phi_1 E[z_{t-1}] + \dots + \phi_p E[z_{t-p}] + c + E[a_t] \\ \mu &= \phi_1 \mu + \dots + \phi_p \mu + c \end{aligned}$$

por lo que :

$$\mu = \frac{c}{1 - \phi_1 - \dots - \phi_p}$$

La ecuación (9) la podemos describir haciendo $\tilde{z} = z_t - \mu$, siendo μ la media del proceso estacionario z_t , como:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t \Leftrightarrow (1 - \phi_1 L - \dots - \phi_p L^p) \tilde{z}_t = a_t \quad (10)$$

Llamando $\phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ al polinomio de grado p en el operador de retardo, cuyo primer término es la unidad, tenemos:

$$\phi_p(L) \tilde{z}_t = a_t$$

Que es la expresión general de un proceso autorregresivo.

Llamaremos *ecuación característica* del proceso a la ecuación:

$$\phi_p(L) = 0$$

Considerada como función de L . Esta ecuación tendrá p raíces $G_1^{-1}, \dots, G_p^{-1}$, en general distintas, y podemos escribir:

$$\phi_p(L) = \prod_{i=1}^p (1 - G_i L)$$

De manera que los coeficientes G_i son los factores de la ecuación característica. Se demuestra que el proceso es estacionario si $|G_i| < 1$, para todo i .

Para calcular la **varianza** y las **covarianzas** multiplicamos (10) por \tilde{z}_{t-k} ($k > 0$) y tomando esperanzas a ambos lados tenemos:

$$\begin{aligned} E[\tilde{z}_t \tilde{z}_{t-k}] &= E[(\phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t) \tilde{z}_{t-k}] \\ &= E[\phi_1 \tilde{z}_{t-1} \tilde{z}_{t-k} + \dots + \phi_p \tilde{z}_{t-p} \tilde{z}_{t-k} + a_t \tilde{z}_{t-k}] \\ &= \phi_1 E[\tilde{z}_{t-1} \tilde{z}_{t-k}] + \dots + \phi_p E[\tilde{z}_{t-p} \tilde{z}_{t-k}] + E[a_t \tilde{z}_{t-k}] \end{aligned}$$

Por lo que :

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma^2 \\ \gamma_k &= \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} \quad k > 0 \end{aligned} \quad (11)$$

El comportamiento de las **funciones de autocorrelación** será la siguiente: La FAS de un proceso $AR(p)$ tendrá siempre una tendencia amortiguada hacia cero. En caso que las raíces del polinomio sean complejas, la forma de la FAS es sinusoidal, pero también amortiguada hacia cero. Las FAS de un $AR(p)$ tendrá los primeros p polos diferentes de cero.

3.1.4. Procesos de Media Móvil.

Los procesos autorregresivos que se han introducido tienen una característica común, todos ellos tienen memoria larga. La memoria larga aparece en que los polos de la FAS van decreciendo lentamente. Así, un proceso autorregresivo tarda bastante tiempo en absorber los impactos externos. Por ejemplo, si una serie es autorregresiva y se produce una perturbación externa, la serie tarda mucho tiempo en absorber el impacto. La familia de procesos matemáticos que representa los procesos de memoria corta se denomina genéricamente procesos de media móvil (MA). Estos procesos MA son funciones de un número finito, y generalmente pequeño, de las innovaciones pasadas.

Proceso de Media Móvil de primer orden. MA(1)

El proceso de media móvil de primer orden tiene por ecuación:

$$z_t = \mu + a_t - \theta a_{t-1} \quad (12)$$

En esta ecuación, a_t denota la innovación y representa los efectos externos a la serie. Si un proceso es MA(1), sus valores estarán formados por los efectos externos muy recientes, el actual a_t y el inmediatamente anterior a_{t-1} . El proceso MA tiene memoria corta, y absorbe rápidamente los impactos. En las aplicaciones de este proceso supondremos que $|\theta| < 1$, de manera que la innovación pasada tenga menos peso que la presente.

La media del proceso se calcula tomando esperanzas en (12) y considerando que $E[a_t] = E[a_{t-1}] = 0$, obtenemos:

$$\begin{aligned}
E[z_t] &= E[\mu + a_t - \theta a_{t-1}] \\
&= E[\mu] + E[a_t] + \theta E[a_{t-1}] \\
&= \mu
\end{aligned}$$

La ecuación (12) puede ser rescrita llamando $\tilde{z} = z_t - \mu$ al proceso de desviaciones a su media como:

$$\tilde{z}_t = a_t - \theta a_{t-1} \Leftrightarrow \tilde{z}_t = (1 - \theta L)a_t \quad (13)$$

La **varianza** del proceso se calcula a partir de (13). Elevando al cuadrado y tomando esperanzas, se obtiene:

$$\begin{aligned}
E[\tilde{z}_t^2] &= E[(a_t - \theta a_{t-1})^2] \\
&= E[a_t^2 + \theta^2 a_{t-1}^2 - 2\theta a_t a_{t-1}] \\
&= E[a_t^2] + \theta E[a_{t-1}^2] - 2\theta E[a_t a_{t-1}]
\end{aligned}$$

como $E[a_t a_{t-1}] = 0$, por ser un proceso de ruido blanco, y $E[a_t^2] = E[a_{t-1}^2] = \sigma^2$, tenemos que:

$$\sigma_z^2 = \sigma^2(1 + \theta^2).$$

Esta ecuación nos indica que la varianza marginal del proceso, σ_z^2 , es siempre mayor que la varianza de las innovaciones, σ^2 , y por tanto mayor cuanto mayor sea θ^2 .

La **covarianza** de primer orden se obtiene multiplicando la ecuación (14) por \tilde{z}_{t-1} y tomando esperanzas, se obtiene:

$$\gamma_1 = E[\tilde{z}_t \tilde{z}_{t-1}] = E[a_t \tilde{z}_{t-1}] - \theta E[a_{t-1} \tilde{z}_{t-1}]$$

En esta expresión el primer término $E[a_t \tilde{z}_{t-1}]$ es cero, ya que \tilde{z}_{t-1} según (13) depende de su innovación, a_{t-1} , y de la innovación previa, a_{t-2} , pero no de las innovaciones futuras, como a_t . Para calcular el segundo término, sustituyendo \tilde{z}_{t-1} por su expresión según (14), resulta:

$$E[a_{t-1} \tilde{z}_{t-1}] = E[a_{t-1}(a_{t-1} - \theta a_{t-2})] = \sigma^2$$

Con lo que se obtiene:

$$\gamma_1 = -\theta \sigma^2 \quad (14)$$

De la misma forma, se calcula la covarianza de orden dos:

$$\gamma_2 = E[\tilde{z}_t \tilde{z}_{t-2}] = E[a_t \tilde{z}_{t-2}] - \theta E[a_{t-1} \tilde{z}_{t-2}] = 0$$

Ya que al estar la serie incorrelada con sus innovaciones futuras, los dos términos son nulos. El mismo resultado se obtiene para las covarianzas de orden superior a dos, ya que al multiplicar la ecuación (14) por \tilde{z}_{t-j} , donde $j > 1$, tendremos productos de la serie por sus innovaciones futuras y las esperanzas de estos términos son cero. En conclusión:

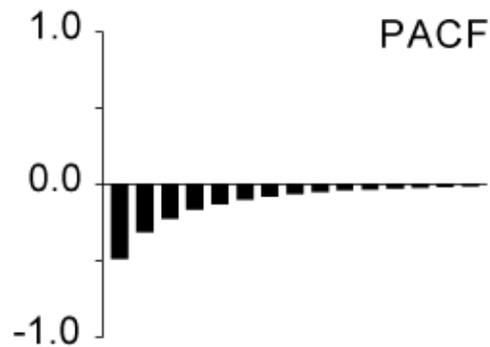
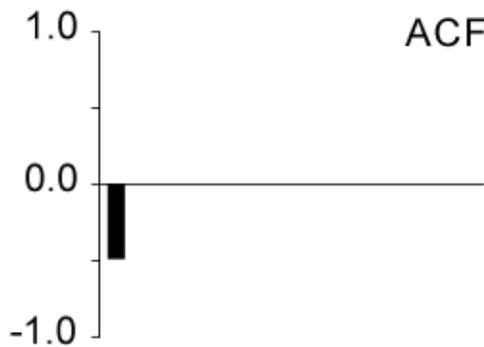
$$\gamma_j = 0, j > 1 \quad (15)$$

En general, el comportamiento de las **funciones de autocorrelación** de un MA es la misma que de un AR pero intercambiadas. Por lo que para un MA(1) las FAS tendrá un único palo, y la FAP tendrá un comportamiento exponencial.

A continuación presentemos algunos ejemplos:

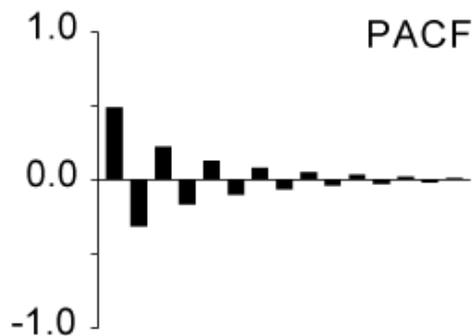
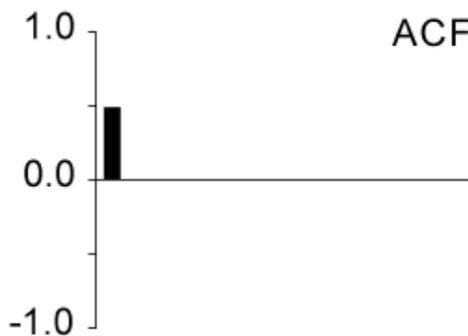
FAS $\theta > 0$

FAP $\theta > 0$



FAS $\theta < 0$

FAP $\theta < 0$



Proceso de Media Móvil de segundo orden . MA(2)

El proceso de media móvil de segundo orden tiene por ecuación:

$$z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (16)$$

En esta ecuación, a_t denota la innovación y representa los efectos externos a la serie.

La **media** del proceso MA(2) la calculamos tomando en (16) esperanzas y considerando que $E[a_t] = E[a_{t-1}] = E[a_{t-2}] = 0$, obtenemos:

$$\begin{aligned} E[z_t] &= E[\mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}] \\ &= E[\mu] + E[a_t] - \theta_1 E[a_{t-1}] - \theta_2 E[a_{t-2}] \\ &= \mu \end{aligned}$$

La ecuación (16) puede ser rescrita llamando $\tilde{z}_t = z_t - \mu$ al proceso de desviaciones a su media como:

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \Leftrightarrow \tilde{z}_t = (1 - \theta_1 L - \theta_2 L^2) a_t \quad (17)$$

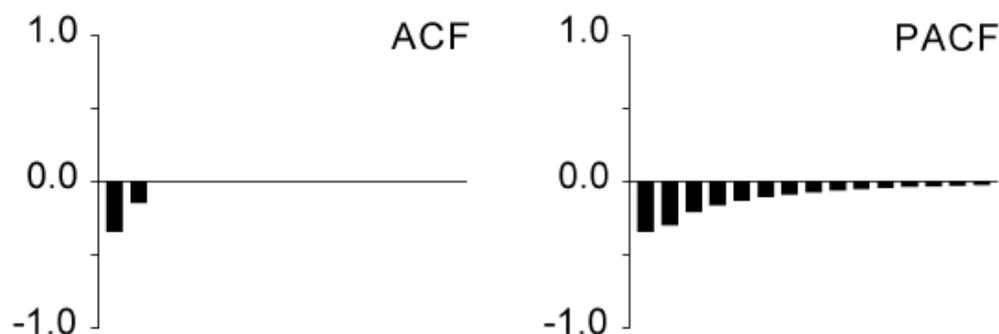
De manera similar que en el AR(2) tenemos que la ecuación característica para el modelo MA(2) es: $1 - \theta_1 L - \theta_2 L^2 = 0$

Multiplicando (17) por \tilde{z}_{t-k} para $k \geq 0$ y tomando esperanzas se obtienen la **varianza** y las **covarianzas**:

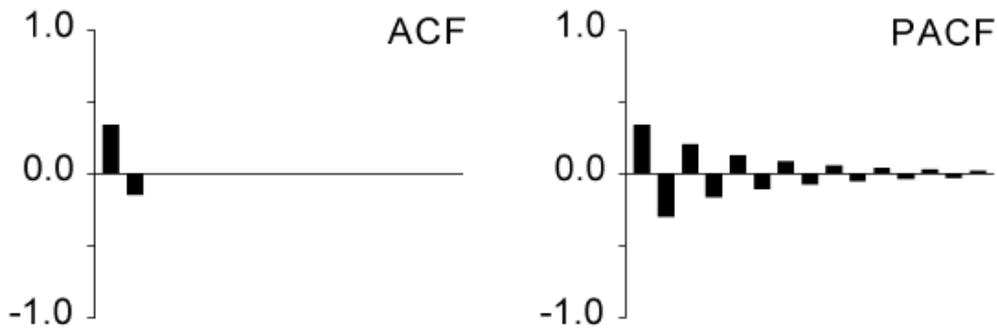
$$\begin{aligned} \gamma_0 &= (1 + \theta_1^2 + \theta_2^2) \sigma^2 \\ \gamma_1 &= (-\theta_1 + \theta_1 \theta_2) \sigma^2 \\ \gamma_2 &= -\theta_2 \sigma^2, \\ \gamma_k &= 0, k > 2 \end{aligned}$$

Las **funciones de autocorrelación** de un MA(2) se comportan de manera similar que las FAS y la FAP de un AR(2) pero de manera intercambiadas, y dependen de las raíces de la ecuación característica. Presentamos algunos ejemplos:

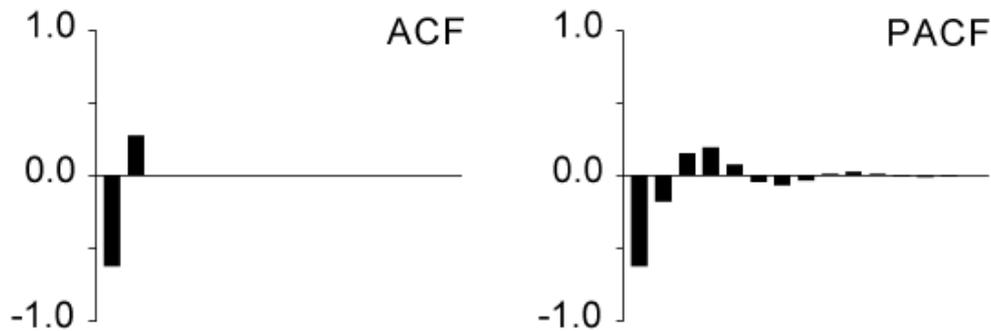
- Raíces positivas



- Raíces Negativas



- Raíces complejas



Proceso de Media Móvil generalizado MA(q).

Generalizando la idea de un MA(1) y de un MA(2), podemos escribir procesos cuyo valor actual dependa no sólo de la última innovación, sino de las q últimas innovaciones. Se obtiene entonces el proceso MA(q), con representación general:

$$z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_p a_{t-p} \quad (18)$$

La **media** para un MA(q) se calcula tomando esperanzas en (18), y considerando que $E[a_t] = E[a_{t-1}] = \dots = E[a_{t-p}] = 0$, obtenemos:

$$\begin{aligned} E[z_t] &= E[\mu + a_t - \theta_1 a_{t-1} - \dots - \theta_p a_{t-p}] \\ &= E[\mu] + E[a_t] - \theta_1 E[a_{t-1}] - \dots - \theta_p E[a_{t-p}] \\ &= \mu \end{aligned}$$

Haciendo $\tilde{z}_t = z_t - \mu$, podemos reescribir la ecuación (18) como:

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \Leftrightarrow \tilde{z}_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) a_t \quad (19)$$

La ecuación anterior puede escribirse de forma más compacta como:

$$\tilde{z}_t = \theta_q(L) a_t$$

Un proceso MA(q) es siempre estacionario, por ser la suma de procesos estacionarios. Diremos que el proceso es invertible si las raíces del operador $\theta_q(L)=0$ son, en módulo, mayores que la unidad.

Las propiedades de este proceso se obtienen con el mismo método utilizando para el MA(1).

Multiplicando (19) por \tilde{z}_{t-k} para $k \geq 0$ y tomando esperanzas se obtienen la **varianza** y las **covarianzas**:

$$\begin{aligned} \gamma_0 &= (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2 \\ \gamma_k &= (-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma^2 \quad k = 1, \dots, q \\ \gamma_k &= 0, \quad k > q \end{aligned}$$

Para la determinación de las **funciones de autocorrelación** de un MA(q), podemos generalizar lo que pasa con el MA(1) y el MA(2). Existe una dualidad entre procesos AR y MA, de manera que la FAP de un MA(q) tiene la estructura de la FAP de un AR(q), y la FAS de un MA(q) tiene la estructura de la FAP de un AR(q)

3.1.5. Procesos ARMA(p,q)

En la práctica se observa que las series no son puras AR o MA, sino que presentan parte AR y parte MA. Los procesos ARMA(p,q) son combinación de estructuras autorregresivas y de media móvil que tienen una parte AR(p) y una parte MA(q). Su FAS y su FAP serán combinación de ambos procesos y tendrán la siguiente estructura:

- **FAS:** Los primeros q palos de la FAS vendrán establecidos por la parte MA. Estos palos le van a permitir a la serie absorber rápidamente los impactos externos, tal como corresponde a la parte MA. A partir del retardo q se producirá un decrecimiento de los palos que vendrá dado por la estructura AR.
- **FAP:** Los primeros p palos de la FAP vendrán establecidos por la parte AR. A partir del retardo p se producirá un decrecimiento de los palos que vendrá dado por la estructura MA.

3.1.6. Series No Estacionarias. Procesos ARIMA(p,d,q)

Los modelos ARMA únicamente sirven para ajustar series estacionarias. Si la serie no lo es y se le aplican diferenciaciones para transformarla en una serie estacional, el modelo que diremos que hemos ajustado a la serie sin diferenciar será un modelo ARIMA(p,d,q), donde:

- p representa el orden de la parte autorregresiva de la serie estacionaria.
- q representa el orden de la media móvil de la serie estacionaria.
- d representa el número de diferencias que ha habido que tomar para que la serie que, inicialmente no era estacionaria, sea finalmente estacionaria.

Podemos concluir por tanto que la representación general de una serie será el modelo ARIMA(p,d,q). Así, si una serie requiere 2 diferencias para ser estacionaria en tendencia, y hay que transformarla a Logaritmos para estabilizar la varianza, y se le identifica un modelo AR(2) una vez que se ha vuelto estacional, decimos que la serie $\log z_t$ sigue un modelo ARIMA(2,2,0). Es decir:

$$\nabla \log(z_t) = \phi_1 \nabla \log(z_{t-1}) + \phi_2 \nabla \log(z_{t-2}) + a_t.$$

3.1.7. Modelos Estacionales. Procesos ARIMA(p,d,q)(ps, ds,qs)s.

Muchas series con periodicidad menor que la anual tienen estacionalidad. La estacionalidad aparece en forma de ciclo, de modo que la serie tiene una pauta que se repite año tras año. El orden de la estacionalidad indica cada cuantos valores de la variable se repite el ciclo. A este valor se le conoce como periodo. Así, en datos mensuales es normal encontrar estacionalidad de orden 12. En datos trimestrales la estacionalidad será de orden 4, en datos cuatrimestrales la estacionalidad será de orden 3 y así sucesivamente.

Las series estacionales tienen dos tipos de estructura de dependencia.

- Estructura regular. Se ajustan modelos ARIMA(p,d,q) .

- Estructura estacional. Se ajustan modelos ARIMA estacionales. Denotados como $ARIMA(p_s, d_s, q_s)_s$, donde s es el período.

Para ajustar los parámetros de la parte estacional se utiliza la misma metodología que en la parte regular, pero solo me debo fijar en los palos separados por s posiciones. Es decir, si es una serie con periodo 12, me debo fijar en los palos, 12, 24, 36, 48, ... de las dos funciones de autocorrelación.

Si es preciso realizar diferenciaciones estacionales, por no ser una serie estacionaria en la parte estacional, simplemente diferenciamos teniendo en cuenta el periodo, es decir, la nueva serie será, $w_t = z_t - z_{t-s} = (1 - L^s) z_t$

En conclusión en un modelo $ARIMA(p,d,q)(p_s,d_s,q_s)_s$ los parámetros son:

- p representa el orden de la parte autorregresiva de la serie estacionaria en la estructura regular.
- q representa el orden de la media móvil de la serie estacionaria en la parte regular.
- d representa el número de diferencias que ha habido que tomar para que la serie que, inicialmente no era estacionaria, sea finalmente estacionaria en la estructura regular.
- p_s representa el orden de la parte autorregresiva de la serie estacionaria en la estructura estacional.
- q_s q representa el orden de la media móvil de la serie estacionaria en la estructura estacional.
- d_s representa el número de diferencias que ha habido que tomar para que la serie que, inicialmente no era estacionaria, sea finalmente estacionaria en la estructura estacional.

3.2. Análisis Multivariante. Regresión Dinámica.

Un modelo de regresión dinámica es una representación de la relación existente entre dos o más series temporales. Para su determinación se deben considerar dos factores:

- Las series temporales que pueden ser utilizadas en el modelo. Utilizaremos las funciones de correlaciones cruzadas para tal efecto.
- El tipo de modelo en el que se integrarán estas variables. Utilizaremos el modelo de Función de Transferencia.

3.2.1. Funciones de correlaciones cruzadas.

Antes de definir las correlaciones cruzadas es necesario definir las covarianzas cruzadas. Que posteriormente utilizaremos para definir la función de correlaciones cruzadas.

Covarianzas cruzadas.

Supongamos que tenemos dos procesos estacionarios x_t , y_t , con medias μ_x, μ_y .

Definimos las *covarianzas cruzadas*, mediante:

$$\gamma_{xy}(t, t+k) = E[(x_t - \mu_x)(y_{t+k} - \mu_y)].$$

Esta función depende del orden en que se tomen las variables y de los instantes en que se consideran cada una de ellas.

La definición establece que la variable que aparece en primer lugar en el subíndice, x , se supone en t y la segunda, y , en $t+k$

Función de correlación cruzada.

Definimos la función de correlación cruzada, $\rho_{xy}(k)$, de dos procesos estocásticos conjuntamente estacionarios, x_t , y_t , por:

$$\rho_{xy} = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \quad (k = 0, \pm 1, \pm 2, \dots)$$

$\rho_{xy}(t, t+k)$ mide la relación lineal entre las dos variables x_t y y_{t+k} . El valor de k puede ser positivo o negativo. Se verifica que $\rho_{xy}(t, t+k)$ es idéntica a $\rho_{yx}(t+k, t)$.

La función de autocorrelación cruzada verifica las siguientes propiedades:

1. $|\rho_{xy}(k)| \leq 1$;

2. $\rho_{xy}(k) = \rho_{yx}(-k)$, pero en general $\rho_{xy}(k) \neq \rho_{xy}(-k)$, y la función no es simétrica alrededor del origen. Para $k > 0$ mide la relación desde x hacia y , mientras que para $k < 0$ la relación desde y hacia x .

3.2.2. Modelo de Función de Transferencia.

En modelo de Función de Transferencia se toman dos series, esto para el caso más simple, porque se puede disponer de más de dos. Para el caso más sencillo, 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 :¹

$$Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + N_t = v(L) X_t + N_t, \text{ donde, } v(L) = v_0 + v_1 L + v_2 L^2 + \dots$$

Siendo N_t 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:

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

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

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

Donde:

¹ Denotaremos a las series originales con letras mayúsculas, dejando la notación de letras minúsculas a las series estacionarias.

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

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

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 ello 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 autorregresivos y de medias móviles, con lo cual el modelo de función de transferencia se escribe de la siguiente forma:

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

Donde:

$$y_t = (1 - L)^{d'} Y_t$$

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

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 σ_u^2 lo que se conoce como ruido blanco.

El grado del denominador determina la complejidad del modelo de transferencia creado. A mayor grado del denominador mayor complejidad del modelo. Por este motivo intentaremos definir modelos en los que el grado del polinomio del denominador sea 0. Es decir:

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

3.3. Comparación de series temporales. Índice cualitativo de similitud.

Comparar dos o más series temporales es un problema con muchas aplicaciones, en diferentes campos. Por ejemplo: En Geología, interesa la detección de diferencias en la forma de las ondas producidas por terremotos ó por explosiones nucleares. En Medicina, la comparación de diferentes secciones de una señal biomédica se utiliza como procedimiento de diagnóstico. En Economía, habitualmente, es interesante comparar los tipos de interés, las tasas de inflación en diferentes regiones o países, o los distintos datos macroeconómicos. Todas estas comparativas son una herramienta clave para el análisis estructural y para conocer cómo funciona la Economía.

La mayoría de los métodos de comparación de series temporales necesitan: la selección y estimación de modelos; la estimación del espectro de las series; la independencia de las mismas. Presentaremos un par de ideas para abordar la comparativa.

La primera idea es que la comparación de dos series se puede hacer mediante subsampling (ver Alonso y Maharaj, 2006). En este caso, la clave es plantear un contraste en el que la hipótesis nula es que las medias poblacionales son iguales y el estadístico se obtiene tras comparar submuestras de k observaciones consecutivas de ambas series, variando k .

La segunda idea es que la comparación se base en características cualitativas de las series. Dos ejemplos pueden encontrarse en Martín (2005), que compara series de PIB de diferentes países, y en Ortega et al. (2001). En este último caso, se asignan etiquetas cualitativas a las series y se define un índice de similitud en función de la similitud entre las cadenas de etiquetas, conocido como índice cualitativo de similitud (QSI). Los principales inconvenientes de este método son su fuerte dependencia del proceso de normalización y la complejidad que subyace si se quiere realizar la comparación con un mínimo de rigor, sobre todo cuando existe algún tipo de irregularidad en la periodicidad con la que se han obtenido los datos.

De estos dos inconvenientes el primero de ellos se puede minimizar eligiendo adecuadamente el proceso de normalización. Por ello pasaremos a exponer los principales métodos de normalización existentes.

3.3.1. Normalización

En realidad, el primer problema que nos podemos encontrar para comparar series es que los datos de dos series X_n y Y_n sean heterogéneos o que las escalas en las que estén medidos sean muy diferentes. Para obtener valores homogéneos, o con escalas comparables, y así operar con ellos, vamos a utilizar la técnica de normalización, que es la operación mediante la cual un conjunto de valores de una determinada magnitud son transformados en otros, de forma que estos pertenezcan a una escala predeterminada. La normalización puede realizarse de los siguientes modos: sin cambio de magnitud; con cambio de magnitud a escala libre; o con cambio de magnitud a escala fija. Según Barba et al. (1997), los procedimientos más destacados en la actualidad se pueden resumir en cuatro:

1. $\tilde{x}_i = \frac{x_i}{\max(x_i)}$
2. $\tilde{x}_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}$
3. $\tilde{x}_i = \frac{x_i}{\sum_{i=1}^n x_i}$
4. $\tilde{x}_i = \frac{x_i}{\sqrt{\sum_{i=1}^n x_i}}$

siendo $\max.$ y $\min.$ las operaciones que devuelven los valores máximo y mínimo de la serie temporal X_n . Los procedimientos primero, tercero y cuarto corresponden a una linealización pura, mientras que el segundo corresponde a una linealización con ordenada en el origen. Los métodos de normalización citados pueden llegar a concentrar los valores normalizados, lo cual en ocasiones puede ser un problema añadido. Para intentar solucionar dicho problema se define otro procedimiento de normalización, propuesto por Cloquell (1999), que se resume con: $\tilde{x}_i = P_{F(x)} [x \leq x_i]$.

Es decir, el valor normalizado es igual a la probabilidad de que un número del intervalo sea menor o igual que el valor a normalizar, conocida la función de distribución del conjunto de valores a normalizar. Un problema que presenta este procedimiento de normalización es conocer o estimar la función de distribución de los valores que se desean normalizar.

Una vez definido los distintos procesos de normalización pasaremos a definir el concepto de Subsecuencia Común Máxima, concepto necesario para la definición del QSI.

3.3.2 Subsecuencia Común Máxima (LCS)

Antes de definir el concepto de Subsecuencia Común Máxima, debemos definir el concepto de lenguaje de definición de formas. El lenguaje de definición de formas (Shape Definition Language) propuesto en (Agrawal 1995) es muy apropiado para realizar consultas sobre la forma de evolución de valores a lo largo del tiempo.

Dado un conjunto de valores registrados durante un período, la primera idea en *SDL* es dividir el intervalo de los posibles valores de las variaciones entre sucesivos valores de la serie en rangos disjuntos y asignarle a cada uno una etiqueta.

El comportamiento de una serie puede describirse considerando las transiciones entre las sucesivas muestras. Se construye una derivada con respecto al tiempo de la serie, calculando la diferencia de amplitud entre las sucesivas muestras. El valor de esas diferencias se encuadra en uno de los diferentes rangos disjuntos definidos proporcionando una etiqueta del alfabeto.

Así la traducción produce una secuencia de transiciones basada en un alfabeto cuyos símbolos describen la magnitud de los incrementos de los valores de la serie temporal. El tamaño del alfabeto es muy pequeño comparado con los diferentes valores reales que la curva toma a lo largo del tiempo. Una traducción de este tipo permite dar prioridad a la forma sobre los valores originales. Cada cadena de símbolos puede describir un infinito número de curvas que cumplan con las restricciones impuestas por los símbolos a las transiciones que representan.

Una vez transformada la serie numérica en una cadena de caracteres, una de las medidas de similitud más usadas entre dos cadenas de caracteres es la Subsecuencia Común Máxima (*LCS*). *LCS* es la mayor colección de elementos que se encuentran en el mismo orden en dos secuencias distintas.

Formalmente. Sean $S = \{s_1, s_2, \dots, s_m\}$ y $T = \{t_1, t_2, \dots, t_n\}$ sobre un alfabeto Σ , entonces $LCS(S, T) = U$, $U = \{u_1, u_2, \dots, u_r\}$ tal que existen índices $i_1 < i_2 < \dots < i_r$, con $1 \leq r \leq m$, y $j_1 < j_2 < \dots < j_r$, con $1 \leq r \leq n$, tal que: $s_{i_z} = t_{j_z} = u_z$

En conclusión, el interés de usar el *LCS* es el siguiente: Como el lenguaje *SDL* produce una cadena de símbolos desde los valores numéricos de la serie temporal, es posible entonces usar este algoritmo para obtener una distancia entre dos series con abstracción en las formas de las curvas.

Llegados a este punto, solo nos resta definir formalmente el índice cualitativo de similitud (*QSI*).

3.3.3 Índice Cualitativo de Similitud (*QSI*)

El índice cualitativo de similitud (*QSI*) es una medida de similitud basada en la coincidencia de las etiquetas cualitativas que presentan la evolución de los valores de las series. Cada etiqueta representa un rango de valores que desde una perspectiva cualitativa podemos entender como similares.

La comparación de series temporales a través del *QSI*, obtiene buenos resultados por dos motivos. Por un lado, el uso de toda la información contenida en la serie temporal maximiza la exactitud. Por otro, la consideración de grupos de evoluciones como similares prioriza la focalización de la comparación en la forma general de las curvas y no de sus valores puntuales.

La aproximación se realiza en tres pasos. Sea $X = \{x_0, \dots, x_f\}$ una serie temporal, primero se realiza una normalización de los valores de la serie temporal, obteniéndose

$\tilde{X} = \langle \tilde{x}_0, \dots, \tilde{x}_f \rangle$ y a partir de ella se obtiene la serie de diferencias $X_D = \langle d_0, \dots, d_{f-1} \rangle$, que se traduce a una cadena de caracteres $S_X = \langle c_1, \dots, c_{f-1} \rangle$. La similitud entre dos series temporales se obtiene comparando las dos cadenas obtenidas de la transformación anterior mediante un logaritmo *LCS*.

Vamos a continuación de manera detallada cada una de estos pasos:

Normalización

En primer lugar y con la intención de poder comparar cualitativamente las series se realiza una normalización de sus valores al intervalo $[0,1]$.

Sea $X = \langle x_0, \dots, x_f \rangle$ una serie temporal. A partir de ella se obtiene la serie temporal normalizada, representada como $\tilde{X} = \langle \tilde{x}_0, \dots, \tilde{x}_f \rangle$ donde:

$$\tilde{x}_i = \frac{x_i - \min(x_0, \dots, x_f)}{\max(x_0, \dots, x_f) - \min(x_0, \dots, x_f)}$$

Siendo *min* y *max* operaciones que devuelven los valores mínimo y máximo de una secuencia de números. Si se considera oportuno se pueden utilizar otros métodos de normalización.

Diferenciación

A partir de esta serie normalizada se obtiene la serie de diferencias $X_D = \langle d_0, \dots, d_{f-1} \rangle$ donde:

$$d_i = \tilde{x}_i - \tilde{x}_{i-1}$$

Esta serie de diferencias será utilizada posteriormente en el etiquetado para obtener la cadena de caracteres correspondientes a la serie temporal. Es interesante comentar que cualquier $d_i \in X_D$ será un número en el intervalo $[-1,1]$ con la normalización indicada.

Etiquetado

La normalización propuesta anteriormente considera la evolución de la pendiente en lugar de los valores de la serie. Con la intención de asignar una etiqueta a cada tipo de pendiente, se divide el intervalo de las posibles pendientes en varios grupos y se le asigna a cada uno una etiqueta cualitativa.

La anterior división se realiza con un parámetro δ . El valor de este parámetro influye en la calidad de los resultados, sin embargo aún no se ha realizado un estudio detallado sobre su importancia en los resultados ni sobre su relación con las series pertenecientes al dominio del problema.

Veamos la siguiente tabla para realizar el etiquetado:

Etiqueta	Intervalo	Símbolo
Fuerte incremento	$[1/\delta, +\infty]$	H
Incremento moderado	$[1/\delta^2, 1/\delta]$	M
Ligero incremento	$[0, 1/\delta^2]$	L
Sin variación	0	0
Ligero decremento	$[-1/\delta^2, 0]$	l
Decremento moderado	$[-1/\delta, -1/\delta^2]$	m
Fuerte decremento	$[-\infty, -1/\delta]$	h

Este alfabeto es el que se utiliza para obtener la cadena de caracteres asociada a la serie temporal $S_x = \langle c_1, \dots, c_{f-1} \rangle$ donde cada c_i representa la evolución de la curva entre dos instantes consecutivos y se obtiene a partir de $X_D = \langle d_0, \dots, d_{f-1} \rangle$ asignándole a cada d_i su carácter.

Esta traducción de la serie temporal a una secuencia de símbolos permite abstraernos de los valores de la curva y centrarnos en su forma. Cada secuencia de símbolos describe toda una familia de curvas con una evolución similar. Evidentemente, si se utiliza otra normalización se debe reconsiderar el modo de asignar etiquetas.

Definición del QSI

Sean las series temporales $X = \{x_0, \dots, x_f\}$ y $Y = \{y_0, \dots, y_f\}$. Normalizando y etiquetando esas series se obtendrán las cadenas S_X, S_Y . Se define la similitud QSI entre las cadenas como:

$$QSI(S_X, S_Y) = \frac{\nabla(LCS(S_X, S_Y))}{m}$$

Donde ∇S es el cuantificador de conteo aplicado a la cadena S , es decir obtiene la longitud de S . Por otro lado, m se define como $m = \max(\nabla S_X, \nabla S_Y)$. De esta manera, se puede interpretar la similitud QSI como el número de símbolos que podemos encontrar en el mismo orden en ambas secuencias dividido por la longitud de la cadena más larga.

La obtención del QSI es un problema NP-Completo, debido a que el cálculo de LCS es NP-Completo. Por ello podemos abordar el cálculo en un problema real mediante dos enfoques:

1. Podemos definir una heurística que nos estime el valor de LCS, con la ventaja de obtener una aproximación razonable pero sin poder aumentar la interpretabilidad del QSI.
2. En el caso de que las series tengan el mismo número de datos, medidos en los mismos periodos de tiempo, podemos definir una cota inferior para el índice QSI, calculable en tiempo polinomial y con una interpretación directa. Esta cota sería:

$$cQSI = \sum_{i=1}^n \delta_i \quad \text{donde} \quad \delta_i = \begin{cases} 1 & \text{si } S_{x_i} = S_{y_i} \\ 0 & \text{en otro caso} \end{cases}$$

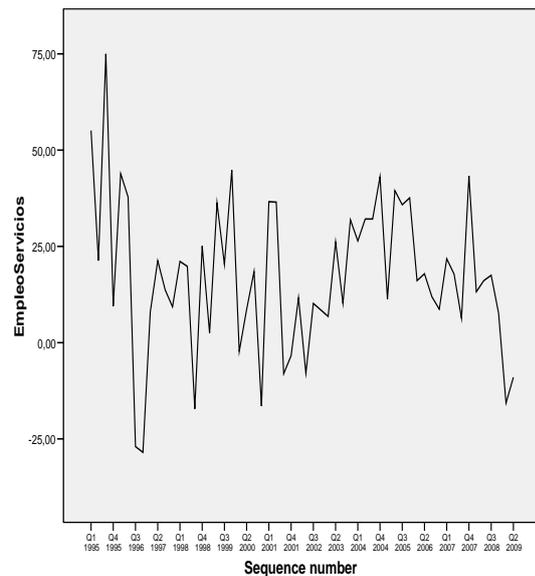
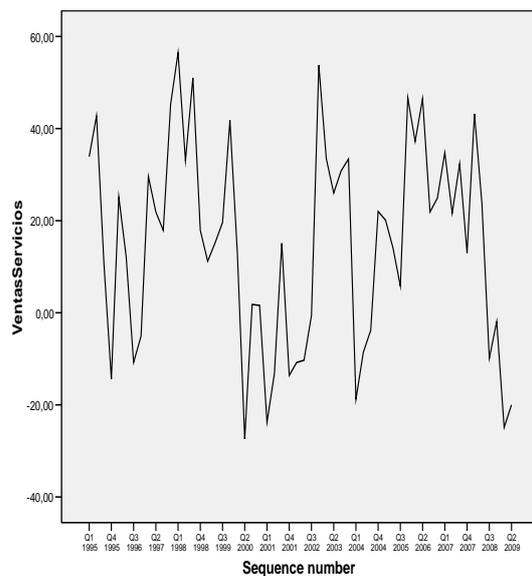
La interpretación de la misma sería el porcentaje de símbolos coincidentes en el total de periodos estudiados.

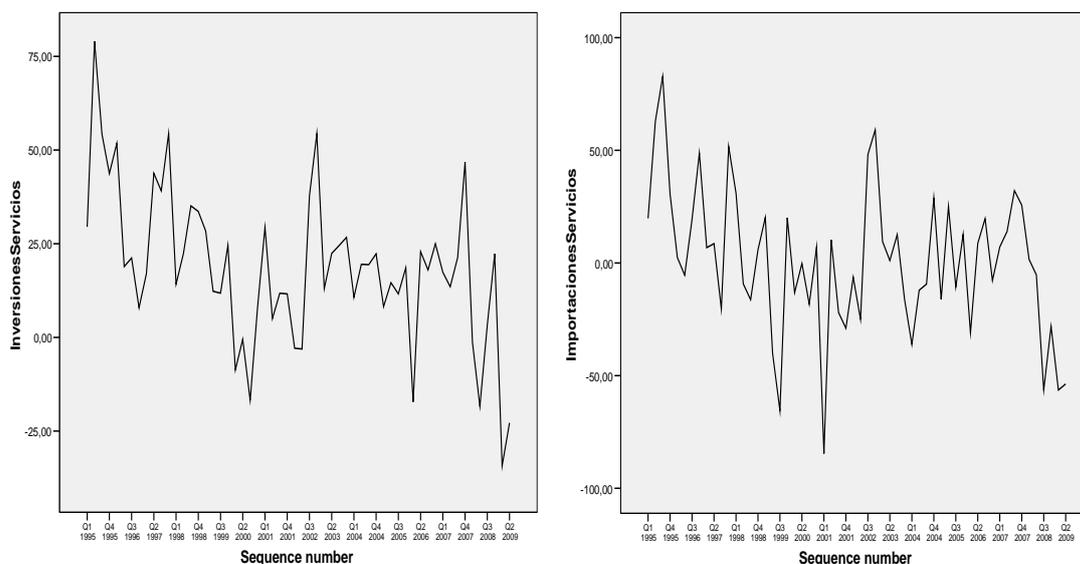
4. ANÁLISIS DESCRIPTIVOS DE LOS DATOS.

En esta sección realizaremos un análisis descriptivo de las 20 variables en estudio. Las 20 variables utilizadas en el estudio son el resultado de estudiar las ventas, las inversiones, las importaciones y el empleo de forma global y en los cuatro sectores productivos: Servicios, Industria, Comercio y Construcción..

4.1. Análisis descriptivo para las variables del sector Servicios.

En primer lugar se muestran los gráficos de secuencias para las cuatro variables del sector Servicio: *VentasServicios*, *EmpleoServicios*, *InversionesServicios* e *ImportacionesServicios*.



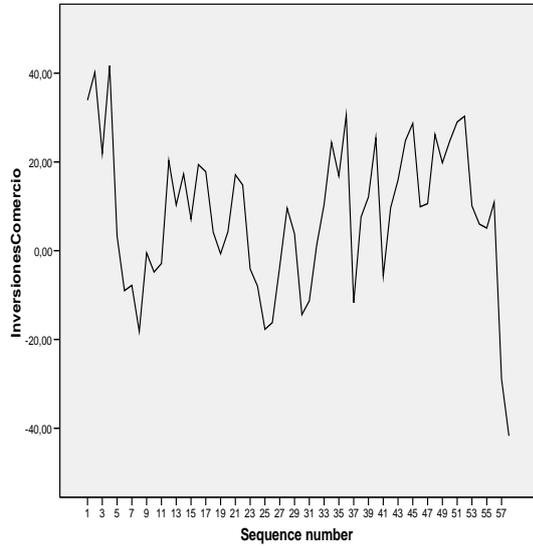
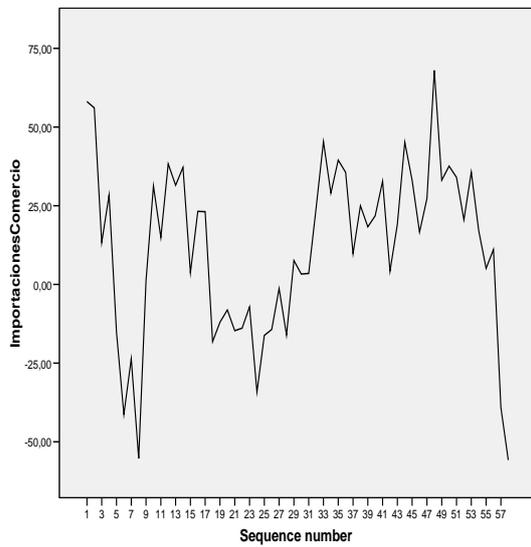
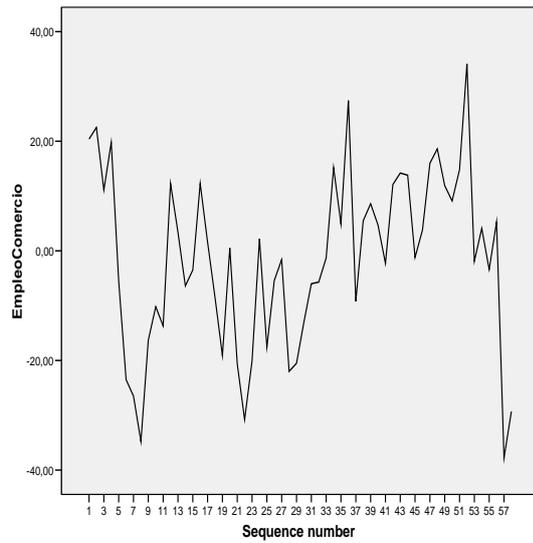
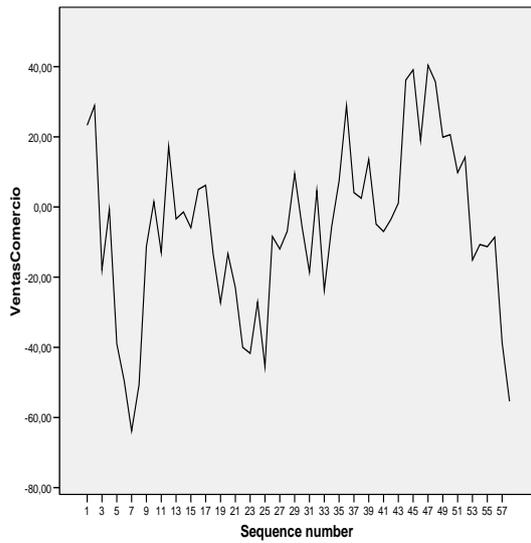


A partir de los gráficos de secuencias se observa que:

- La serie *VentasServicios* tiene un comportamiento estacionario en media cuyo valor es de 15.25, con un rango de variación de 83.90, que es relativamente pequeño, el valor mínimo es de -27.40 y el máximo de 56.50 .La variabilidad es aproximadamente constante en el tiempo con un valor de 492.135.
- La serie *EmpleoServicios* tiene un comportamiento estacionario en media cuyo valor es de 17.00. Tiene un rango de variación igual a 103.50, cuyo valor mínimo es de -28.50 y el máximo de 75.00. La variabilidad es aproximadamente constante en el tiempo con un valor de 399.489. No se observa comportamiento estacional.
- La serie *InversionesServicios* también presenta un comportamiento estacionario en la media cuyo valor es de 18.75. Los datos varían en un rango de 113.30, cuyo valor mínimo es -34.30 y máximo 79.00. Observamos también que la variabilidad de los datos es aproximadamente constante en el tiempo con un valor de 430.240.
- La serie *ImportacionesServicios* presenta un comportamiento estacionario en la media cuyo valor es de 0.6776. Los datos varían en un rango de 167.40, cuyo valor mínimo es -84.60 y máximo 82.80. Observamos también que la variabilidad de los datos es aproximadamente constante en el tiempo con un valor de 430.240.

4.2. Análisis descriptivo para las variables del sector Comercio.

Tenemos gráficos de secuencias para las cuatro variables del sector Comercio: *VentasComercio*, *EmpleoComercio*, *InversionesComercio* e *ImportacionesComercio*.

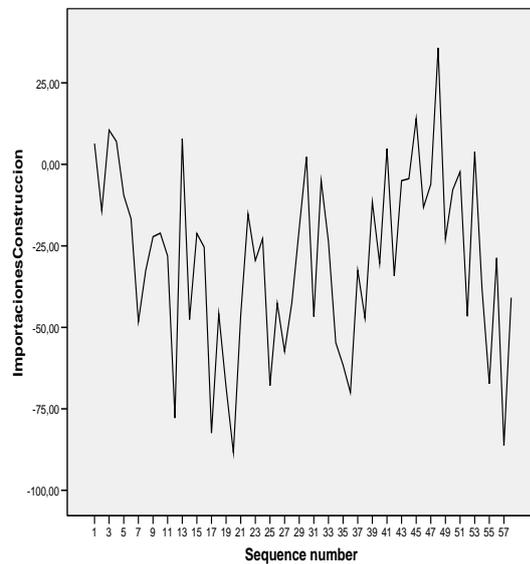
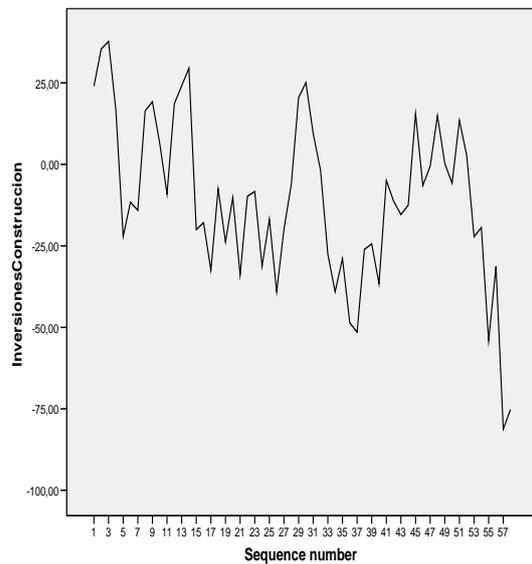
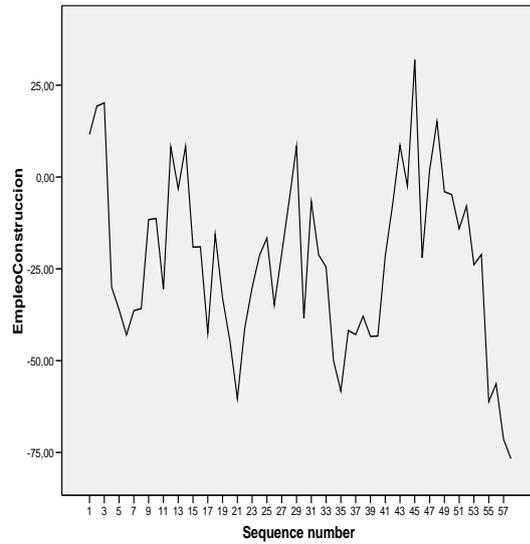
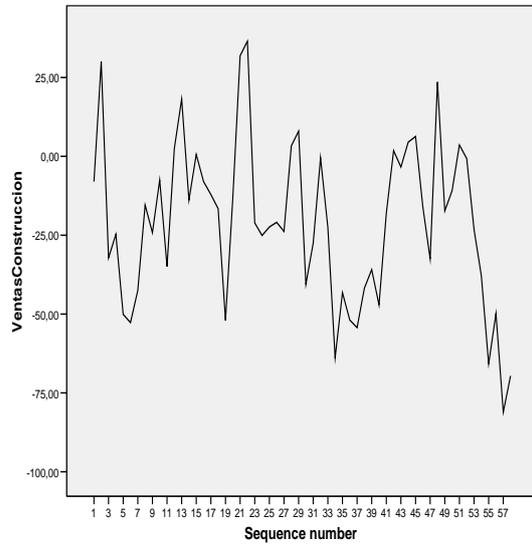


De los gráficos anteriores podemos ver que:

- Para la serie *VentasComercio* se observa un comportamiento estacionario alrededor de la media de valor -5.77. La variabilidad de los datos es constante a lo largo del tiempo con un valor de 592.128. Los datos varían en un rango de 104.30, toman el valor mínimo de -63.90 y un valor máximo de 40.40.
- La serie *EmpleoComercio* también tiene un comportamiento estacionario alrededor de la media de valor -1.508. Los datos tienen una variabilidad constante a lo largo del tiempo de valor 259.033. La variación de los datos está en un rango de 71.90, cuyo valor mínimo es de -37.80 y valor máximo de 40.40.
- La serie *InversionesComercio* tiene un comportamiento estacionario alrededor de la media de valor 8.10. La variabilidad de los datos es constante a lo largo del tiempo con un valor de 193.78. Los datos varían en un rango de 83.40, toman el valor mínimo de -41.70 y un valor máximo de 41.70.
- La serie *ImportacionesComercio* también tiene un comportamiento estacionario alrededor de la media de valor 11.67. Los datos tienen una variabilidad constante a lo largo del tiempo de valor 372.35. La variación de los datos está en un rango de 96.38, cuyo valor mínimo es de -55.80 y valor máximo de 68.00.

4.3. Análisis descriptivo para las variables del sector Construcción.

A continuación presentamos los gráficos de secuencias para las cuatro variables del sector Comercio: *VentasConstrucción*, *EmpleoConstrucción*, *InversionesConstrucción* *ImportacionesConstrucción*.



De los gráficos anteriores podemos concluir lo siguiente:

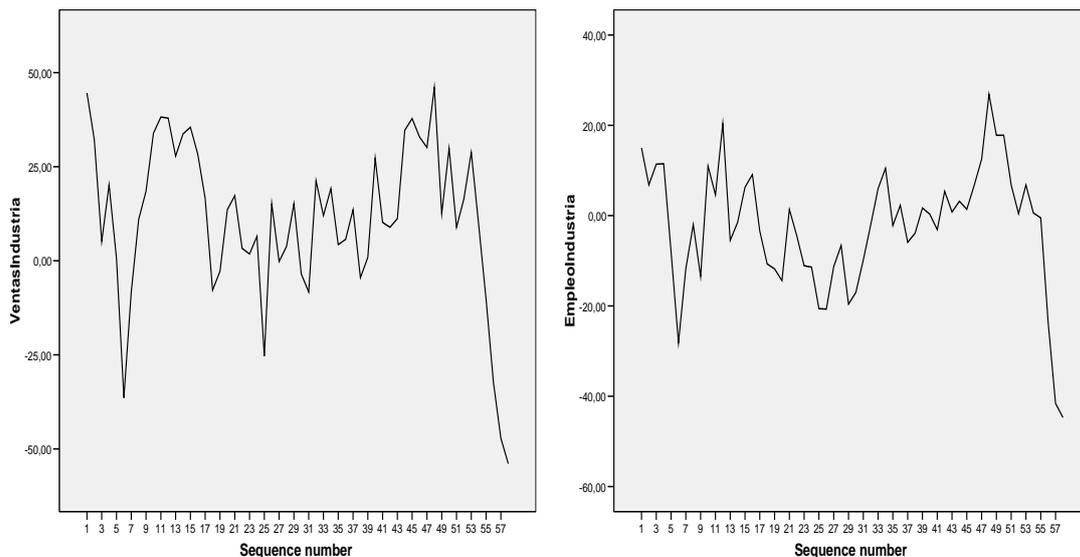
- La serie *VentasConstrucción* presenta un comportamiento estacionario alrededor de la media de valor -20.78. Los datos presentan una variabilidad constante y de valor 671.79. Los datos varían en un rango de 117.60 cuyo valor mínimo es de -81.10 y máximo de 36.50.
- La serie *EmpleoConstrucción* también presenta un comportamiento estacionario alrededor de la media de valor -22.64. Los datos poseen una variabilidad

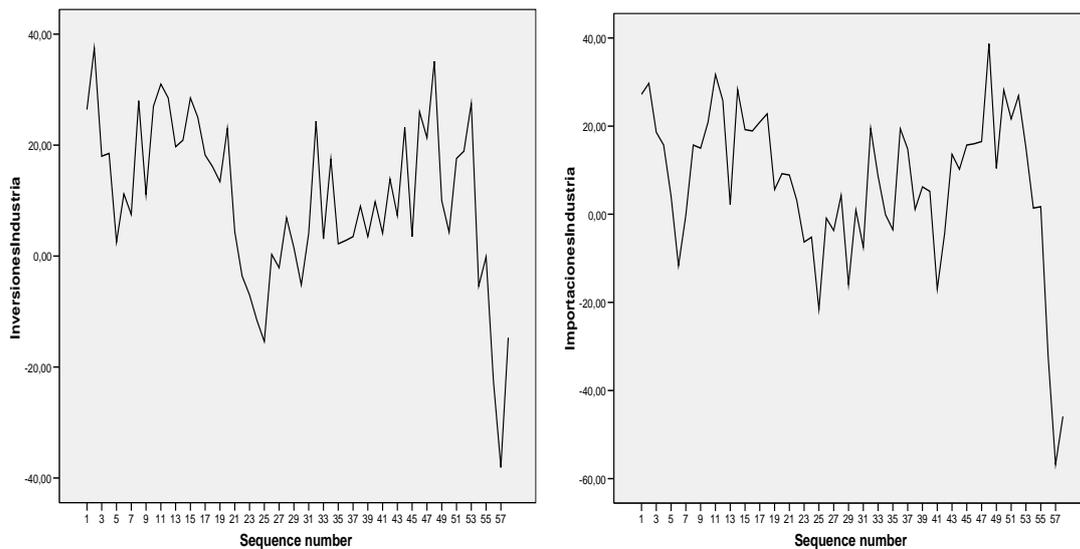
constante igual a 571.668. Los datos varían en un rango de 108.60, tomando el valor mínimo en -76.70 y el valor máximo en 31.90.

- La serie *InversionesConstrucción*, claramente la serie tiene tendencia, por lo que es necesario realizar una diferencia regular para convertirla a estacionaria. La media tiene un valor -10.84. La variabilidad es constante y tiene un valor de 655.09. Los datos varían en un rango de 118.90.
- La serie *ImportacionesConstrucción* presenta un comportamiento estacionario alrededor de la media de valor -29.10. Los datos presentan una variabilidad constante y de valor 766.79. Los datos varían en un rango de 124.10 cuyo valor mínimo es de -88.40 y máximo de 35.70.

4.4. Análisis descriptivo para las variables del sector Industria.

Los gráficos de secuencias para las cuatro variables del sector Industria, *VentasIndustria*, *EmpleoIndustria*, *InversionesIndustria* e *ImportacionesIndustria* son:





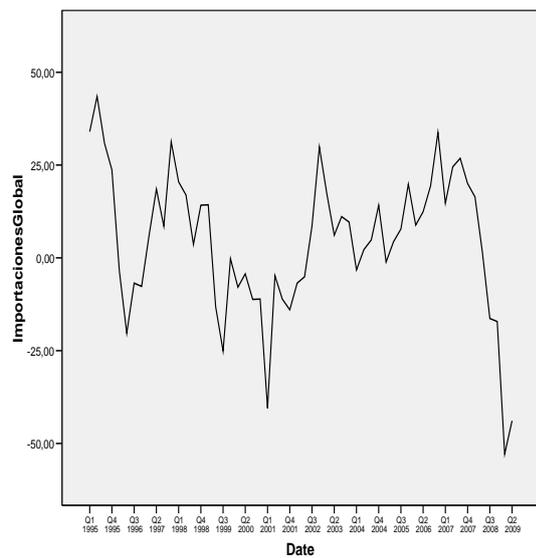
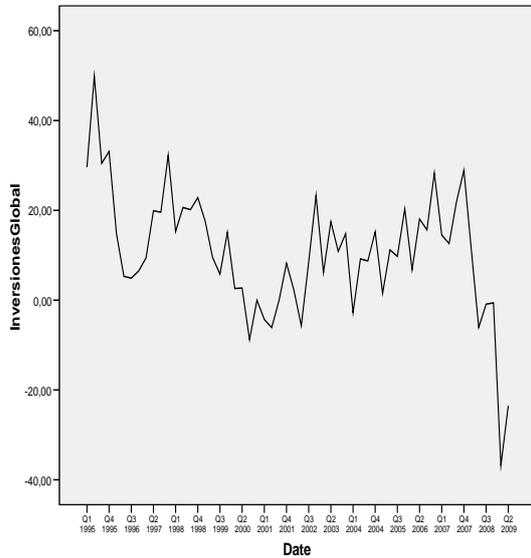
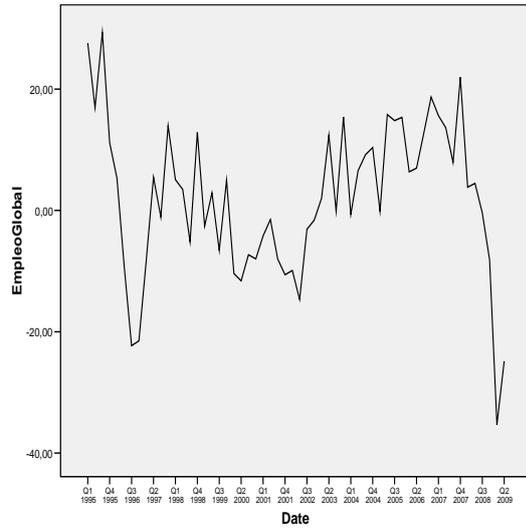
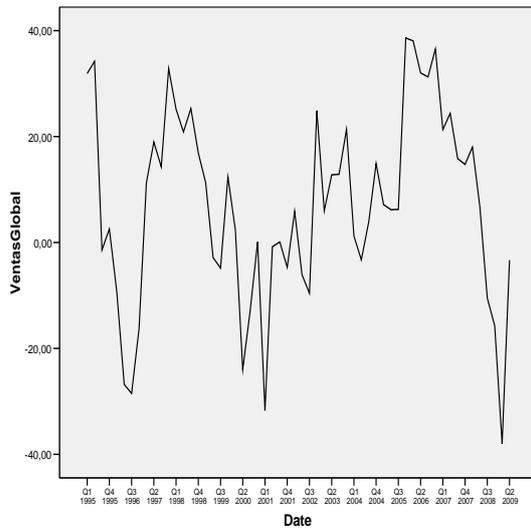
De los gráficos anteriores podemos ver que:

- Para la serie *VentasIndustria* posee un comportamiento estacionario alrededor de la media de valor 11.07. La variabilidad de los datos es constante a lo largo del tiempo con un valor de 455.05. Los datos varían en un rango de 117.60, toman el valor mínimo de -53.90 y un valor máximo de 46.10.
- La serie *EmpleoIndustria* presenta un comportamiento estacionario alrededor de la media de valor -2.50. Los datos tienen una variabilidad constante a lo largo del tiempo de valor 189.610. La variación de los datos está en un rango de 71.60, cuyo valor mínimo es de -44.70 y valor máximo de 26.90.
- La serie *InversionesIndustria* tiene un comportamiento estacionario alrededor de la media de valor 10.20. La variabilidad de los datos es constante a lo largo del tiempo con un valor de 219.058. Los datos varían en un rango de 75.60, toman el valor mínimo de -38.10 y un valor máximo de 37.50.
- La serie *ImportacionesImporataciones* también tiene un comportamiento estacionario alrededor de la media de valor 7.01. Los datos tienen una variabilidad constante a lo largo del tiempo de valor 322.11. La variación de

los datos está en un rango de 95.50, cuyo valor mínimo es de -56.80 y valor máximo de 38.70.

4.5. Análisis descriptivo para las variables Globales.

Tenemos los gráficos de secuencias para las cuatro variables globales: Ventas, Empleo, Inversiones e importaciones.



De los gráficos anteriores podemos ver que:

- Para la serie *Ventas* posee un comportamiento estacionario alrededor de la media de valor 7.08. La variabilidad de los datos es constante a lo largo del tiempo con un valor de 333.09. Los datos varían en un rango de 76.64, toman el valor mínimo de -38.01 y un valor máximo de 38.63.
- La serie *Empleo* presenta un comportamiento estacionario alrededor de la media de valor 1.99. Los datos tienen una variabilidad constante a lo largo del tiempo de valor 162.31. La variación de los datos está en un rango de 64.68, cuyo valor mínimo es de -35.31 y valor máximo de 29.37.
- La serie *Inversiones* tiene un comportamiento estacionario alrededor de la media de valor 10.65. La variabilidad de los datos es constante a lo largo del tiempo con un valor de 193.78. Los datos varían en un rango de 86.73, toman el valor mínimo de -36.83 y un valor máximo de 49.90.
- La serie *Importaciones* también tiene un comportamiento estacionario alrededor de la media de valor 4.3. Los datos tienen una variabilidad constante a lo largo del tiempo de valor 372.35. La variación de los datos está en un rango de 96.38, cuyo valor mínimo es de -52.88 y valor máximo de 43.50.

Después de análisis descriptivo podemos observar que todas las variables tienen casi el mismo comportamiento, excepto las del sector construcción. Esta circunstancia se puede traducir en que el sector de la construcción ha sido el más afectado durante el periodo en estudio. Además, se puede conjeturar que ninguna de las series en estudio requieren transformación, si se desea ajustar un modelo de series temporales ARIMA o de regresión dinámica.

5. MODELADO DE LOS DATOS GLOBALES.

En esta sección ajustaremos a los datos globales un modelo mediante modelos ARIMA y por regresión dinámica. Al final se realizará una comparación entre ambos modelos.

5.1. Modelado mediante modelos ARIMA.

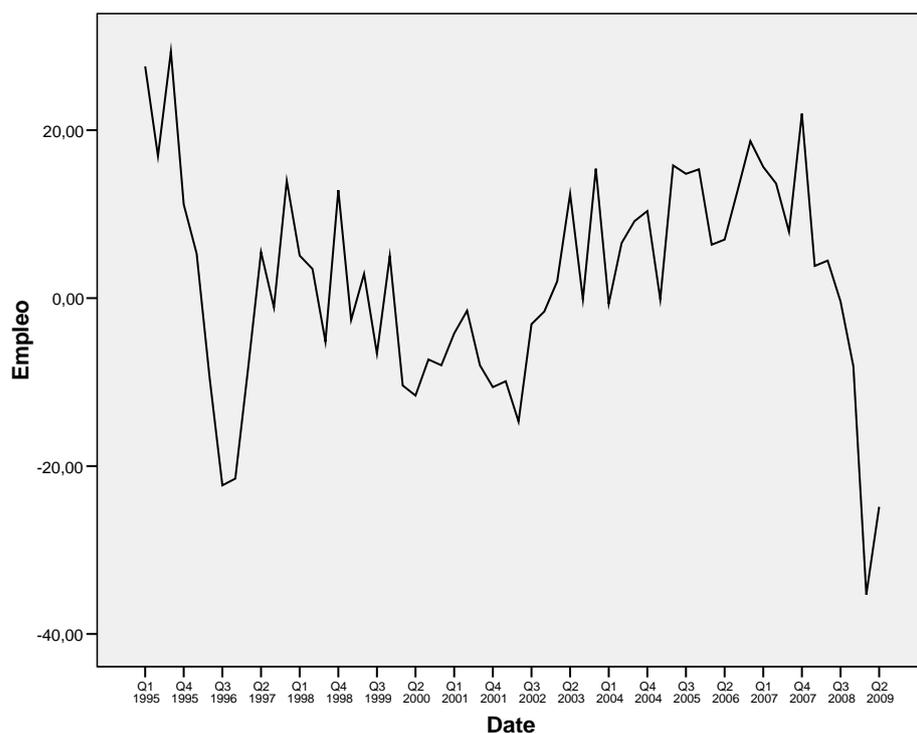
Iniciaremos modelando con modelos ARIMA cada una de las series de datos globales. *Empleo, Inversiones, Importaciones y Ventas.*

5.1.1. Modelado de la serie *Empleo*.

Ajustaremos primero el modelo para la variable *Empleo*. Para ello disponemos de los datos desde 1995 hasta el segundo trimestres de 2009 de forma trimestral. Que los datos sean trimestrales y dado que los comportamientos periódicos de producirse son de forma anual, en caso de existir periodicidad, esta será en periodos de cuatro.

La variable *Empleo* esta medida en función de la variación de las mismas con respecto al periodo anterior, es decir, si el valor de ventas es -12, indica que se ha producido una disminución de un 12% con respecto a las ventas del periodo anterior.

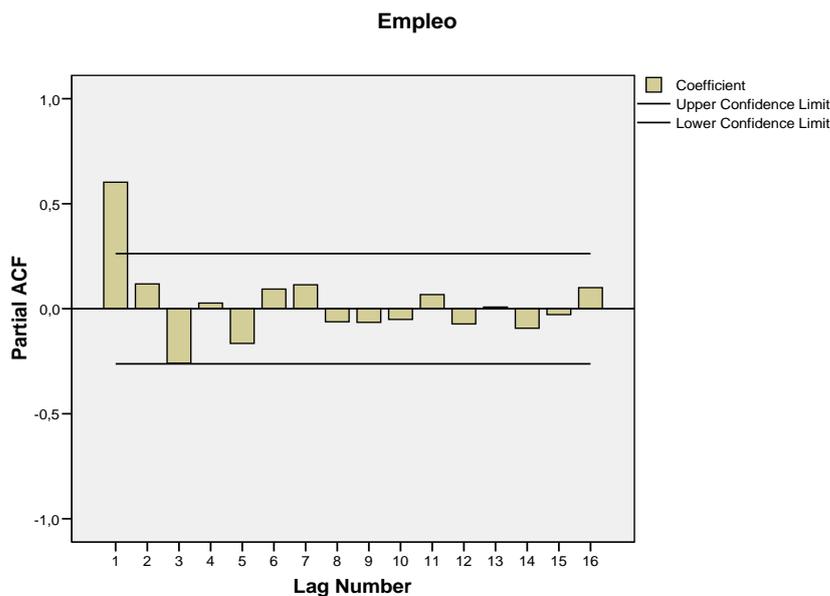
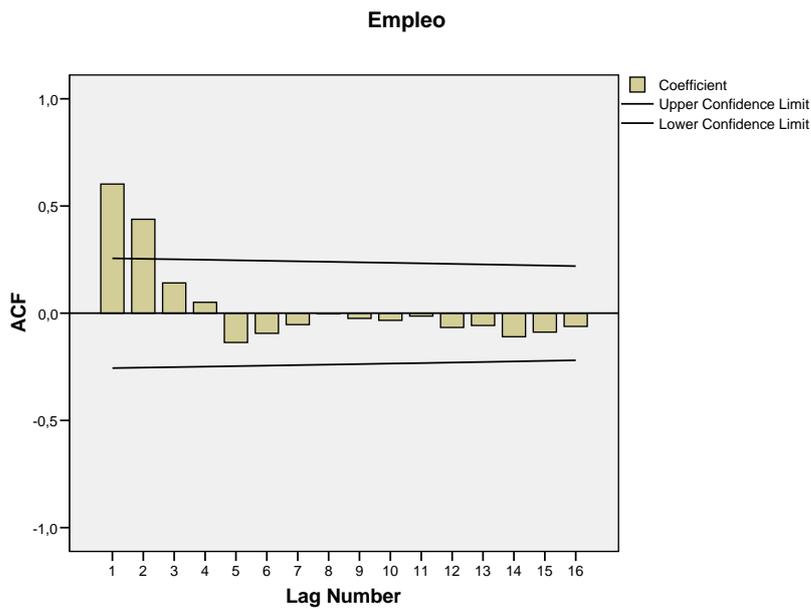
Tenemos el gráfico de secuencia para la serie donde en el eje **X** está representado el año y el trimestre respectivo, y en el eje **Y** el valor que toma la variable *Empleo*



De la grafica podemos observar tres cuestiones:

- La serie no tiene tendencia. Es decir, los valores varían respecto a la media, que parece estar en el entorno del 0. No necesito realizar diferenciación no estacional.
- La serie no tiene tendencia estacional. Es decir, los valores varían indistintamente en el entorno de la media sin depender del trimestre al que se refieran (si tomamos datos cada cuatro no tienen un comportamiento similar). No necesito realizar diferenciación estacional.
- La serie es homocedástica. Es decir, no se observa aumento o disminución de variabilidad a la largo de el tiempo. No necesito realizar ninguna transformación.

Una vez que la serie es estacionaria, podemos pasar a analizar las autocorrelaciones.



De la parte regular podemos presuponer que es un AR(1), ya que en un modelo AR(1) las autocorrelaciones simples presentan una tendencia de decaimiento exponencial y la primera autocorrelación parcial debe ser significativa, lo cual es claro en nuestro grafico.

No parece que existan otros valores significativos de las autocorrelaciones que no queden explicados con el modelo AR(1) en la parte regular. Recordar que las

autocorrelaciones significativas son las que se salen de las bandas de confianza, dado que sobre ellas puedo garantizar que con una probabilidad alta su valor es distinto de 0. De la parte estacional no podemos presuponer modelo alguno dado que no se observa valores significativos en las autocorrelaciones múltiplos de cuatro.

Por todo ello el modelo inicial para la serie Ventas es un $ARIMA(1,0,0)(0,0,0)$. Pasaremos a analizar dicho modelo.

Modelo ARIMA (1, 0,0) (0, 0,0)

En primer lugar estudiaremos el comportamiento global de modelo.

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	0	,392	21,471	17	,206	0

El valor de R-cuadrado obtenido para este modelo pese a no ser alto si parece ser informativo. El test de Ljung-Box no es significativo ya que su valor es mayor que 0.05, lo que indica que no rechazamos la hipótesis de que todas las autocorrelaciones de los residuos son significativamente iguales a 0.

En segundo lugar estudiaremos el comportamiento de los distintos parámetros.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		1,946	4,027	,483	,631
			AR	Lag 1	,692	,103	6,747	,000

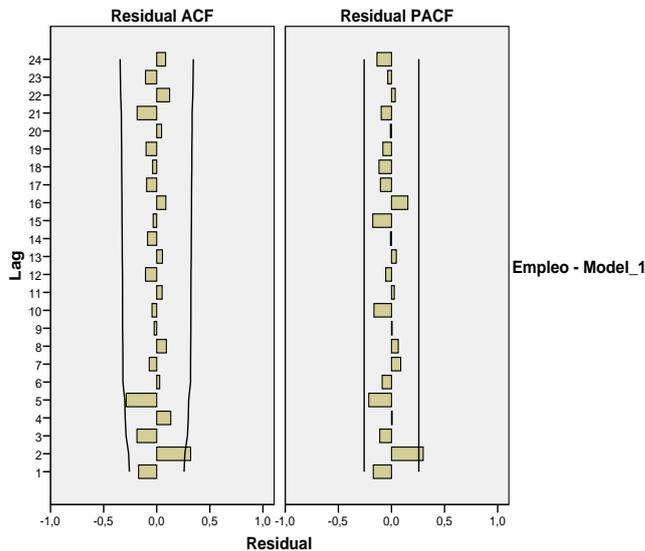
El parámetro de la parte autorregresiva regular es significativa, lo que indica que es distinto de 0 y por lo tanto se debe considerar en el modelo.

Sobre la constante debemos indicar que la conservaremos siempre en los modelos independientemente de su significación. Esto se debe a tres consideraciones:

- No existe forma estadística de demostrar que sea igual a 0 (un valor de Sig superior a 0.05 solo nos indica que no puedo rechazar que sea 0, pero no podemos afirmarlo sin conocer la función de potencia, cuyo cálculo es imposible en este caso).

- La naturaleza de las variables no nos permite decir que deba pasar por el eje de coordenadas.
- No incrementa la dificultad o interpretabilidad del modelo.

Por último estudiaremos el comportamiento de los residuos.



La grafica de las autocorrelaciones nos indica que para todos los retardos menos el retardo 2 no podemos rechazar que sean 0. Para explicar que el retardo 2 sea significativo existen dos posibilidades:

- Pertenece al porcentaje de falsos rechazos, lo que sucederá con una probabilidad α .
- La variable *Empleo* depende del retardo, es decir el empleo de este trimestre dependen depende del empleo de hace 2 trimestres

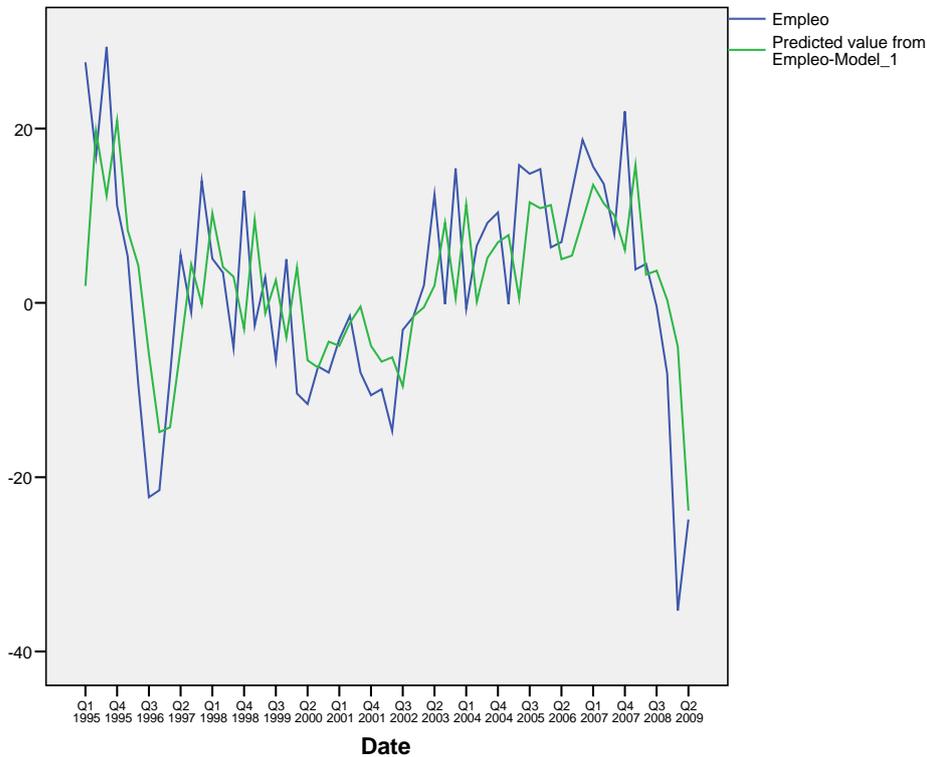
Si planteamos el modelo AR(2), se comprueba que no es significativa el valor del parámetro asociado al segundo retraso, por lo que descartamos esta posibilidad y asumiremos la primera posibilidad.

Por todo ello, podemos concluir que la variable ventas queda ajustada con un modelo AR(1). Concretamente el modelo planteado será:

$$Empleo = 1.946 + 0.692Empleo_{t-1} + a_t$$

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar diferencias notables en algunos casos, esto se debe al valor de R cuadrado. Obsérvese que no identifica los cambios de tendencia. Eso se refleja en que la

serie de las predicciones se encuentra desplazada a la derecha con respecto a la serie real del empleo.



Para modelar las otras tres variables se sigue un proceso similar que el de la variable *Empleo*

5.1.2. Modelado de la serie *Ventas*

El modelo que mejor ajusta a la variable *Ventas* es el ARIMA (1,0,0)(0,0,0)

El comportamiento global de modelo es:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Ventas-Model_1	0	,409	16,773	17	,470	0

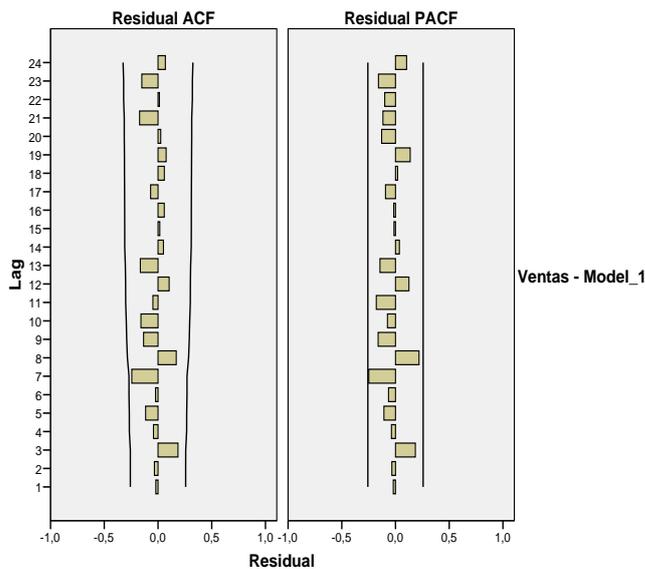
El valor de R-cuadrado para este modelo pese a no ser alto si parece ser informativo. El test de Ljung-Box no es significativo ya que su valor es mayor que 0.05, lo que indica que no rechazamos la hipótesis de que todas las autocorrelaciones de los residuos son significativamente iguales a 0.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Ventas-Model_1	Ventas	No Transformation	Constant	7,518	5,107	1,472	,147
			AR Lag 1	,651	,101	6,463	,000

El parámetro de la parte autorregresiva regular es significativa, lo que indica que es distinto de 0 y por lo tanto se debe considerar en el modelo.

Por último estudiaremos el comportamiento de los residuos.

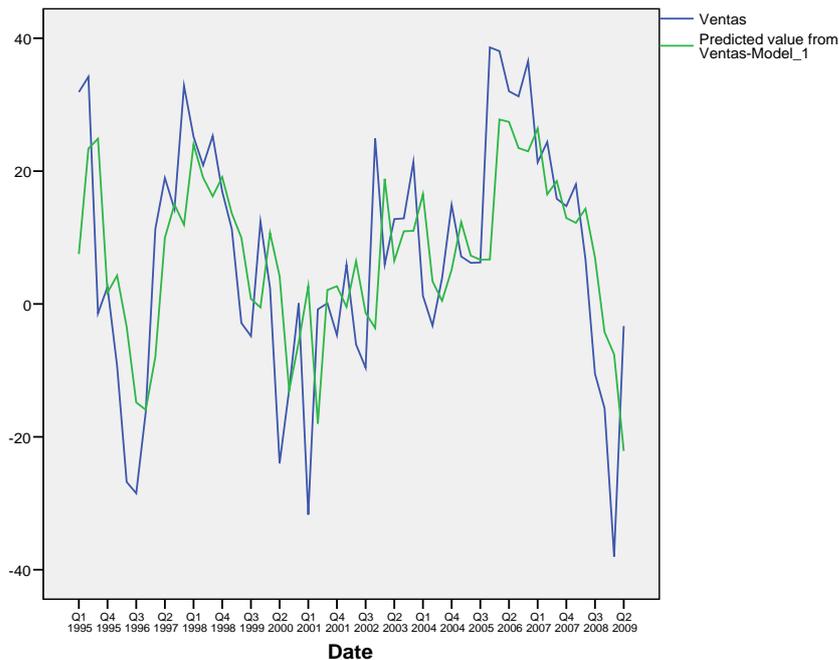


Podemos observar que todas las autocorrelaciones están dentro de los márgenes de confianza (no son significativamente distintos de 0), lo que indica que los residuos se pueden explicar mediante un ruido, como debe suceder si el modelo está bien planteado.

Por todo ello, podemos concluir que la variable ventas queda bien ajustada con un modelo AR(1). Concretamente el modelo planteado será:

$$Ventas_t = 5.107 + 0.101Ventas_{t-1} + a_t$$

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar diferencias notables en algunos casos, esto se debe al valor de R cuadrado. Obsérvese que tampoco identifica los cambios de tendencia.



5.1.3. Modelo para la serie *Inversiones*

El modelo que mejor ajusta a la serie *Inversiones* es el ARIMA(1,0,0)(1,0,1)

El comportamiento global de modelo es el siguiente:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Inversiones-Model_1	0	,537	10,149	15	,810	0

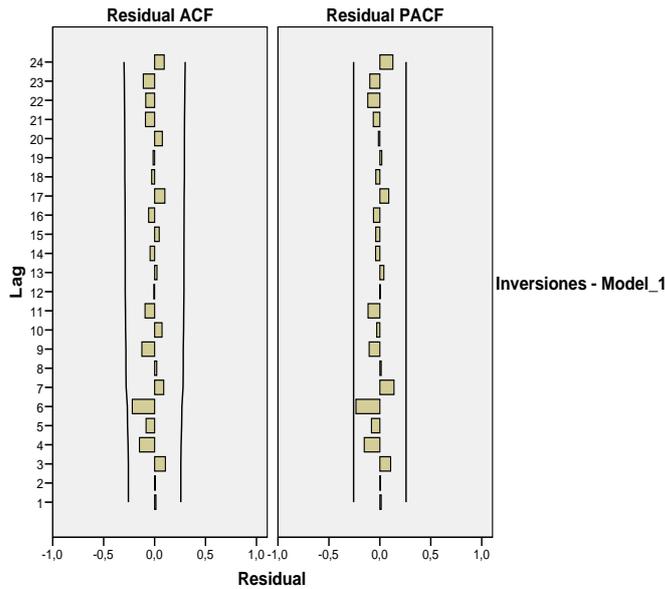
El R-cuadrado toma un valor aceptable, el El test de Ljung-Box no es significativo ya que su valor es mayor que 0.05.

ARIMA Model Parameters

			Estimate	SE	t	Sig.		
Inversiones-Model_1	Inversiones	No Transformation	Constant	10,130	21,751	,466	,643	
			AR	Lag 1	,800	,106	7,542	,000
			AR, Seasonal	Lag 1	,973	,067	14,448	,000
			MA, Seasonal	Lag 1	,798	,234	3,409	,001

Todos los parámetros son significativos (salvo la constante), por lo que deben considerarse en el modelo.

Las graficas para los residuos:



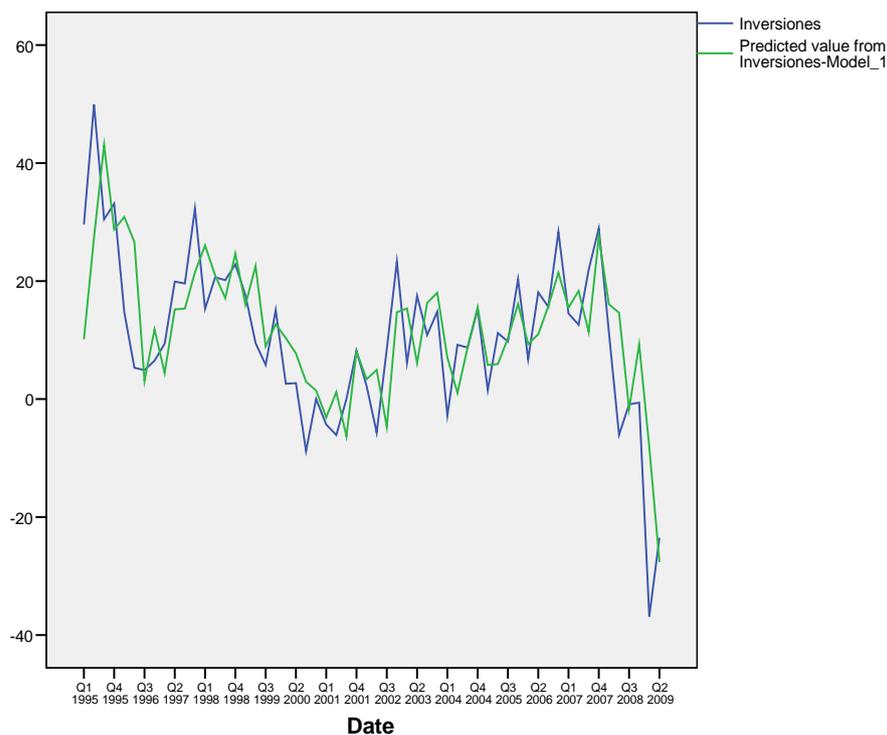
Observamos que todos son significativamente iguales a cero.

Por todo lo anterior concluimos que el modelo es el ARIMA(1,0,0)(1,0,1):

$$Inversiones_t = 10.13 + 0.8Inversiones_{t-1} + 0.973Inversiones_{t-4} + 0.798a_{t-4} + a_t$$

Lo que nos dice que la inversión actual depende de la inversión del trimestre anterior y de la inversión de hace un año, y de perturbaciones de hace un año.

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar diferencias notables en algunos casos, esto se debe al valor de R cuadrado. Obsérvese que identifica mejor los cambios de tendencia.



5.1.4. Modelo para la serie *Importaciones*.

El modelo que mejor ajusta a la serie es el ARIMA(1,0,0)(0,0,0).

El comportamiento global del modelo es el siguiente:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Importaciones-Model_1	0	,487	12,781	17	,751	0

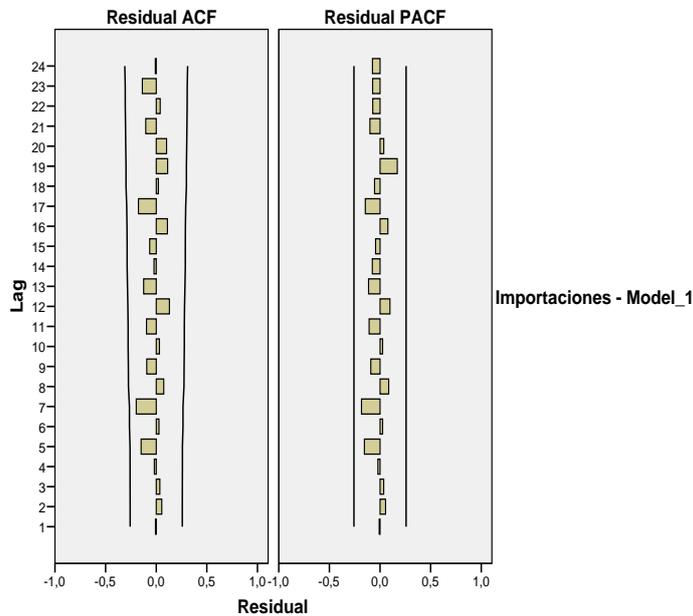
El valor del R_cuadrado nos indica que le modelo es informativo. El valor del test de Ljung-Box es alto, por lo que n debemos rechazar que los autocorrelaciones de los residuos sean todas iguales a cero.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Importaciones-Model_1	Importaciones	No Transformation	Constant	3,431	7,147	,480	,633
			AR Lag 1	,759	,097	7,834	,000

El parámetro de la parte autorregresiva es significativo y diferente de cero. Por lo que debemos considerarlo en el modelo.

La grafica de las autocorrelaciones para los residuos



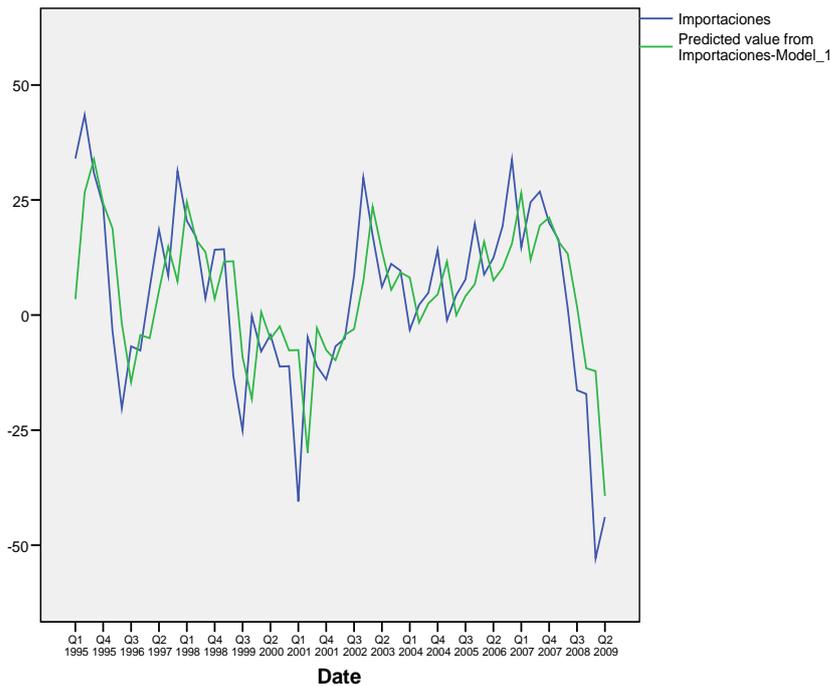
De la grafica observamos que todas las autocorrelaciones son significativamente iguales a cero.

Por lo que el modelo que mejor ajusta a la serie es:

$$Importaciones_t = 3.431 + 0.759Importaciones_{t-1} + a_t$$

El modelo indica que las importaciones actuales dependen de las importaciones del trimestre anterior

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar diferencias notables en algunos casos, esto se debe al valor de R cuadrado. Obsérvese que no identifica los cambios de tendencia.



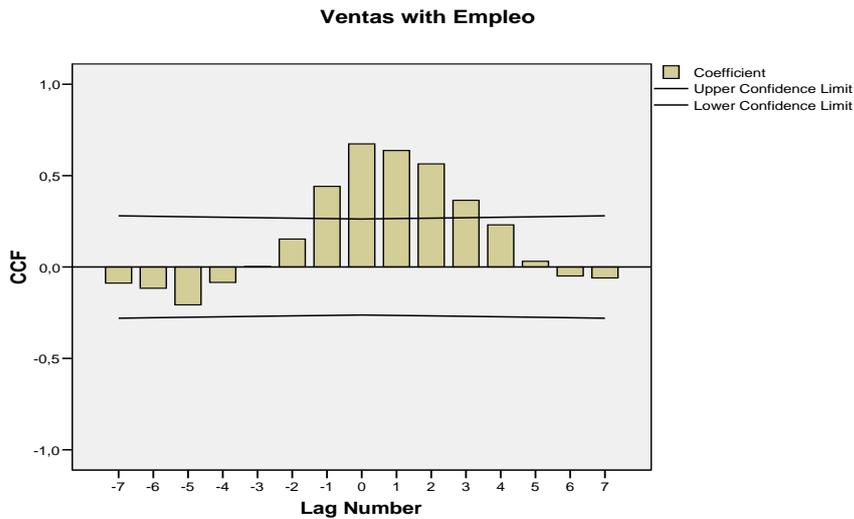
5.2. Modelado mediante regresión dinámica.

Para realizar el análisis de regresión dinámica de las variables globales es necesario verificar que todas tengan relación entre si. Para ello debemos realizar primero un análisis de correlación cruzada entre las variables.

5.2.1. Correlación cruzada

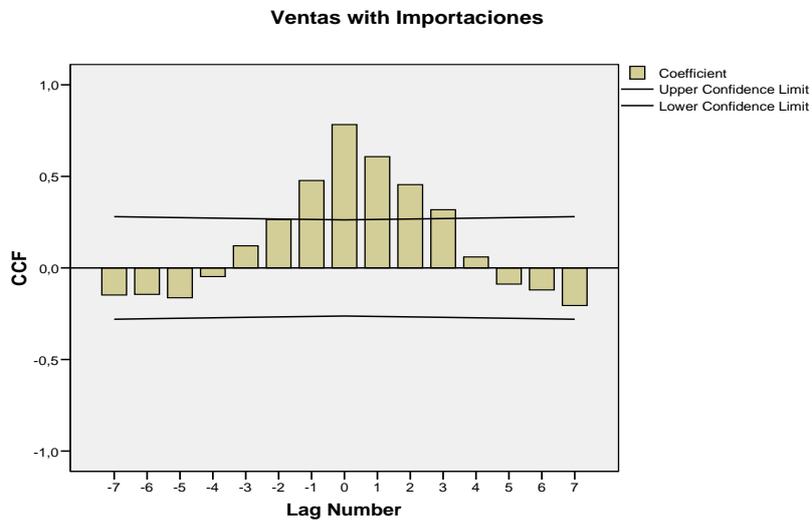
Recordemos que el único requisito para utilizar la correlación cruzada es que las series deben ser estacionarias. Ya que las series en estudio son estacionarias, podemos utilizarla sin ningún problema. Si no hubiese sido así, deberíamos transformar las series a estacionarias y utilizar esas nuevas series transformadas.

Ventas con Empleo:



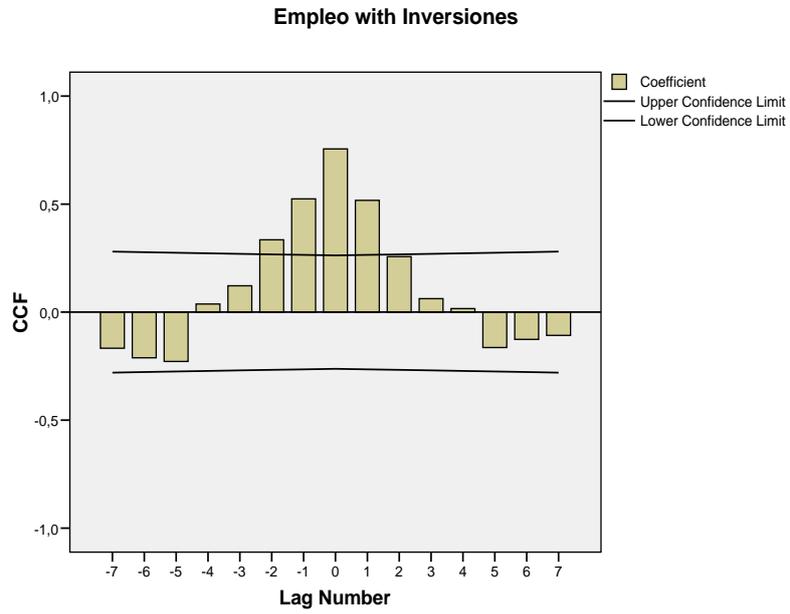
El grafico nos indica que el mayor valor significativo está en el retardo 0. También son significativos los retardos cercanos al retardo 0. Lo que nos indica una clara relación entre *Ventas* y *Empleo*.

Ventas con Importaciones:



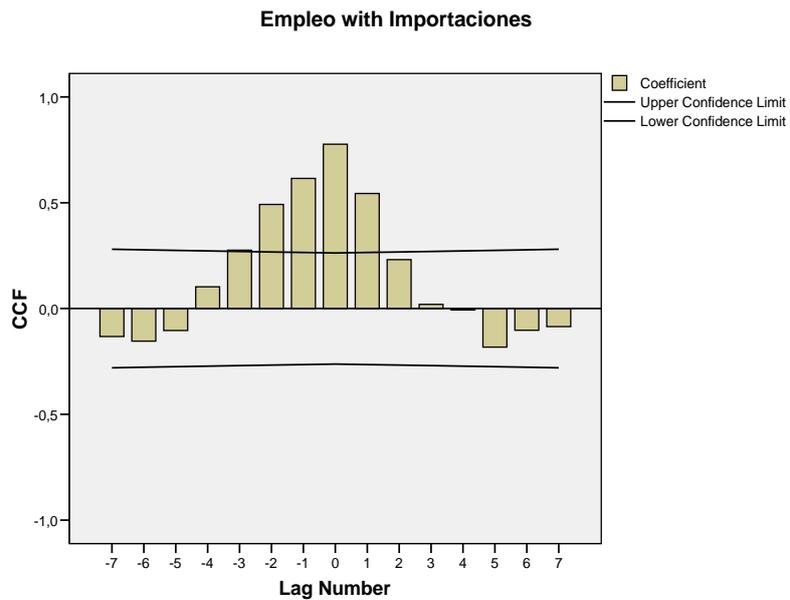
Vemos que ventas con importaciones están muy relacionadas en el retardo 0 y retardos cercanos a 0. Lo que indica que lo que pasa actualmente en *Importaciones* afecta a las *Ventas* cercanas y viceversa.

Empleo con Inversiones:



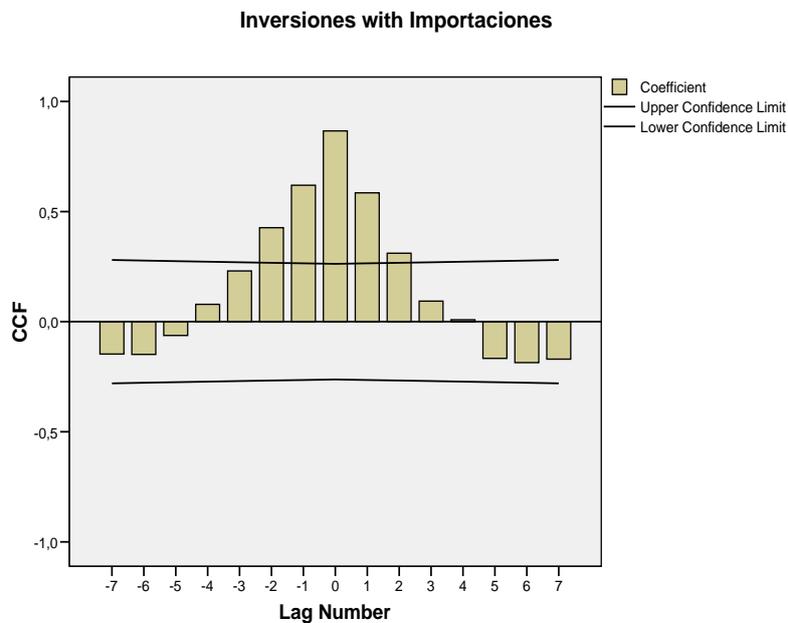
De igual manera el *Empleo* y las *Inversiones* están muy relacionadas en el retardo 0 y retardos cercanos a 0, lo que indica que lo que pasa actualmente en *Inversiones* afecta al *Empleo* cercano y viceversa.

Empleo con Importaciones



Empleo con Importaciones están relacionadas en el retardo 0 y retardo cercanos a 0. Por lo que vemos que lo que pasa actualmente con las *Importaciones* afecta significativamente a lo que está pasando en el entorno actual con el *Empleo* y viceversa.

Inversiones con Importaciones:



Inversiones con Importaciones están relacionadas en el retardo 0 y retardo cercanos a 0. Por lo que vemos que lo que pasa actualmente con las *Importaciones* afecta significativamente a lo que está pasando en el entorno actual con las *Inversiones* y viceversa.

Dado que todas las variables presentan correlaciones cruzadas altas, debemos tener mucho cuidado con la existencia de multicolinealidad en los modelos. Ahora procederemos al modelado mediante regresión dinámica para cada variable.

5.2.2. Modelo para la serie *Empleo*.

MÉTODO FORWARD

Vamos iniciar ajustado el modelo de regresión dinámica con el método forward, de menos a más, es decir, partiremos del modelo mínimo y sobre el debemos ir añadiendo los factores significativos hasta llegar al modelo correcto.

De la sección anterior conocemos que las cuatro variables están correladas, es decir, todas ellas aportan información sobre las otras.

Por lo tanto el modelo mínimo inicial que utilizaremos será el resultante de considerar para la variable dependiente (*Empleo*) un modelo constante y para las otras tres variables el retardo actual.

Modelo 0

En primer lugar estudiaremos el comportamiento global de modelo.

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	3	,643	18,926	18	,396	0

El valor de R-cuadrado obtenido para este modelo es alto y por lo tanto es informativo. El test de Ljung-Box no es significativo ya que su valor es mayor que 0.05, lo que indica que no rechazamos la hipótesis de que todas las autocorrelaciones de los residuos son significativamente iguales a 0.

En segundo lugar estudiaremos el comportamiento de los distintos parámetros.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant	-3,194	1,633	-1,956	,056
	Ventas	No Transformation	Numerator Lag 0	,126	,091	1,384	,172
	Inversiones	No Transformation	Numerator Lag 0	,312	,149	2,092	,041
	Importaciones	No Transformation	Numerator Lag 0	,224	,129	1,740	,087

A la vista de los parámetros vemos que la variable ventas no es significativa para predecir el empleo, lo que implica que este no es el modelo más simple, ya que al modelo más simple no le debe de sobrar nada. Debemos eliminar la variable Ventas del modelo para obtener el más simple.

Modelo 1

Comportamiento global de modelo.

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	2	,630	22,482	18	,211	0

Nuevamente el valor de R-cuadrado obtenido para este modelo es alto y por lo tanto es informativo. El test de Ljung-Box no es significativo ya que su valor es mayor que 0.05.

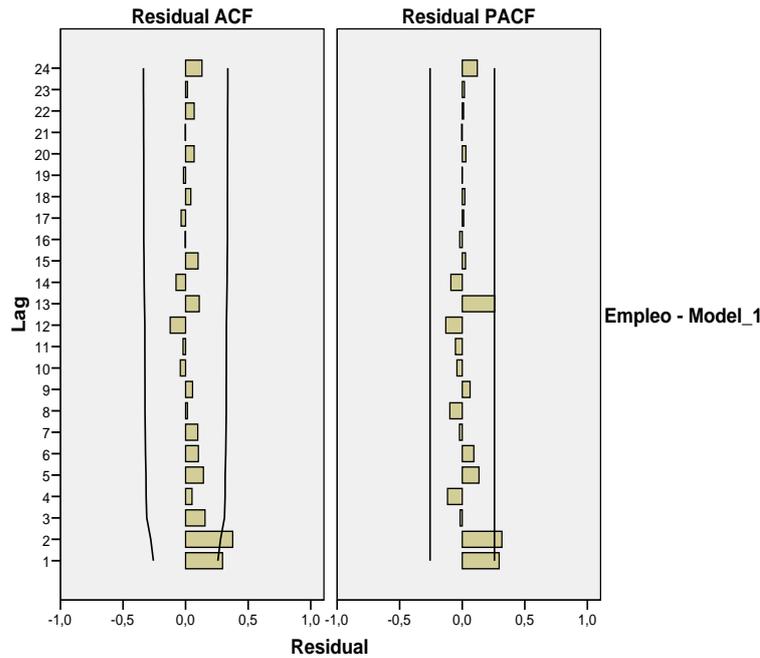
Comportamiento de los distintos parámetros:

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-2,641	1,596	-1,655	,104
	Inversiones	No Transformation	Numerator	Lag 0	,304	,150	2,022	,048
	Importaciones	No Transformation	Numerator	Lag 0	,323	,108	2,982	,004

Ahora podemos observar que todos los parámetros son significativos, ya no le sobra nada a nuestro modelo. Ya que tenemos el modelo más simple debemos ver que es lo que le hace falta, para ellos analizaremos las funciones de autocorrelación simple y parcial de los residuos.

Graficas de las autocorrelaciones residuales:



De la autocorrelación simple notamos que en los dos primeros retardos son significativamente diferentes de cero lo que nos indica que debemos considerar los modelos MA(1) y MA(2).

De las autocorrelaciones parciales también las primeras dos son significativamente diferentes de cero, lo que nos indica que debemos considerar también los modelos AR(1) y AR(2).

Ambas correlaciones simple y parcial nos indican que no hay nada que ajustar en la parte estacionaria.

De aquí se desprenden varios modelos, pero usando el principio de simplicidad debemos empezar a probar con las más sencillos, que son el AR(1) y el MA(1).

Modelo 2: AR(1)

Comportamiento global del modelo.

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	2	,665	19,360	17	,308	0

Vemos que el valor de R-cuadrado es alto. También el valor para del test de Ljung-Box es mayor que 0.05, lo que nos indica que no rechazamos que las autocorrelaciones de los residuos son iguales a 0.

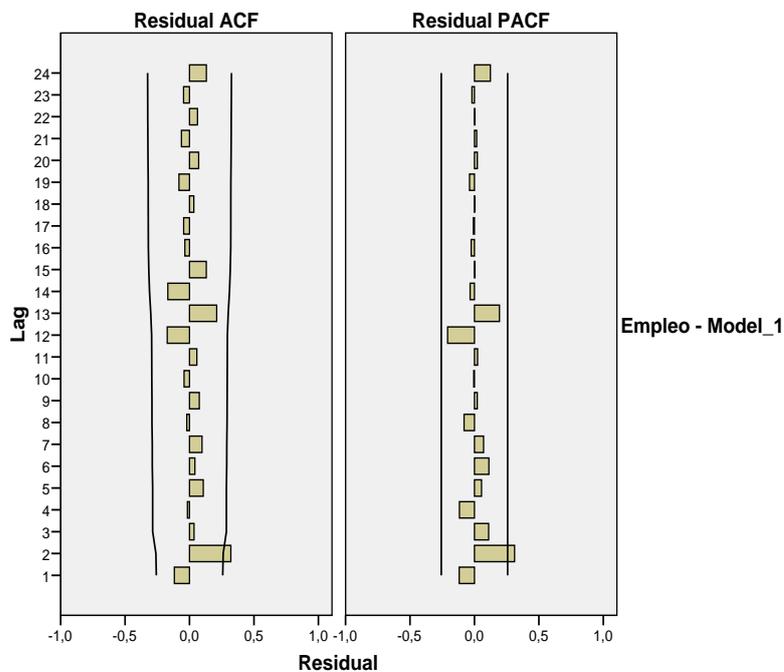
Comportamiento de los distintos parámetros

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-2,782	1,895	-1,468	,148
			AR	Lag 1	,329	,129	2,549	,014
	Inversiones	No Transformation	Numerator	Lag 0	,352	,146	2,412	,019
	Importaciones	No Transformation	Numerator	Lag 0	,254	,109	2,332	,023

Con este modelo vemos que todos los parámetros son significativos. Este modelo es un candidato para nuestro modelo final. Con este modelo estamos diciendo que el empleo depende del empleo del trimestre anterior, de las inversiones e Importaciones actuales.

Graficas de las autocorrelaciones residuales:



En las graficas de autocorrelación simple y parcial vemos que el modelo ajusta bien, aunque dado que la segunda autocorrelación es distinta de cero, deberíamos estudiar la posibilidad de considerar un modelo AR(2). Ahora analicemos el otro modelo simple el MA(1).

Modelo 3: MA(1)

Comportamiento global del modelo

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	2	,651	18,937	17	,332	0

Con los valores de R-cuadrado y del test de Ljung- Box el modelo está bien Comportamiento de los diferentes parámetros.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-2,698	1,719	-1,570	,122
			MA	Lag 1	-,192	,135	-1,422	,161
	Inversiones	No Transformation	Numerator	Lag 0	,325	,150	2,159	,035
	Importaciones	No Transformation	Numerator	Lag 0	,294	,110	2,661	,010

Observamos que el parámetro referido a la parte móvil MA(1) no es significativo. Por lo cual considerar modelos con parte media móvil no parece apropiado. Por último debemos ver si el modelo AR(2) es mejor que el modelo 2.

Modelo 4. AR(2)

Comportamiento global del modelo

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	2	,702	7,441	16	,964	0

El R-cuadro de este modelo es superior al del modelo 1, aunque por poco (0.665 frente a 0.702) . También el test de Ljung Box toma un valor no significativo en este modelo.

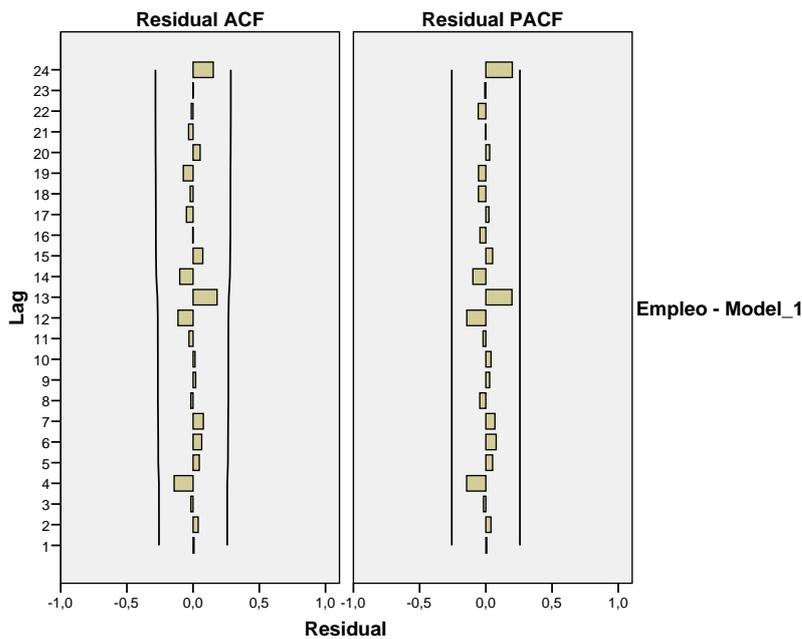
Comportamiento de los diferentes parámetros.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-3,425	2,379	-1,440	,156
			AR	Lag 1	,205	,126	1,629	,109
				Lag 2	,366	,129	2,839	,006
	Inversiones	No Transformation	Numerator	Lag 0	,428	,132	3,232	,002
	Importaciones	No Transformation	Numerator	Lag 0	,193	,097	1,995	,051

De la tabla observamos que no todos los parámetros de este modelo son significativos. Concretamente el parámetro relativo al retardo 1 de la variable Empleo no es significativo y que por lo tanto debemos eliminarlo. Pero si lo eliminamos estamos asegurando que el empleo del trimestre actual depende del empleo de hace dos trimestres, que en la práctica no parece razonable.

Graficas de las autocorrelaciones residuales:



La graficas de las autocorrelaciones nos indica que todas son significativamente iguales a cero.

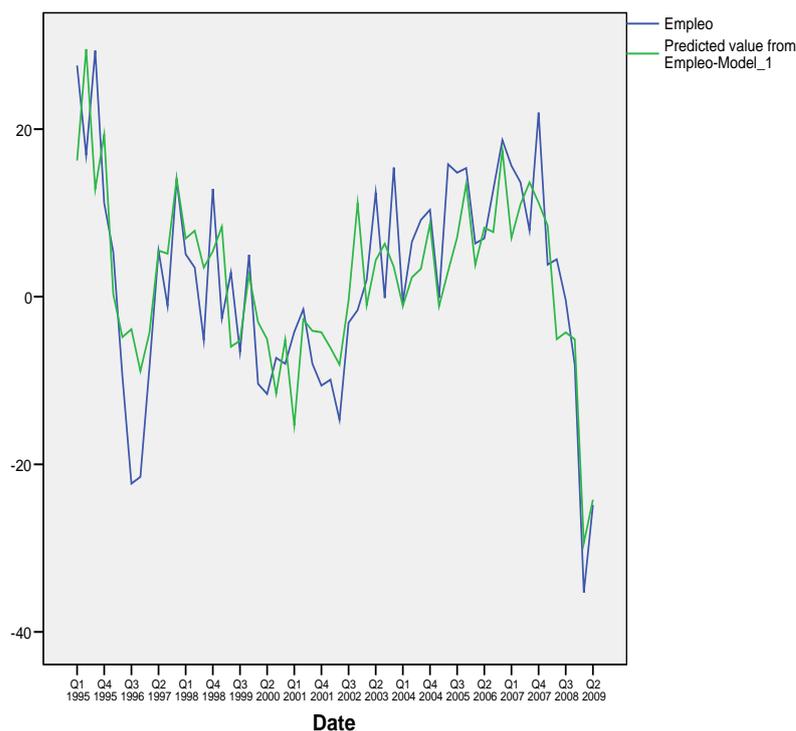
En conclusión, dado que el modelo con parte autoregresiva de orden dos no presenta grandes ventajas en cuanto a la bondad del ajuste y presenta serios inconvenientes en la interpretación y selección de los parámetros a incluir. Podemos afirmar que el modelo 2 es mejor que el modelo 4.

Modelo Final con el método Forward.

Por lo tanto nuestro modelo ajustado por el método Forward es el modelo 2, es decir:

$$Empleo_t = 0.329Empleo_{t-1} + 0.352Inversiones_t + 0.254Importaciones - 2.782 + a_t .$$

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar un ajuste bastante bueno de las predicciones.



METODO BACKWARD

Ahora ajustaremos el modelo de regresión dinámica de más a menos, es decir, partiremos del modelo máximo y sobre el debemos ir eliminando los factores no significativos hasta llegar al modelo correcto.

De la sección anterior sabemos que:

- La variable *Empleo* es un modelo $ARIMA(1,0,0)(0,0,0)$, lo que quiere decir que empleo depende solo de un retardo en la parte autorregresiva.
- La variable *Inversiones* es un modelo $ARIMA(1,0,0)(1,0,1)$, lo que quiere decir, que inversiones depende solo de un retardo en la parte autorregresiva no estacional, y del retardo 1 en la parte estacional.
- La variable *Importaciones* es un modelo $ARIMA(1,0,0)(0,0,0)$, lo que quiere decir, que importaciones depende solo de un retardo en la parte autorregresiva.
- La variable *Ventas* es un modelo $ARIMA(1,0,0)(0,0,0)$, lo que quiere decir, que ventas depende solo de un retardo en la parte autorregresiva.

Por lo tanto el modelo máximo que utilizaremos será el resultante de considerar para la variable dependiente (*Empleo*) un modelo $ARIMA(1,0,0)(0,0,0)$, para las otras tres variables un retardo en la parte regular y además para la variable *Inversiones* un retardo en la parte estacional. Hemos supuesto que la variable empleo no puede depender en más retardos que lo que dependían originalmente las otras series.

Modelo 0

En primer lugar estudiaremos el comportamiento global de modelo.

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	3	,724	18,453	17	,361	0

El valor de R-cuadrado obtenido para este modelo es alto y por lo tanto es informativo. El test de Ljung-Box no es significativo ya que su valor es mayor que 0.05, lo que indica que no rechazamos la hipótesis de que todas las autocorrelaciones de los residuos son significativamente iguales a 0.

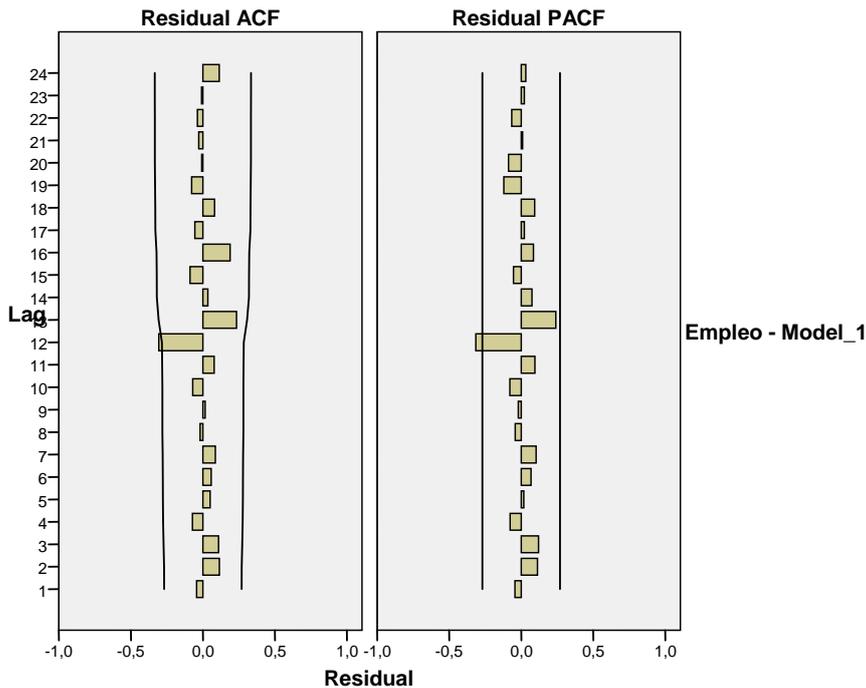
En segundo lugar estudiaremos el comportamiento de los distintos parámetros.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-2,834	2,518	-1,125	,267
			AR	Lag 1	,338	,147	2,292	,027
	Ventas	No Transformation	Numerator		,116	,104	1,110	,273
				Lag 1	-,148	,110	-1,347	,185
	Inversiones	No Transformation	Numerator		,256	,141	1,820	,076
				Lag 1	,189	,103	1,833	,074
	Importaciones	No Transformation	Numerator, Seasonal		-,674	,613	-1,100	,277
				Lag 0	,145	,120	1,203	,235
				Lag 1	-,081	,126	-,644	,523

En la tabla observamos que varios parámetros no son significativos, el menos significativo es el parámetro asociado al primer retardo de la variable *Importaciones*, que nos indica que las importaciones del trimestre anterior no afecta al empleo en el trimestre actual. Procederemos a eliminarlo del modelo.

En tercer lugar estudiaremos las graficas de las autocorrelaciones residuales:



La grafica de las autocorrelaciones nos indica que para todos los retardos menos el retardo 12 no podemos rechazar que sean 0. Para explicar que el retardo 12 sea significativo existen dos posibilidades:

- Pertenece al porcentaje de falsos rechazos, lo que sucederá con una probabilidad α .
- La variable ventas depende del retardo 12, es decir las ventas de este trimestre dependen de las ventas de hace tres años.

Parece más asumible la primera posibilidad.

En conclusión, tenemos un modelo que explica bien, pero que tiene parámetros no significativos. Por ello eliminaremos del modelo el parámetro menos significativo (el retardo uno de la variable *Importaciones*).

Modelo 1

Comportamiento global del modelo:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	3	,722	19,946	17	,277	0

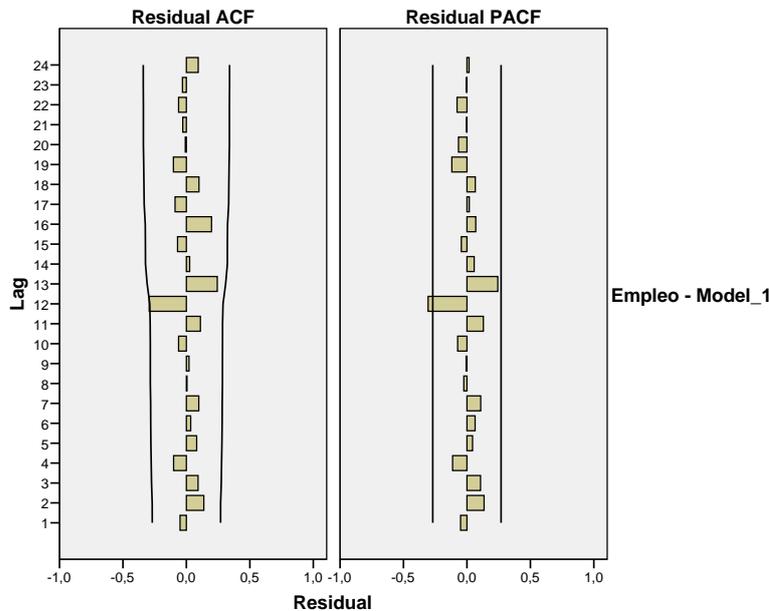
Nuevamente el test de Ljung-Box nos indica que el modelo es significativo, ya que su valor es mayor que 0.05. Al igual hemos obtenido un valor de R-cuadrado alto.

Comportamiento de los distintos parámetros:

ARIMA Model Parameters					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-3,416	2,411	-1,417	,163
			AR	Lag 1	,350	,145	2,406	,020
	Ventas	No Transformation	Numerator	Lag 0	,109	,104	1,048	,300
				Lag 1	-,197	,078	-2,538	,015
	Inversiones	No Transformation	Numerator	Lag 0	,265	,145	1,831	,074
				Lag 1	,166	,091	1,832	,074
				Numerator, Seasonal	Lag 1	-,639	,602	-1,061
	Importaciones	No Transformation	Numerator	Lag 0	,156	,118	1,318	,194

De la tabla anterior observamos algo importante, el retardo 0 de la variable ventas no es significativo, pero en cambio el retardo 1 de la misma variable si es significativo. Y esto nos indica que lo que pasa con *Ventas* en el trimestre actual no afecta al empleo, solo es afectado por las ventas del trimestre anterior. Debemos eliminar el retardo 0 de *Ventas* de nuestro modelo y dejar solo el retardo 1, si deseamos continuar con el modelo Backward.

Grafica de las correlaciones residuales:



Nuevamente la grafica de las correlaciones nos indica que para todos los retardos menos el retardo 12 no podemos rechazar que sean 0, asumimos que pertenece al porcentaje de rechazos falsos.

Seguimos eliminando lo que está de sobre en el modelo, el retardo 0 de la variable ventas.

Modelo 2

Comportamiento global del modelo

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	3	,714	25,272	17	,089	0

El valor de R_cuadrado sigue siendo alto y valor del test de Ljung-Box sigue siendo mayor que 0.05.

Comportamiento de los parámetros:

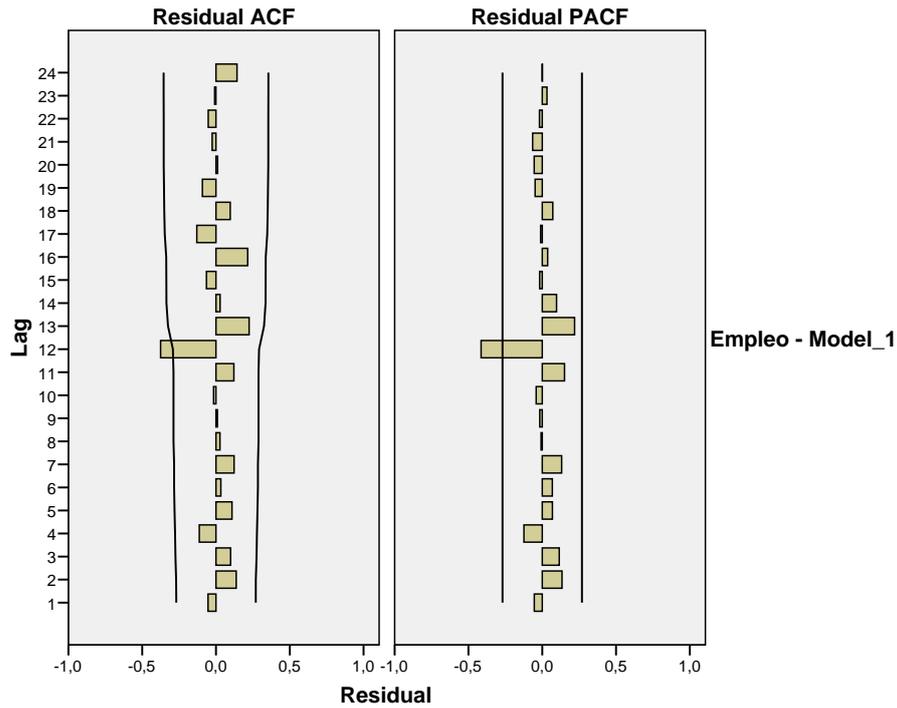
ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-2,828	2,331	-1,213	,231
			AR	Lag 1	,362	,143	2,541	,014
	Ventas	No Transformation	Delay		1			
			Numerator	Lag 0	,212	,078	2,723	,009
			Numerator	Lag 0	,285	,144	1,975	,054
Inversiones	No Transformation		Numerator	Lag 1	,185	,092	2,017	,050
			Numerator, Seasonal	Lag 1	-,540	,509	-1,062	,294
Importaciones	No Transformation		Numerator	Lag 0	,229	,095	2,405	,020

Ahora en nuestro modelo obtenemos que el retardo 1 de ventas es muy significativo como era de esperarse, pero el que es menos significativo es el retardo 1 de la parte estacionaria de la variable *Inversiones*, que es el que debemos eliminar como siguiente paso.

Que el retardo 1 de la parte estacionaria de *Inversiones* no es significativo nos indica que lo que se invirtió hace 1 año (4 trimestres) no afecta al empleo en el trimestre actual.

Grafica de las autocorrelaciones residuales



Nuevamente aparece el problema con el retardo 12, que lo asumimos de igual manera que en el caso anterior.

Modelo 3

Comportamiento global del modelo:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	3	,693	16,677	17	,476	0

El R-cuadrado sigue siendo alto, al igual que el valor para el test de Ljung-Box sigue siendo mayor que 0.05, por lo cual no rechazamos que las autocorrelaciones sean significativamente iguales a 0.

Comportamiento de los parámetros:

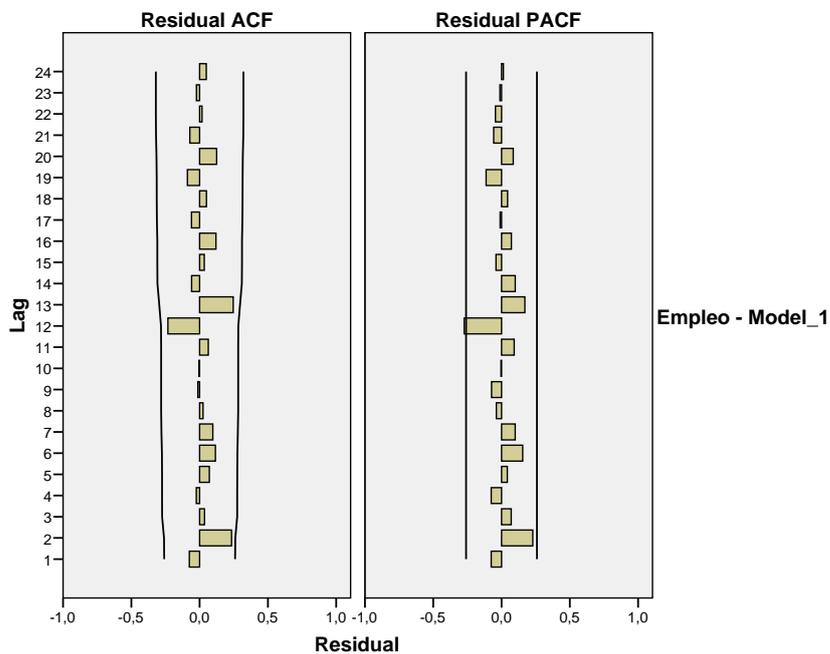
ARIMA Model Parameters

					Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant		-3,761	1,982	-1,897	,063
			AR	Lag 1	,291	,135	2,149	,036
	Ventas	No Transformation	Delay		1			
			Numerator	Lag 0	,168	,084	2,007	,050
	Inversiones	No Transformation	Numerator	Lag 0	,358	,140	2,550	,014
				Lag 1	,031	,113	,275	,785
	Importaciones	No Transformation	Numerator	Lag 0	,177	,109	1,617	,112

Ahora vemos que el retardo 1 de la variable *Inversiones* no es significativa, es el siguiente que debemos eliminar del modelo.

El retardo 1 de la variable *Inversiones* no es significativo, pero el retardo 0 de inversiones si los es. Esto nos indica que lo que se invirtió en el trimestre actual si influye sobre el empleo del mismo periodo, no así lo que se invirtió en el trimestre anterior.

Comportamiento de la grafica de las autocorrelaciones residuales



El grafico de las autocorrelaciones nos indica que el modelo está muy bien ya que todas las autocorrelaciones no son significativamente distintas de 0.

Procederemos a eliminar el retardo 1 de *Inversiones*.

Modelo 4.

Comportamiento global del modelo

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	3	,693	16,879	17	,463	0

El R-cuadrado sigue siendo alto, al igual que el valor para el test de Ljung-Box sigue siendo mayor que 0.05, por lo cual no rechazamos que las autocorrelaciones sean significativamente iguales a 0.

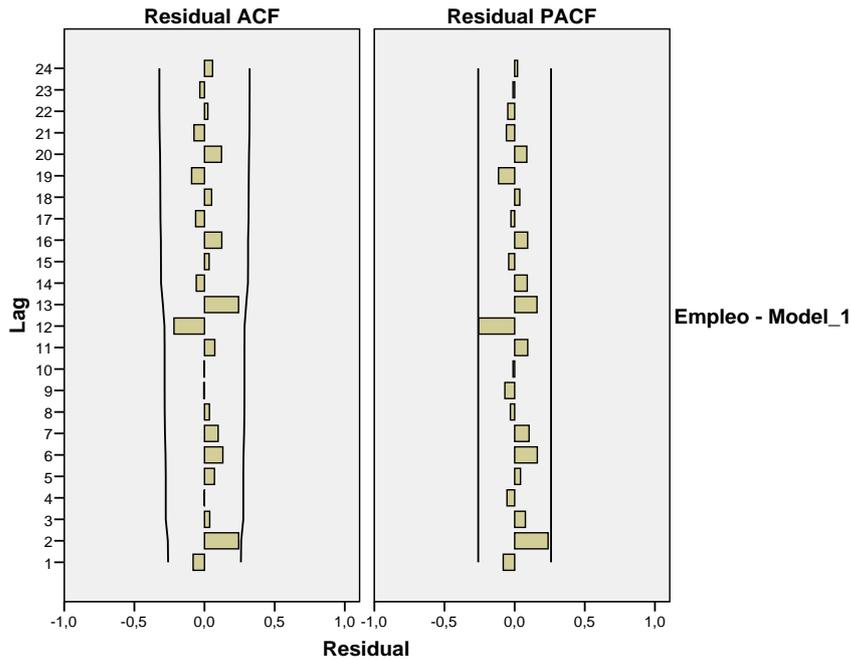
Comportamiento de los parámetros

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant	-4,000	1,789	-2,237	,030
			AR Lag 1	,298	,134	2,229	,030
	Ventas	No Transformation	Delay	1			
			Numerator Lag 0	,154	,068	2,249	,029
	Inversiones	No Transformation	Numerator Lag 0	,360	,139	2,584	,013
	Importaciones	No Transformation	Numerator Lag 0	,171	,107	1,602	,115

De la tabla podemos observar que definitivamente las *Importaciones* no dan mucha información para explicar el *Empleo*, debemos eliminar la variable *Importaciones* del modelo.

Grafica de las autocorrelaciones residuales:



El grafico de las autocorrelaciones nos indica que todas son no significativamente distintas de 0.

Modelo 5

Comportamiento global del modelo:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	2	,678	18,694	17	,346	0

Para este modelo observamos que el valor del R-cuadrado es alto, al igual que el valor para el test de Ljung-Box sigue siendo mayor que 0.05, por lo cual no rechazamos que las autocorrelaciones sean significativamente iguales a 0.

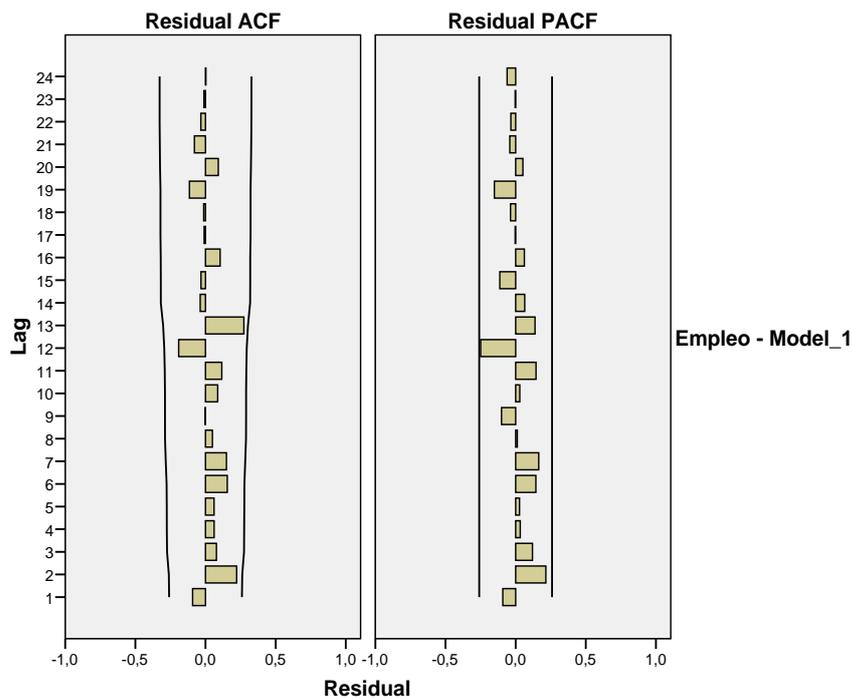
Comportamiento de los parámetros:

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant	-5,345	1,661	-3,219	,002
			AR Lag 1	,330	,131	2,517	,015
	Ventas	No Transformation	Delay	1			
			Numerator Lag 0	,174	,068	2,550	,014
	Inversiones	No Transformation	Numerator Lag 0	,536	,088	6,101	,000

En la tabla observamos que todas son significativas y por lo tanto deben pertenecer al modelo final. Es decir, el *Empleo* del trimestre actual depende del *Empleo* y de las *Ventas* del trimestre anterior y de la *Inversión* actual. Y la que mayor influencia tiene a igual variación es la inversión hecha en el trimestre actual.

Gráfica de las autocorrelaciones residuales:



Observamos que las autocorrelaciones residuales del modelo son todas no significativamente distintas de cero.

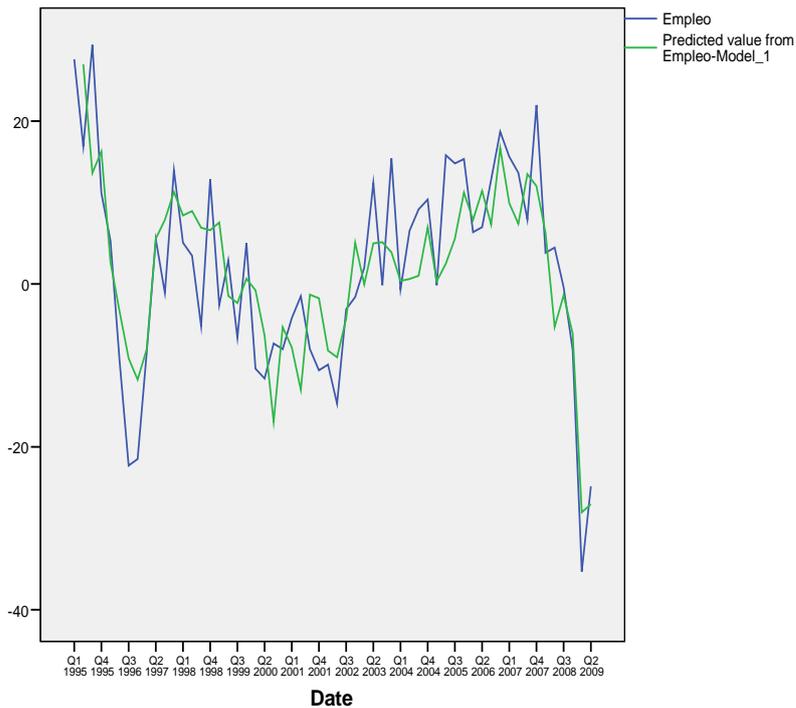
Modelo Final por el método Backward

El modelo 5 será el resultante por el método Backward para la variable Empleo:

$$Empleo_t = 0.330Empleo_{t-1} + 0.174Ventas_{t-1} + 0.536Inversion_t - 5.345 + a_t$$

Lo que nos indica que el empleo actual depende del empleo y ventas del trimestre anterior, y de la inversión hecha en el trimestre actual.

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar un buen ajuste.



Comparación de los modelos Forward y Backward y determinación del modelo final para empleo.

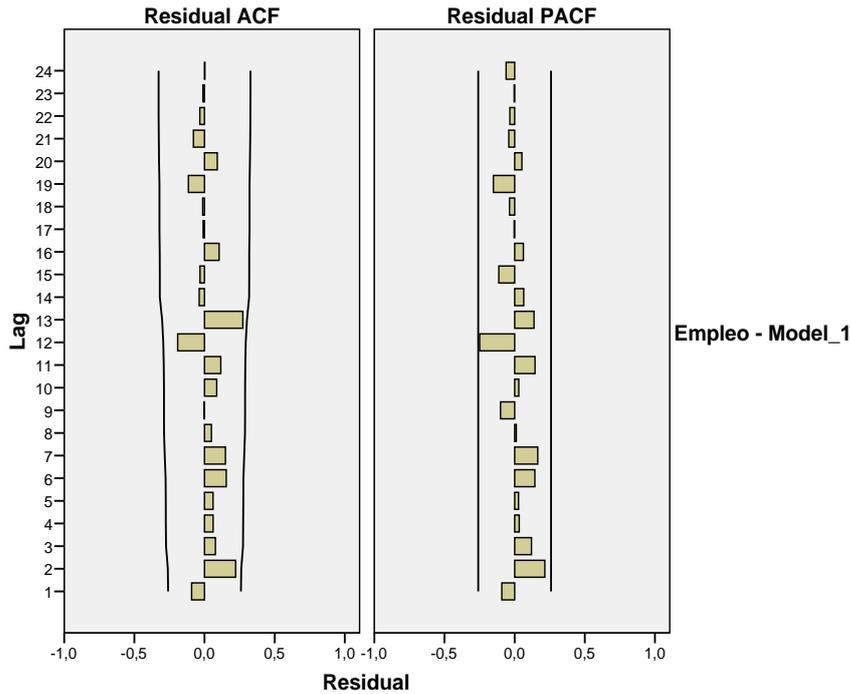
El método Backward, da el siguiente modelo:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	2	,678	18,694	17	,346	0

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant	-5,345	1,661	-3,219	,002
			AR Lag 1	,330	,131	2,517	,015
	Ventas	No Transformation	Delay	1			
			Numerator Lag 0	,174	,068	2,550	,014
	Inversiones	No Transformation	Numerator Lag 0	,536	,088	6,101	,000



$$Empleo_t = 0.330Empleo_{t-1} + 0.174Ventas_{t-1} + 0.536Inversion_t - 5.345 + a_t$$

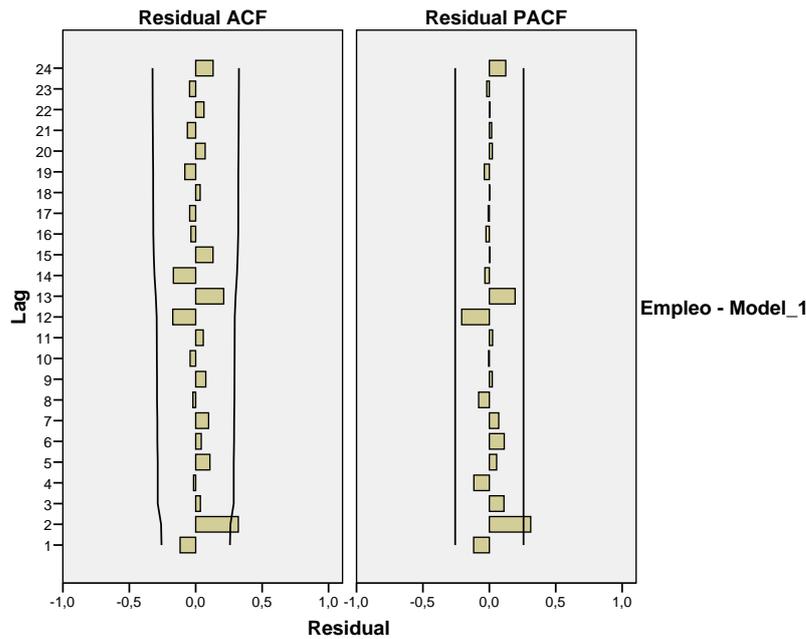
El método Forward, da el siguiente modelo:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	2	,665	19,360	17	,308	0

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant	-2,782	1,895	-1,468	,148
			AR Lag 1	,329	,129	2,549	,014
	Inversiones	No Transformation	Numerator Lag 0	,352	,146	2,412	,019
	Importaciones	No Transformation	Numerator Lag 0	,254	,109	2,332	,023



$$Empleo_t = 0.329Empleo_{t-1} + 0.352Inversiones_t + 0.254Importaciones - 2.782 + a_t$$

En primer lugar debemos constatar que no resulta el mismo modelo por ambos métodos. Cuando esto sucede, se debe plantear por qué se produce esto y analizar las posibles consecuencias.

El motivo más normal por el que los métodos Forward y Backward no coinciden, es por la presencia de multicolinealidad en el modelo. Además, hemos visto en el cálculo de las correlaciones cruzadas, que existe correlación entre todas las variables, por lo que es bastante plausible la existencia de multicolinealidad en el modelo.

Para eliminar esta multicolinealidad, debemos ver el funcionamiento del modelo eliminando alguna de las variables. Comenzaremos estudiando el modelo en el que están presentes la variable Empleo mediante un modelo AR(1) y la variable inversiones sin retardos (dado que es la variable presente en los dos modelos).

Con esta consideración el modelo obtenido es el siguiente.

$$Empleo_t = -4.323 + 0.419Empleo_{t-1} + 0.602Inversiones_t + a_t$$

Comportamiento global del modelo:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Empleo-Model_1	1	,632	22,227	17	,176	0

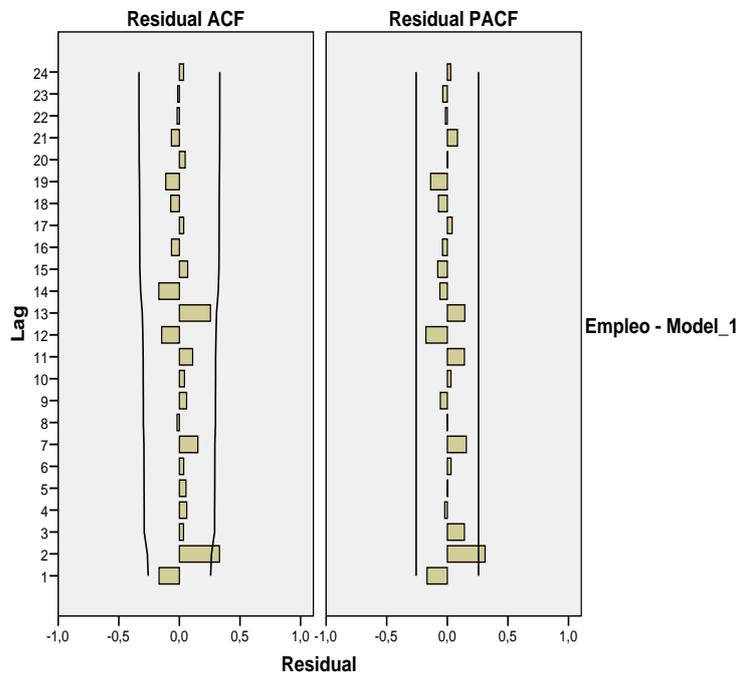
Observamos que el valor del R-cuadrado es bueno, por lo que el modelo es informativo. El test de Ljung-Box también tiene un valor mayor que 0.05, por lo que no rechazamos que las autocorrelaciones de los residuos son iguales a cero.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Empleo-Model_1	Empleo	No Transformation	Constant	-4,323	2,001	-2,161	,035
			AR Lag 1	,419	,125	3,347	,001
	Inversiones	No Transformation	Numerator Lag 0	,602	,094	6,397	,000

Todos los parámetros son significativos.

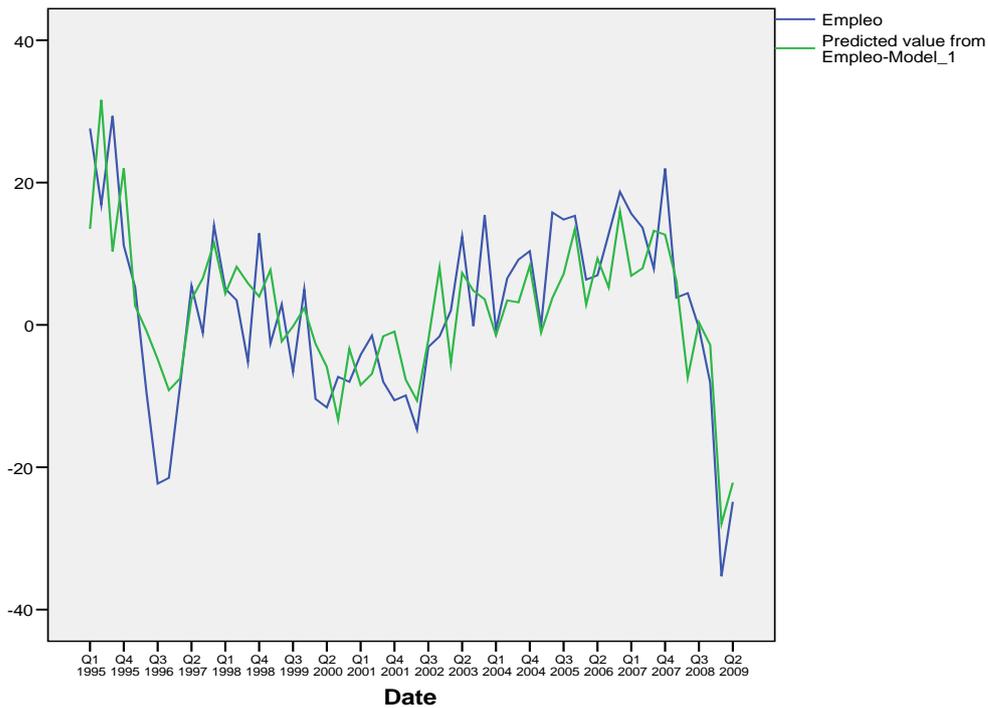
La grafica de las autocorrelaciones de los residuos:



Solo el retardo 2 es significativo, pero si intentamos ajustar un modelo con AR(2) el resultado es peor (aparecen parámetros no significativos), por lo que asumiremos que ese retardo puede deberse al azar.

Este modelo presenta un R-cuadrado similar al de los modelos Forward y Backward. Además no presenta problemas de multicolinealidad por depender solo de una variable externa. Por ello será el modelo final.

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar que el ajuste es bueno. Notablemente mejor que el obtenido sin el uso de la variable inversiones.



Para ajustar las demás variables se ha utilizado el mismo método que en la variable Empleo y los modelos son los siguientes:

5.2.3. Modelo para la serie *Ventas*

Siguiendo el mismo procedimiento que para la serie *Empleo*, obtenemos el modelo:

$$Ventas_t = 4.490 + 0.535Ventas_{t-1} + 0.719Importaciones_t + a_t$$

El comportamiento global del modelo es el siguiente:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Ventas-Model_1	1	,718	11,326	17	,839	0

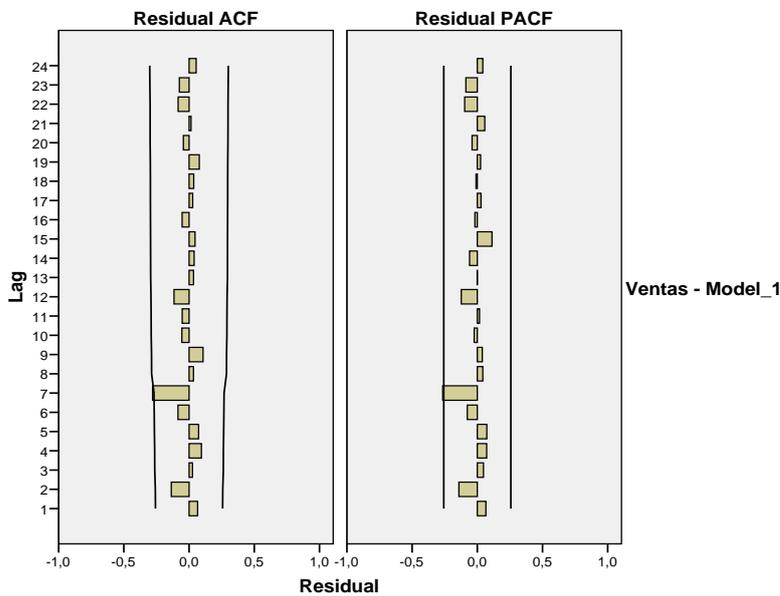
Observamos que el valor del R- cuadrado es alto, por lo que el modelo es informativo. El test de Ljung- Box también tiene un valor alto, por lo que no rechazamos que las autocorrelaciones de los residuos son iguales a cero.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Ventas-Model_1	Ventas	No Transformation	Constant	4,490	2,772	1,620	,111
			AR Lag 1	,535	,121	4,409	,000
	Importaciones	No Transformation	Numerator Lag 0	,719	,093	7,694	,000

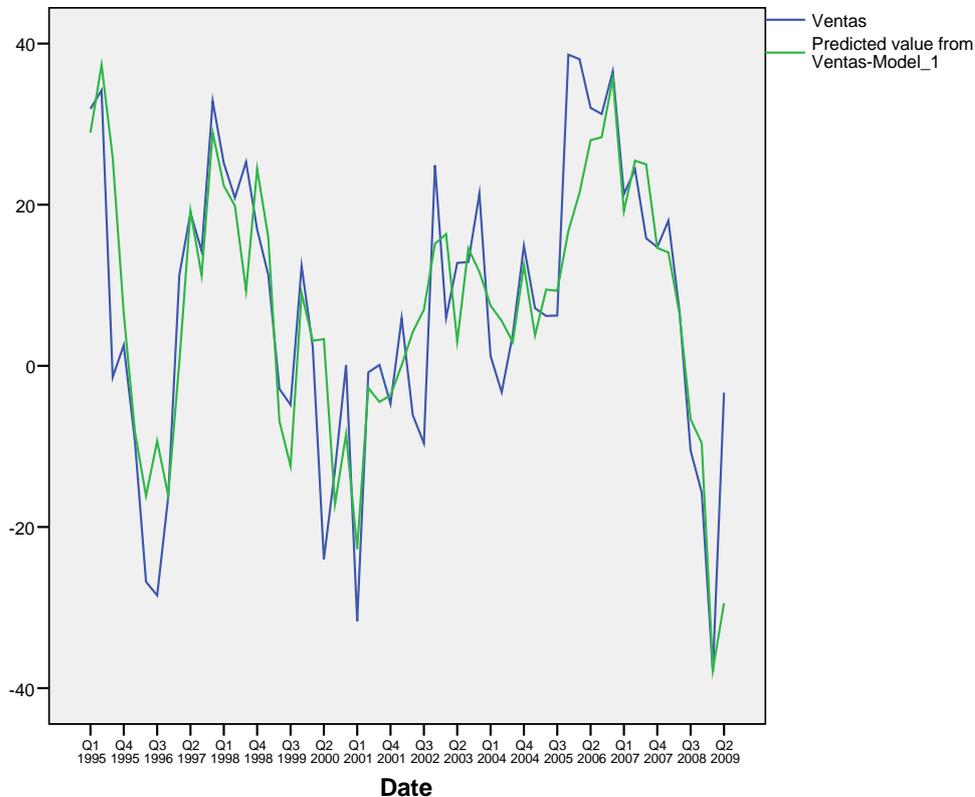
De la tabla de los parámetros observamos que todos son significativos, por lo que la parte autorregresiva de ventas y el retardo actual de Importaciones deben aparecer en el modelo.

La grafica para las autocorrelaciones de los residuos:



La grafica muestra que las autocorrelaciones de los residuos son todas no significativamente distintas de cero.

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar diferencias notables en algunos casos, esto se debe al valor de R cuadrado.



5.2.4. Modelo para la serie *Inversiones*

El modelo para la serie *Inversiones* es el siguiente:

$$Inversiones_t = 8.378 + 0.581Inversiones_{t-4} + 0.322Empleo_t + 0.421Importaciones_t + a_t$$

El comportamiento global del modelo es el siguiente:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Inversiones-Model_1	2	,821	20,321	17	,258	0

Observamos que el valor del R- cuadrado es alto, por lo que el modelo es informativo.

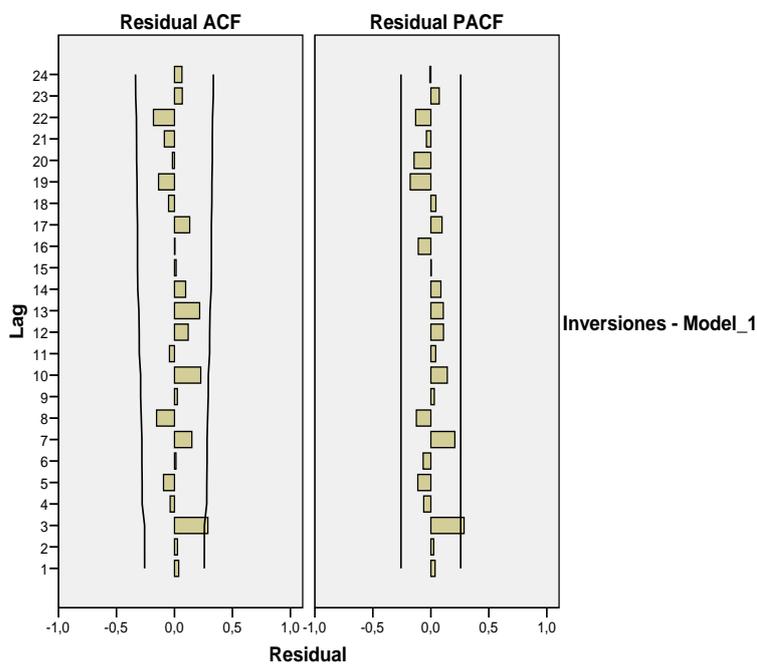
El test de Ljung- Box también tiene un valor mayor que 0.05, por lo que no rechazamos que las autocorrelaciones de los residuos son iguales a cero.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Inversiones-Model_1	Inversiones	No Transformation	Constant		8,378	1,683	4,977	,000
			AR, Seasonal	Lag 1	,581	,113	5,124	,000
	Empleo	No Transformation	Numerator	Lag 0	,322	,083	3,888	,000
	Importaciones	No Transformation	Numerator	Lag 0	,421	,054	7,808	,000

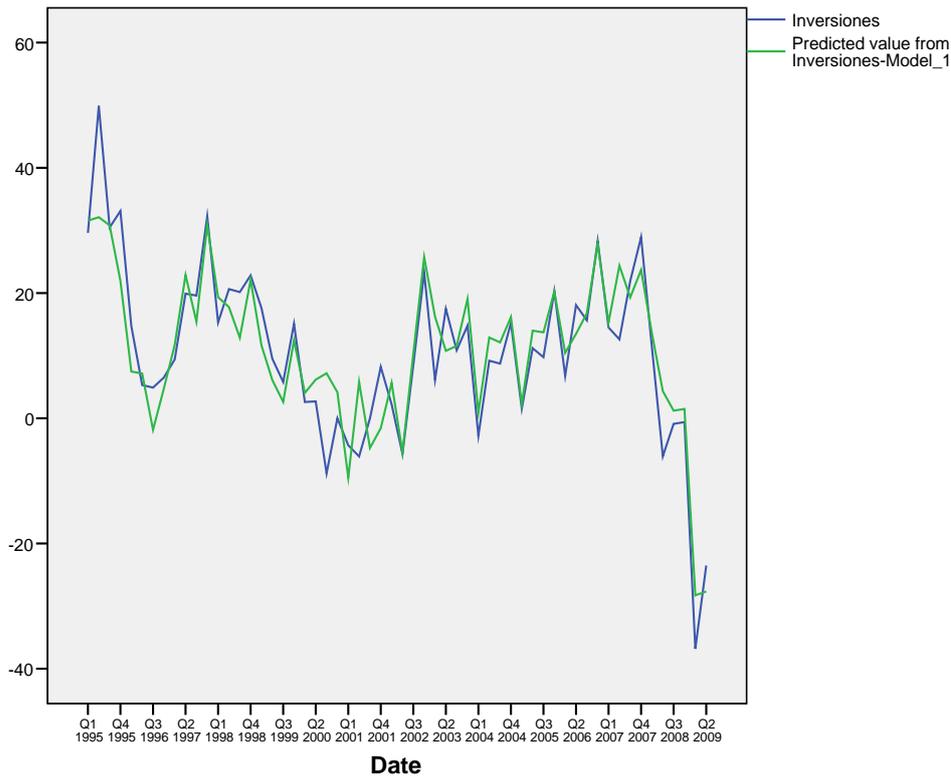
De la tabla de los parámetros observamos que todos son significativos y diferentes de cero, por lo que deben aparecer en el modelo.

La grafica para las autocorrelaciones de los residuos:



De la grafica observamos que la única autocorrelación que diferente de cero es la tercera, asumiremos que se debe a los falsos positivos dado que al ajustar modelos como un AR(3) o MA(3) los resultados son incoherentes.

Por último se presenta el gráfico de los valores observados y los valores predichos. Se pueden observar un buen ajuste.



5.2.5. Modelo para la serie *Importaciones*

El modelo para la serie *Importaciones* es el siguiente:

$$Importaciones_t = -7.58 + 0.392Ventas_t + 0.859Inversiones_t + a_t$$

El comportamiento global del modelo es el siguiente:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Importaciones-Model_1	2	,827	20,149	18	,324	0

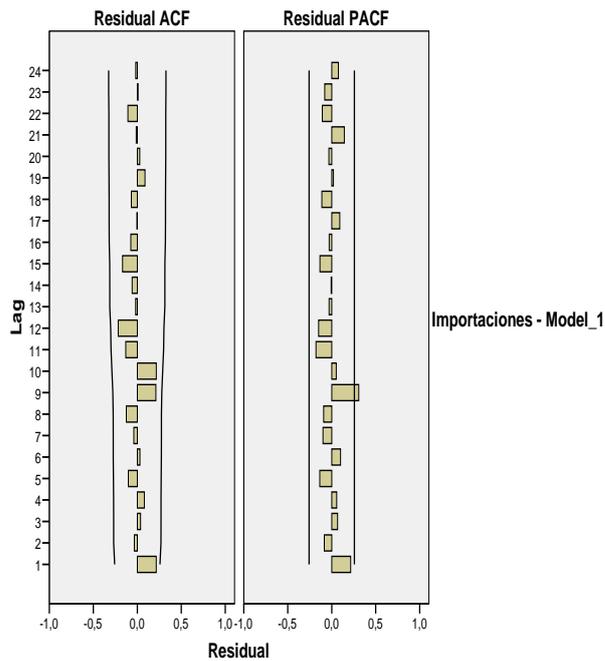
Observamos que el valor del R- cuadrado es alto, por lo que el modelo es informativo.

El test de Ljung- Box también tiene un valor mayor que 0.05, por lo que no rechazamos que las autocorrelaciones de los residuos son iguales a cero.

ARIMA Model Parameters

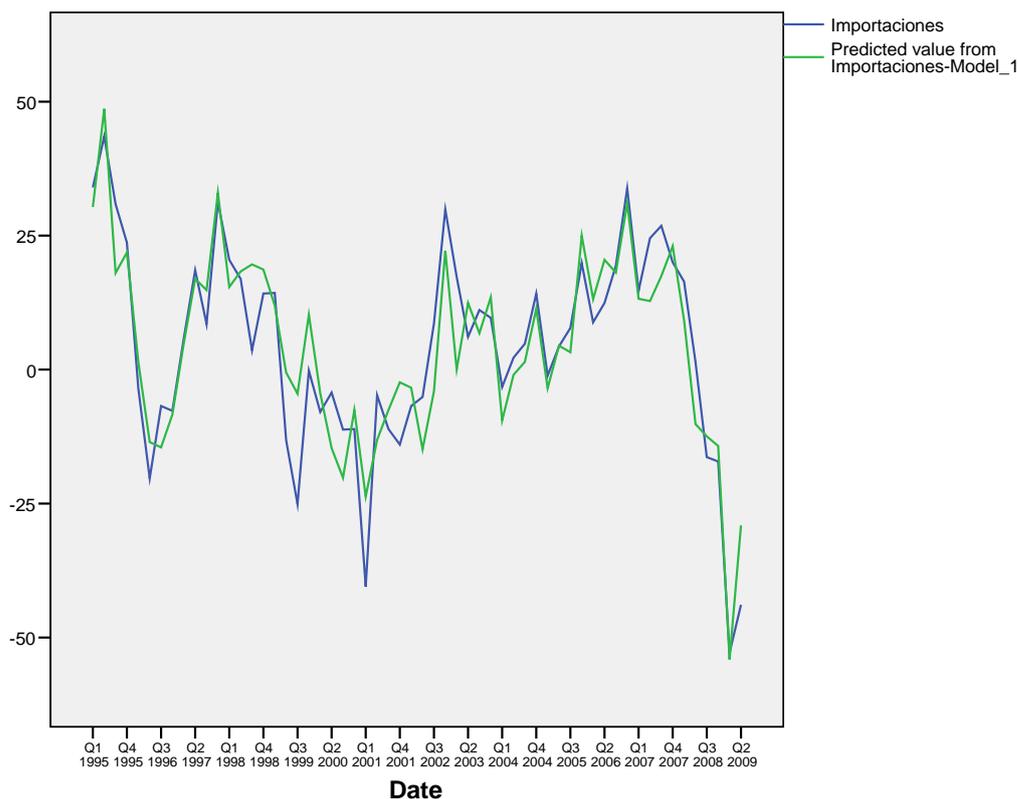
					Estimate	SE	t	Sig.
Importaciones-Model_1	Importaciones	No Transformation	Constant		-7,580	1,368	-5,539	,000
	VentasGlobal	No Transformation	Numerator	Lag 0	,392	,080	4,920	,000
	Inversiones	No Transformation	Numerator	Lag 0	,859	,104	8,232	,000

De la tabla de los parámetros observamos que todos son significativos y diferentes de cero, por lo que deben aparecer en el modelo.



De la grafica observamos que las autocorrelaciones son no significativamente distintas de cero.

Por último se presenta el gráfico de los valores observados y los valores predichos. . Se pueden observar un buen ajuste.



5.3. Comparación de los modelos ARIMA y los modelos de regresión dinámica.

5.3.1. Serie *Empleo*

El modelo ARIMA que obtuvimos fue el siguiente:

$$Empleo_t = 1.946 + 0.692Empleo_{t-1} + a_t$$

Con un valor de R^2 de 0.392.

El modelo de regresión dinámica es el siguiente:

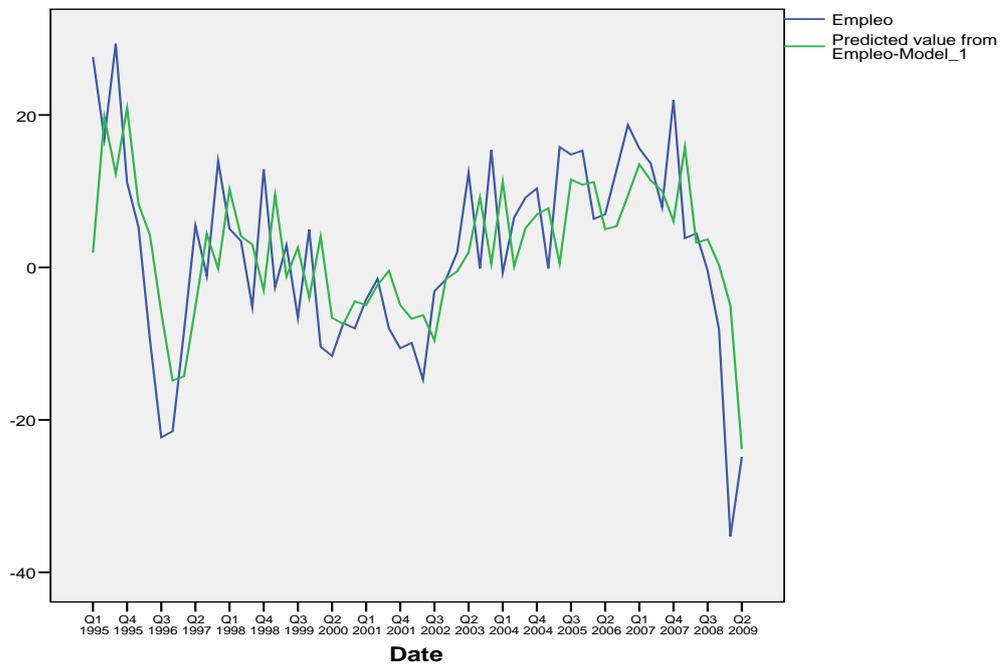
$$Empleo_t = -4.323 + 0.419Empleo_{t-1} + 0.602Inversiones_t + a_t$$

Con un valor de R^2 de 0.632.

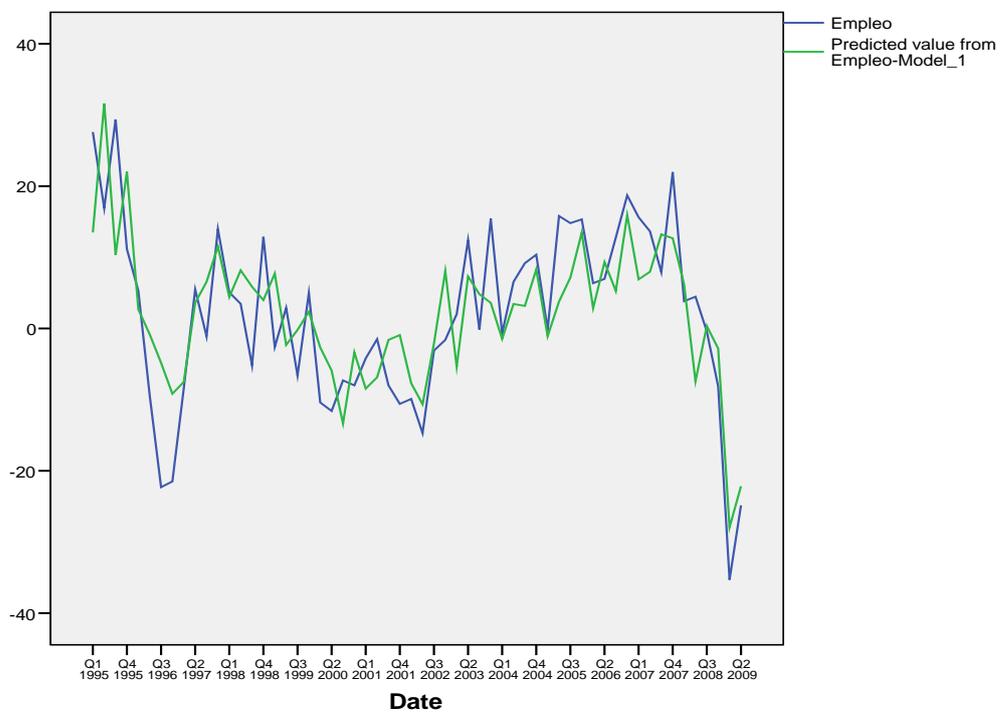
Podemos observar claramente que el modelo ajustado con la regresión dinámica ajusta mejor que el del modelo ARIMA, ya que obtenemos un valor más alto de R^2 . Esto

implica que los valores ajustados son mejores. Lo cual lo podemos observar comparando las graficas:

Ajuste modelo ARIMA



Ajuste modelo de regresión dinámica



5.3.2 Serie Ventas

El modelo ARIMA que obtuvimos fue el siguiente:

$$Ventas_t = 5.107 + 0.101Ventas_{t-1} + a_t$$

Con un valor de R^2 de 0.409.

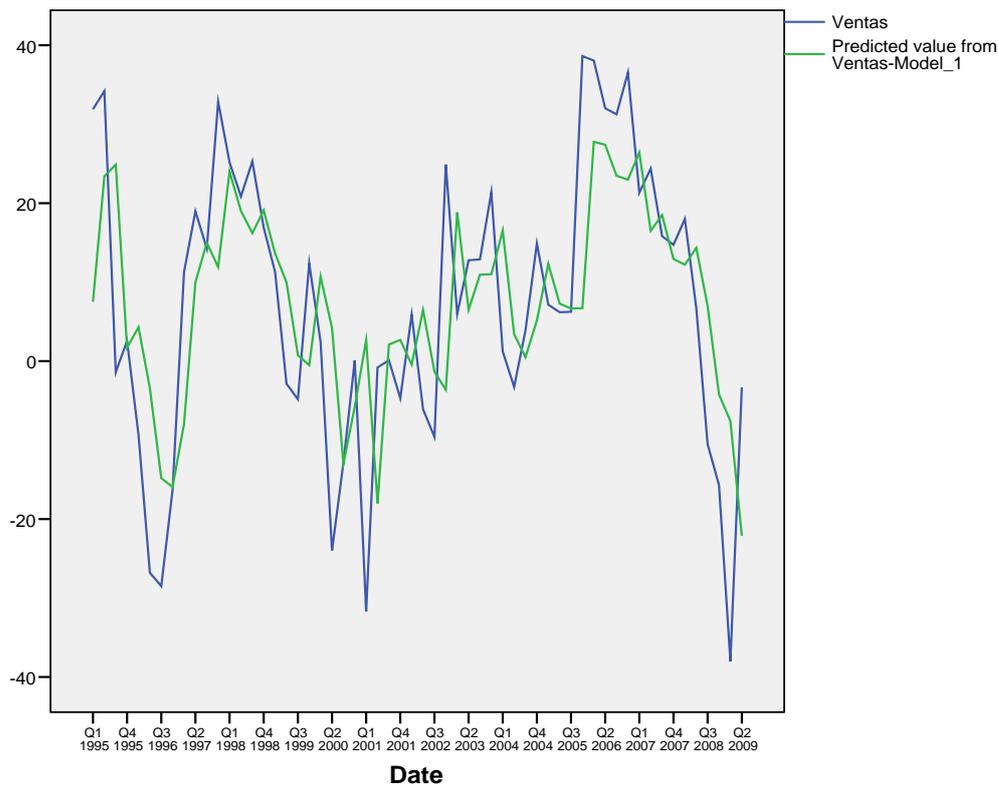
El modelo de regresión dinámica es el siguiente:

$$Ventas_t = 4.490 + 0.535Ventas_{t-1} + 0.719Importaciones_t + a_t$$

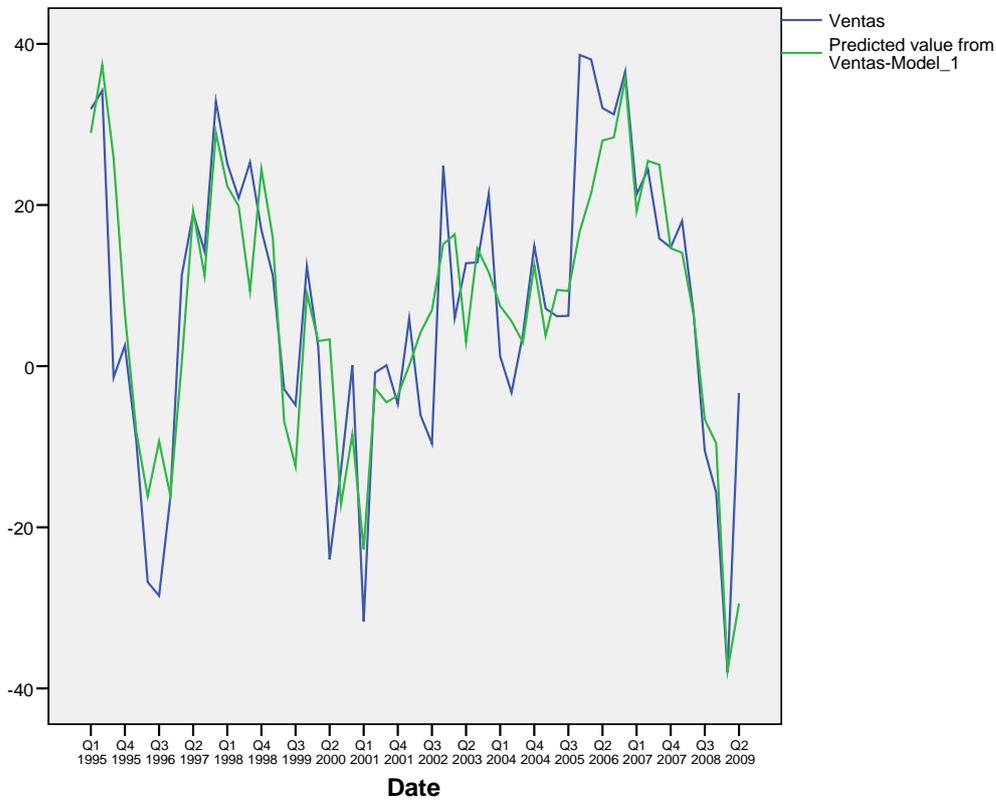
Con un valor de R^2 de 0.718.

Las conclusiones son similares a las obtenidas para la variable Empleo, el modelo de regresión dinámica presenta mejor ajuste. Las gráficas de ajuste para los dos modelos son:

Ajuste modelo ARIMA



Ajuste modelo de regresión dinámica



5.3.3. Serie *Inversiones*

El modelo ARIMA que obtuvimos fue el siguiente:

$$Inversiones_t = 10.13 + 0.8Inversiones_{t-1} + 0.973Inversiones_{t-4} + 0.798a_{t-4} + a_t$$

Con un valor de R^2 de 0.537.

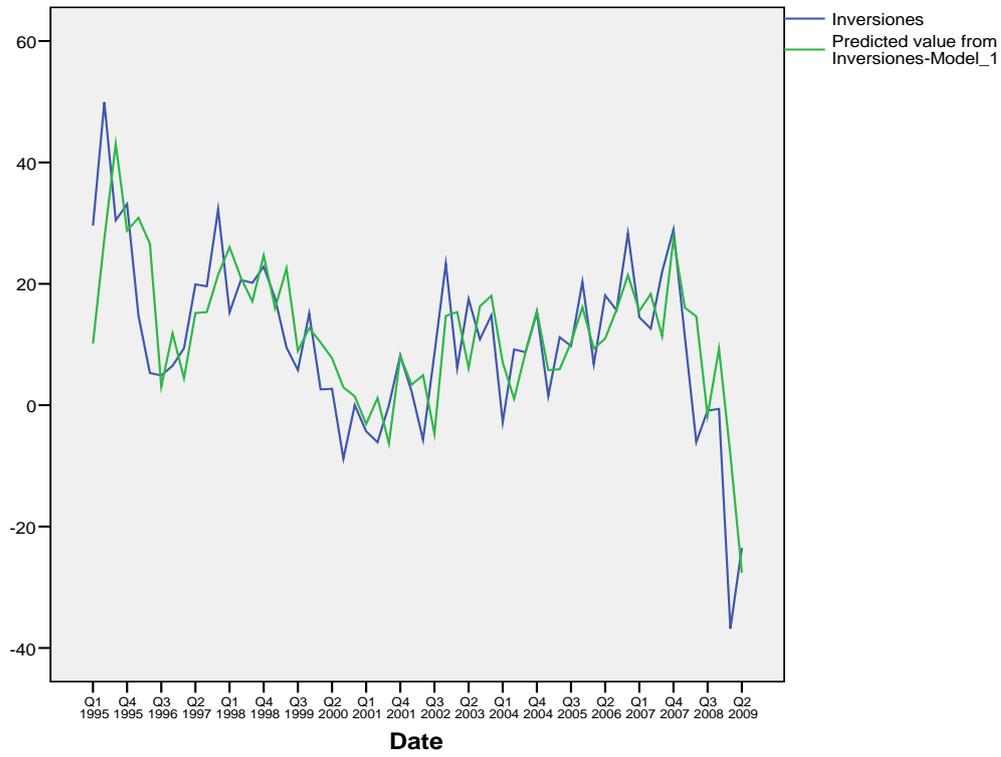
El modelo de regresión dinámica es el siguiente:

$$Inversiones_t = 8.378 + 0.581Inversiones_{t-4} + 0.322Empleo_t + 0.421Importaciones_t + a_t$$

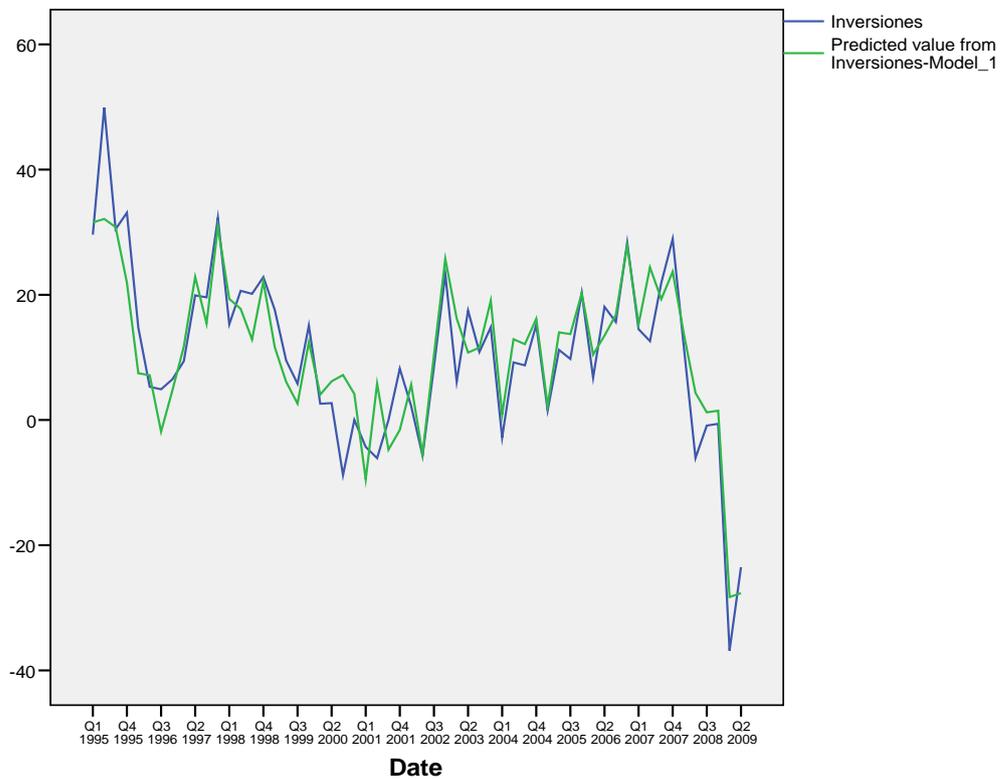
Con un valor de R^2 de 0.810.

Las conclusiones son similares a las obtenidas para las otras dos variables, el modelo de regresión dinámica presenta mejor ajuste. Las gráficas de ajuste para los dos modelos son:

Ajuste modelo ARIMA



Ajuste modelo de regresión dinámica.



5.3.4. Serie *Importaciones*

El modelo ARIMA que obtuvimos fue el siguiente:

$$Importaciones_t = 3.431 + 0.759Importaciones_{t-1} + a_t$$

Con un valor de R^2 de 0.487.

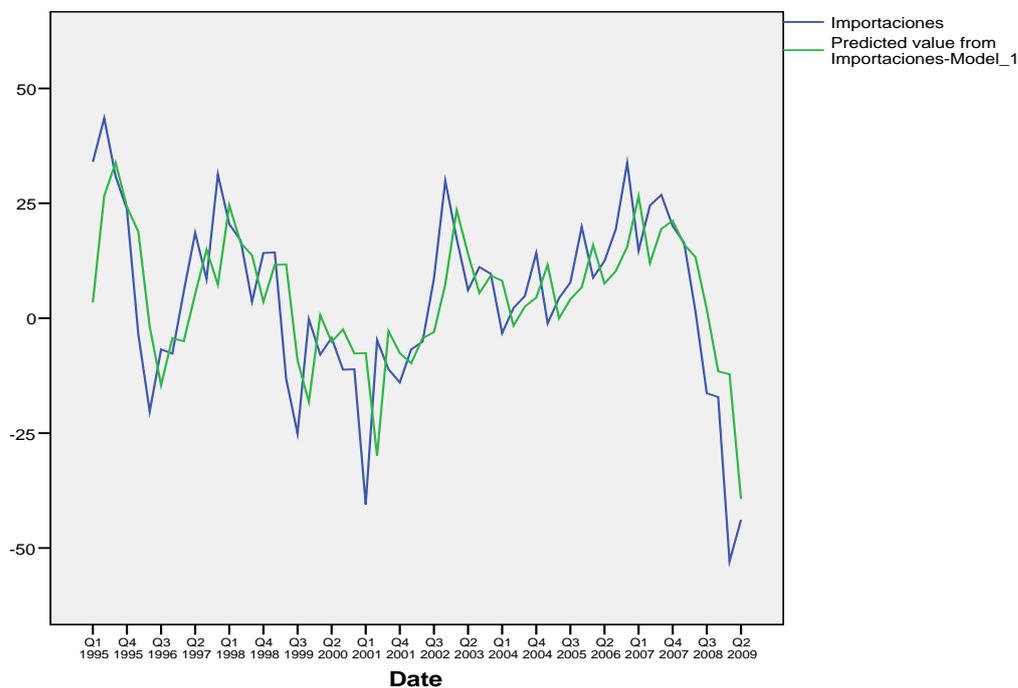
El modelo de regresión dinámica es el siguiente:

$$Importaciones_t = -7.58 + 0.392Ventas_t + 0.859Inversiones_t + a_t$$

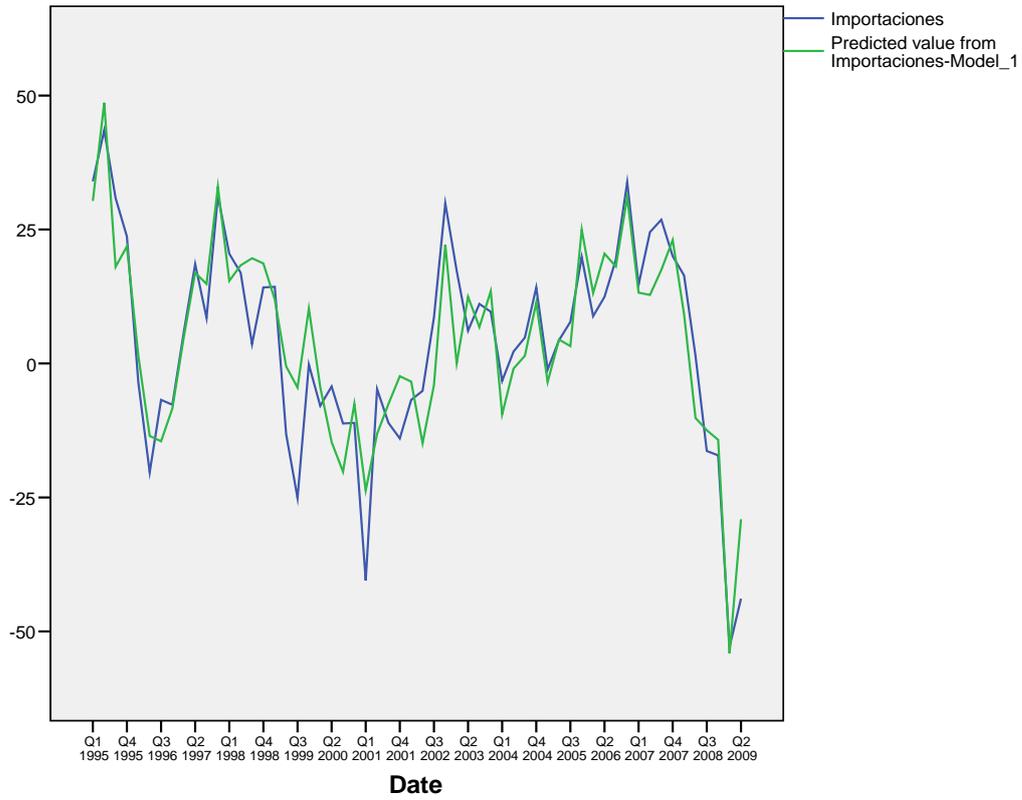
Con un valor de R^2 de 0.827.

Las conclusiones son similares a las obtenidas para las otras tres variables, el modelo de regresión dinámica presenta mejor ajuste. Las gráficas de ajuste para los dos modelos son:

Ajuste modelo ARIMA



Ajuste modelo de regresión dinámica



En conclusión, todas las variables ven mejorados sus ajustes si pueden contar con la información actualizada de otras variables.

6. COMPARACIÓN DE DATOS

En este capítulo haremos una comparación cualitativa de las 20 series de datos que tenemos en total. La comparación se hará por sector y por variables. Dicha comparación se realizará con un índice basado en el índice cualitativo de similitud (QSI), concretamente una cota inferior del QSI (cQSI). Recordemos que este índice (cQSI) se calcula entre dos variables, da un valor entre 0 y 1 y mide el porcentaje de datos con un perfil coincidente en las dos series.

Los pasos para calcularlo son cuatro:

1. Normalizar las variables.
2. Diferenciar las variables una vez normalizadas.
3. Etiquetar las variables diferenciadas.
4. Enumerar las coincidencias entre dos series ya etiquetadas.

Todos estos pasos se realizarán como se indicó en el marco teórico. Solo debemos especificar el valor que debe tomar el parámetro δ , en el etiquetado de las series.

Para nuestro estudio vamos a tomar $\delta = 4$. Hemos decidido este valor por presentar una distribución bastante equitativa en las seis categorías que define (En nuestro caso serán seis categoría en lugar de siete, dado que en ninguna serie el valor permanece constante entre dos años).

Veamos la tabla para realizar el etiquetado con δ igual a cuatro:

Etiqueta	Intervalo	Símbolo
Fuerte incremento	$[0.25, 1]$	1
Incremento moderado	$[0.0625, 0.25]$	2
Ligero incremento	$[0, 0.0625]$	3
Ligero decremento	$[-0.0625, 0]$	4
Decremento moderado	$[-0.25, -0.0625]$	5
Fuerte decremento	$[-1, -0.25]$	6

Una vez realizado el proceso de etiquetado para ambas variables, solo nos resta buscar los instantes donde las etiquetas de las dos series coincidan. Para ello nos auxiliaremos de la tabla de contingencia de ambas variables, ya que la diagonal de dicha tabla nos dará el número de coincidencias para cada etiqueta.

Una vez calculado el valor de la cota del QSI, sería deseable poder indicar si es significativamente diferente de lo que encontraríamos como fruto del azar.

Si asumimos que las seis categorías son igualmente probables (que es el criterio usado para determinar el valor de δ) y asumimos independencia entre las mismas, la cota del QSI sigue una distribución Binomial con parámetros $n=57$ (dado que tenemos 57 datos etiquetados) y $p=1/6$ (dato que de las 36 posibles combinaciones, seis de ellas son coincidentes).

Si planteamos el contraste de hipótesis destinado a determinar si existe una correlación positiva entre las dos series (en términos del valor de $cQSI$), debemos estudiar si la probabilidad asociada a la binomial es mayor o menor que $1/6$. Formalmente:

$$H_0 : p \leq 1/6 \quad H_1 : p > 1/6$$

Para este contraste y en función de los distintos valores del error tipo I, la región crítica será:

$$R.C : \begin{cases} Si & \alpha = 0.01, & cQSI \geq 16/57 \\ Si & \alpha = 0.05, & cQSI \geq 14/57 \\ Si & \alpha = 0.10, & cQSI \geq 13/57 \end{cases}$$

6.1. Comparación por Sectores.

Primero compararemos las cuatro variables: *Ventas*, *Empleo*, *Inversiones* e *Importaciones* para cada uno de los cuatro sectores y la unión de todos ellos: *Servicios*, *Industria*, *Comercio*, *Construcción* y *Global*.

6.1.1. Sector Servicios.

Ventas vs Empleo

Luego de realizar los tres pasos: normalización, diferenciación y etiquetado, para calcular la cota del QSI solo debemos buscar el número de coincidencias para cada etiqueta, para ello mostramos la tabla de contingencia para las variables etiquetadas:

Etiqueta_VentasServicios * Etiqueta_EmpleoServicios Crosstabulation

Count		Etiqueta_EmpleoServicios						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasServicios	1,00	2	3	1	2	0	2	10
	2,00	0	3	2	2	2	2	11
	3,00	1	3	1	1	3	0	9
	4,00	0	0	0	0	1	2	3
	5,00	2	2	1	4	2	0	11
	6,00	3	1	2	3	1	3	13
Total		8	12	7	12	9	9	57

De esta tabla es fácil observar que por ejemplo, para la etiqueta 1 hay dos valores que coinciden, para la etiqueta 2 hay 3 valores que coinciden y así sucesivamente, por lo que la cota para el QSI es:

$$c_{QSI} = \frac{2+3+1+0+2+3}{57} = 0.1929$$

Esta cota nos indica que el 19.29% de los símbolos son coincidentes, o lo que es lo mismo que la series coinciden en un 19.29% en su comportamiento en los mismos instantes temporales. Es un porcentaje bajo, por lo que las *Ventas* y el *Empleo* del Sector servicios presentaron un comportamiento diferente en el periodo de estudio.

Ventas vs Inversiones

Etiqueta_VentasServicios * Etiqueta_InversionesServicios Crosstabulation

Count		Etiqueta_InversionesServicios						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasServicios	1,00	0	5	3	0	1	1	10
	2,00	3	4	0	0	3	1	11
	3,00	0	1	3	3	2	0	9
	4,00	0	1	0	1	1	0	3
	5,00	0	3	1	3	1	3	11
	6,00	0	4	1	3	3	2	13
Total		3	18	8	10	11	7	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+4+3+1+1+2}{57} = 0.1929$$

Lo que indica que la series coinciden en un 19.29% en su comportamiento en los mismos instantes.

Ventas vs Importaciones

Etiqueta_VentasServicios * Etiqueta_ImportacionesServicios Crosstabulation

Count		Etiqueta_ImportacionesServicios						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasServicios	1,00	2	3	0	0	4	1	10
	2,00	3	6	0	1	1	0	11
	3,00	0	2	2	0	4	1	9
	4,00	0	1	0	0	1	1	3
	5,00	0	2	2	4	1	2	11
	6,00	0	5	0	1	4	3	13
Total		5	19	4	6	15	8	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{2+6+2+0+1+3}{57} = 0.2456$$

Lo que indica que la series coinciden en un 24.56% en su comportamiento en los mismos instantes.

Empleo vs Inversiones.

Etiqueta_EmpleoServicios * Etiqueta_InversionesServicios Crosstabulation

Count		Etiqueta_InversionesServicios						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	4	1	1	2	0	8
EmpleoServicios	2,00	1	5	2	0	3	1	12
	3,00	1	2	1	2	1	0	7
	4,00	0	3	1	3	3	2	12
	5,00	0	2	1	4	0	2	9
	6,00	1	2	2	0	2	2	9
Total		3	18	8	10	11	7	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+5+1+3+0+2}{57} = 0.1929$$

Lo que indica que la series coinciden en un 19.29% en su comportamiento en los mismos instantes.

Empleo vs Importaciones

Etiqueta_EmpleoServicios * Etiqueta_ImportacionesServicios Crosstabulation

Count		Etiqueta_ImportacionesServicios						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	3	0	1	1	3	8
EmpleoServicios	2,00	2	4	2	1	3	0	12
	3,00	0	3	1	2	0	1	7
	4,00	2	3	1	1	4	1	12
	5,00	0	4	0	0	4	1	9
	6,00	1	2	0	1	3	2	9
Total		5	19	4	6	15	8	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+4+1+1+4+2}{57} = 0.2105$$

Lo que indica que la series coinciden en un 21.05% en su comportamiento en los mismos instantes.

Inversiones vs Importaciones

Etiqueta_InversionesServicios * Etiqueta_ImportacionesServicios Crosstabulation

Count		Etiqueta_ImportacionesServicios						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	2	1	0	0	0	0	3
InversionesServicios	2,00	2	6	2	3	2	3	18
	3,00	0	5	0	0	3	0	8
	4,00	0	3	2	1	4	0	10
	5,00	1	4	0	1	2	3	11
	6,00	0	0	0	1	4	2	7
Total		5	19	4	6	15	8	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{2+6+0+1+2+2}{57} = 0.2280$$

Lo que indica que la series coinciden en un 22.80% en su comportamiento en los mismos instantes.

Resumen.

En la siguiente tabla se resume la información del sector Servicios, indicando la significación de cada una de las relaciones.

	<i>Ventas</i>	<i>Empleo</i>	<i>Inversiones</i>	<i>Importaciones</i>
<i>Ventas</i>		0.1929(0.2325)	0.1929(0.2325)	0.2456 (0.0435)
<i>Empleo</i>	0.1929(0.2325)		0.2105(0.1437)	0.1929 (0.2325)
<i>Inversiones</i>	0.1929 (0.2325)	0.2105 (0.1437)		0.2280 (0.0822)
<i>Importaciones</i>	0.2456 (0.0435)	0.1929 (0.2325)	0.2280(0.0822)	

Se mostrarán en color rojo los valores significativos con un $\alpha > 0.01$. Con color naranja los valores significativos con un $\alpha > 0.05$. Y con color verde los valores significativos con un $\alpha > 0.10$.

Se puede concluir que pese a observarse cierta relación entre todas las variables en el sector Servicios, solo es posible determinar la significación entre importaciones con ventas e inversiones.

6.1.2. Sector Comercio

Ventas vs Empleo

Etiqueta_VentasComercio * Etiqueta_EmpleoComercio Crosstabulation

Count		Etiqueta_EmpleoComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasComercio	1,00	2	1	0	1	0	0	4
	2,00	3	5	3	0	2	0	13
	3,00	1	3	2	1	3	1	11
	4,00	0	2	3	0	2	0	7
	5,00	0	5	0	3	5	4	17
	6,00	0	0	1	0	1	3	5
Total		6	16	9	5	13	8	57

Por lo que la cota para el QSI es

$$c_{QSI} = \frac{2+5+2+0+5+3}{57} = 0.2982$$

Lo que indica que la series coinciden en un 29.82% en su comportamiento en los mismos instantes.

Ventas vs Inversiones

Etiqueta_VentasComercio * Etiqueta_InversionesComercio Crosstabulation

Count		Etiqueta_InversionesComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasComercio	1,00	1	2	1	0	0	0	4
	2,00	0	5	3	2	3	0	13
	3,00	0	6	3	2	0	0	11
	4,00	0	3	0	1	2	1	7
	5,00	0	2	4	2	8	1	17
	6,00	0	1	0	0	2	2	5
Total		1	19	11	7	15	4	57

Por lo que la cota para el QSI es

$$c_{QSI} = \frac{1+5+3+1+8+2}{57} = 0.3508$$

Lo que indica que la series coinciden en un 35.08% en su comportamiento en los mismos instantes.

Ventas vs Importaciones

Etiqueta_VentasComercio * Etiqueta_ImportacionesComercio Crosstabulation

Count		Etiqueta_ImportacionesComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	2	1	0	0	0	4
VentasComercio	2,00	0	7	1	2	2	1	13
	3,00	0	1	3	2	5	0	11
	4,00	1	3	1	0	1	1	7
	5,00	0	2	4	4	5	2	17
	6,00	0	2	0	0	0	3	5
Total		2	17	10	8	13	7	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+7+3+0+5+3}{57} = 0.3333$$

Lo que indica que la series coinciden en un 33.33% en su comportamiento en los mismos instantes.

Empleo vs Inversiones

Etiqueta_EmpleoComercio * Etiqueta_InversionesComercio Crosstabulation

Count		Etiqueta_InversionesComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	2	2	1	0	0	6
EmpleoComercio	2,00	0	6	4	2	4	0	16
	3,00	0	6	1	0	2	0	9
	4,00	0	2	3	0	0	0	5
	5,00	0	1	1	4	6	1	13
	6,00	0	2	0	0	3	3	8
Total		1	19	11	7	15	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+6+1+0+6+3}{57} = 0.2982$$

Lo que indica que la series coinciden en un 29.82% en su comportamiento en los mismos instantes.

Empleo vs Importaciones

Etiqueta_EmpleoComercio * Etiqueta_ImportacionesComercio Crosstabulation

Count		Etiqueta_ImportacionesComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	1	1	1	2	0	6
EmpleoComercio	2,00	0	5	4	2	5	0	16
	3,00	1	5	0	2	0	1	9
	4,00	0	2	2	0	1	0	5
	5,00	0	2	3	2	2	4	13
	6,00	0	2	0	1	3	2	8
Total		2	17	10	8	13	7	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{1+5+0+0+2+2}{57} = 0.1754$$

Lo que indica que la series coinciden en un 17.54% en su comportamiento en los mismos instantes.

Inversiones vs Importaciones

Etiqueta_InversionesComercio * Etiqueta_ImportacionesComercio Crosstabulation

Count		Etiqueta_ImportacionesComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	0	0	0	1
InversionesComercio	2,00	2	8	3	3	3	0	19
	3,00	0	2	4	2	3	0	11
	4,00	0	1	2	1	3	0	7
	5,00	0	4	1	2	3	5	15
	6,00	0	1	0	0	1	2	4
Total		2	17	10	8	13	7	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+8+4+1+3+2}{57} = 0.3157$$

Lo que indica que la series coinciden en un 31.57% en su comportamiento en los mismos instantes.

Resumen.

En la siguiente tabla se resume la información del sector Comercio, indicando la significación de cada una de las relaciones.

	<i>Ventas</i>	<i>Empleo</i>	<i>Inversiones</i>	<i>Importaciones</i>
<i>Ventas</i>		0.2982(0.0041)	0.3508(0.0002)	0.3333(0.0005)
<i>Empleo</i>	0.2982(0.0041)		0.2982(0.0041)	0.1754(0.3484)
<i>Inversiones</i>	0.3508(0.0002)	0.2982(0.0041)		0.3157(0.0016)
<i>Importaciones</i>	0.3333(0.0005)	0.1754(0.3484)	0.3157(0.0016)	

Hemos observado que hay relación entre todas las variables del sector comercio, excepto entre *Empleo* e *Importaciones*. Estas relaciones son muy significativas dado que el p-valor es menor que 0.01.

6.1.3. Sector Construcción

Ventas vs Empleo

Etiqueta_VentasConstruccion * Etiqueta_EmpleoConstruccion Crosstabulation

Count		Etiqueta_EmpleoConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasConstruccion	1,00	1	2	0	0	2	0	5
	2,00	1	3	5	1	5	2	17
	3,00	1	2	1	2	1	0	7
	4,00	0	4	1	1	2	0	8
	5,00	0	3	3	2	2	2	12
	6,00	0	2	1	0	4	1	8
Total		3	16	11	6	16	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+3+1+1+2+1}{57} = 0.1578$$

Lo que indica que la series coinciden en un 15.78% en su comportamiento en los mismos instantes.

Ventas vs Inversiones

Etiqueta_VentasConstruccion * Etiqueta_InversionesConstruccion Crosstabulation

Count		Etiqueta_InversionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	4	0	0	1	0	5
VentasConstruccion	2,00	2	5	3	2	4	1	17
	3,00	0	4	1	1	1	0	7
	4,00	0	3	0	2	3	0	8
	5,00	0	0	4	0	6	2	12
	6,00	0	0	4	0	3	1	8
Total		2	16	12	5	18	4	57

Por lo que la cota para el QSI es

$$cQSI = \frac{0+5+1+2+6+1}{57} = 0.2631$$

Lo que indica que la series coinciden en un 25.42% en su comportamiento en los mismos instantes.

Ventas vs Importaciones

Etiqueta_VentasConstruccion * Etiqueta_ImportacionesConstruccion Crosstabulation

Count		Etiqueta_ImportacionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	2	0	0	0	2	1	5
VentasConstruccion	2,00	5	3	3	2	1	3	17
	3,00	2	4	0	0	0	1	7
	4,00	2	1	1	1	1	2	8
	5,00	1	1	1	2	6	1	12
	6,00	0	2	0	0	3	3	8
Total		12	11	5	5	13	11	57

Por lo que la cota para el QSI es

$$cQSI = \frac{2+3+0+1+6+3}{57} = 0.2631$$

Lo que indica que la series coinciden en un 26.31% en su comportamiento en los mismos instantes.

Empleo vs Inversiones

Etiqueta_EmpleoConstruccion * Etiqueta_InversionesConstruccion Crosstabulation

Count		Etiqueta_InversionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	2	0	0	1	0	3
EmpleoConstruccion	2,00	1	7	4	2	2	0	16
	3,00	1	3	3	1	3	0	11
	4,00	0	0	2	2	1	1	6
	5,00	0	4	2	0	9	1	16
	6,00	0	0	1	0	2	2	5
Total		2	16	12	5	18	4	57

Por lo que la cota para el QSI es

$$c_{QSI} = \frac{0+7+3+2+9+2}{57} = 0.4035$$

Lo que indica que la series coinciden en un 40.35% en su comportamiento en los mismos instantes.

Empleo vs Importaciones

Etiqueta_EmpleoConstruccion * Etiqueta_ImportacionesConstruccion Crosstabulation

Count		Etiqueta_ImportacionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	0	0	2	3
EmpleoConstruccion	2,00	4	4	2	0	4	2	16
	3,00	1	2	1	1	2	4	11
	4,00	3	1	0	0	2	0	6
	5,00	4	1	2	3	3	3	16
	6,00	0	2	0	1	2	0	5
Total		12	11	5	5	13	11	57

Por lo que la cota para el QSI es

$$c_{QSI} = \frac{0+4+1+0+3+0}{57} = 0.1403$$

Lo que indica que la series coinciden en un 14.03% en su comportamiento en los mismos instantes.

Inversiones vs Importaciones

Etiqueta_InversionesConstruccion * Etiqueta_ImportacionesConstruccion Crosstabulation

Count		Etiqueta_ImportacionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	1	0	0	0	0	2
InversionesConstruccion	2,00	4	3	1	2	4	2	16
	3,00	3	3	2	1	1	2	12
	4,00	1	2	0	0	0	2	5
	5,00	3	1	2	2	6	4	18
	6,00	0	1	0	0	2	1	4
Total		12	11	5	5	13	11	57

Por lo que la cota para el QSI es

$$c_{QSI} = \frac{1+3+2+0+6+1}{57} = 0.2280$$

Lo que indica que la series coinciden en un 22.80% en su comportamiento en los mismos instantes.

Resumen.

En la siguiente tabla se resume la información del sector Construcción, indicando la significación de cada una de las relaciones.

	<i>Ventas</i>	<i>Empleo</i>	<i>Inversiones</i>	<i>Importaciones</i>
<i>Ventas</i>		0.1578(0.4840)	0.2631(0.0213)	0.2631(0.0213)
<i>Empleo</i>	0.1578(0.4840)		0.4035(0.0000)	0.1403(0.6252)
<i>Inversiones</i>	0.2631(0.0213)	0.4035(0.0000)		0.2280(0.0821)
<i>Importaciones</i>	0.2631(0.0213)	0.1403(0.6252)	0.2280(0.0821)	

Podemos observar que existen relaciones significativas a distintos niveles. En global podemos decir que hay menos relación que entre las variables del sector comercio pero más que entre las variables del sector servicio.

6.1.4. Sector Industria.

Ventas vs Empleo

Etiqueta_VentasIndustria * Etiqueta_EmpleoIndustria Crosstabulation

Count		Etiqueta_EmpleoIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasIndustria	1,00	0	2	0	2	0	0	4
	2,00	1	4	3	1	4	0	13
	3,00	0	5	1	3	2	0	11
	4,00	0	5	0	0	1	0	6
	5,00	0	2	3	3	7	4	19
	6,00	0	1	0	0	2	1	4
Total		1	19	7	9	16	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+4+1+0+7+1}{57} = 0.228$$

Lo que indica que la series coinciden en un 22.80% en su comportamiento en los mismos instantes.

Ventas vs Inversiones

Etiqueta_VentasIndustria * Etiqueta_InversionesIndustria Crosstabulation

Count		Etiqueta_InversionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasIndustria	1,00	1	2	0	1	0	0	4
	2,00	1	6	3	0	3	0	13
	3,00	0	2	3	2	3	1	11
	4,00	1	2	0	3	0	0	6
	5,00	1	4	0	3	8	3	19
	6,00	0	1	0	1	0	2	4
Total		4	17	6	10	14	6	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+6+3+3+8+2}{57} = 0.4035$$

Lo que indica que la series coinciden en un 40.35% en su comportamiento en los mismos instantes.

Ventas vs Importaciones

Etiqueta_VentasIndustria * Etiqueta_ImportacionesIndustria Crosstabulation

Count		Etiqueta_ImportacionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasIndustria	1,00	1	2	0	1	0	0	4
	2,00	0	3	3	4	3	0	13
	3,00	1	4	3	1	2	0	11
	4,00	0	1	2	1	2	0	6
	5,00	0	2	4	4	7	2	19
	6,00	0	0	0	0	3	1	4
Total		2	12	12	11	17	3	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{1+3+3+1+7+1}{57} = 0.2807$$

Lo que indica que la series coinciden en un 28.07% en su comportamiento en los mismos instantes.

Empleo vs Inversiones

Etiqueta_EmpleoIndustria * Etiqueta_InversionesIndustria Crosstabulation

Count		Etiqueta_InversionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_EmpleoIndustria	1,00	0	1	0	0	0	0	1
	2,00	3	7	1	4	2	2	19
	3,00	0	2	2	1	2	0	7
	4,00	1	4	0	2	1	1	9
	5,00	0	2	3	3	6	2	16
	6,00	0	1	0	0	3	1	5
Total		4	17	6	10	14	6	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+7+2+2+6+1}{57} = 0.3157$$

Lo que indica que la series coinciden en un 31.57% en su comportamiento en los mismos instantes.

Empleo vs Importaciones

Etiqueta_EmpleoIndustria * Etiqueta_ImportacionesIndustria Crosstabulation

Count		Etiqueta_ImportacionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	0	1	0	0	0	1
EmpleoIndustria	2,00	1	6	3	3	6	0	19
	3,00	1	2	0	3	1	0	7
	4,00	0	2	4	1	2	0	9
	5,00	0	2	4	4	5	1	16
	6,00	0	0	0	0	3	2	5
Total		2	12	12	11	17	3	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+6+0+1+5+2}{57} = 0.2456$$

Lo que indica que la series coinciden en un 24.56% en su comportamiento en los mismos instantes.

Inversiones vs Importaciones

Etiqueta_InversionesIndustria * Etiqueta_ImportacionesIndustria Crosstabulation

Count		Etiqueta_ImportacionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	2	1	0	0	0	4
InversionesIndustria	2,00	0	4	4	2	7	0	17
	3,00	1	2	1	2	0	0	6
	4,00	0	1	3	3	3	0	10
	5,00	0	3	2	4	4	1	14
	6,00	0	0	1	0	3	2	6
Total		2	12	12	11	17	3	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{1+4+1+3+4+2}{57} = 0.2631$$

Lo que indica que la series coinciden en un 26.31% en su comportamiento en los mismos instantes.

Resumen

En la tabla siguiente se resume la información del sector Industria, indicando la significación de cada una de las relaciones.

	Ventas	Empleo	Inversiones	Importaciones
Ventas		0.228 (0.0821)	0.4035(0.0000)	0.2807(0.0097)
Empleo	0.228(0.0821)		0.3157(0.0016)	0.2456(0.0435)
Inversiones	0.4035(0.0000)	0.3157(0.0016)		0.2631(0.0213)
Importaciones	0.2807(0.0097)	0.2456(0.0435)	0.2631(0.0213)	

De la tabla podemos observar que todas las relaciones son significativas a un nivel 0.1, por lo que podemos concluir que existe relación entre todas las series temporales.

6.1.5 .Datos Globales.

Ventas vs Empleo

Etiqueta_VentasGlobal * Etiqueta_EmpleoGlobal Crosstabulation

Count		Etiqueta_EmpleoGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasGlobal	1,00	0	2	3	0	0	0	5
	2,00	0	7	4	1	0	0	12
	3,00	0	0	0	2	4	2	8
	4,00	0	5	0	1	4	0	10
	5,00	1	1	3	2	10	0	17
	6,00	0	1	1	1	1	1	5
Total		1	16	11	7	19	3	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+7+0+1+10+1}{57} = 0.3333$$

Lo que indica que la series coinciden en un 33.33% en su comportamiento en los mismos instantes.

Ventas vs Inversiones

Etiqueta_VentasGlobal * Etiqueta_InversionesGlobal Crosstabulation

Count		Etiqueta_InversionesGlobal					Total
		2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasGlobal	1,00	3	1	1	0	0	5
	2,00	7	2	1	2	0	12
	3,00	2	1	3	2	0	8
	4,00	4	1	4	1	0	10
	5,00	3	3	1	10	0	17
	6,00	0	1	1	2	1	5
Total		19	9	11	17	1	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{7+1+4+10+1}{57} = 0.4035$$

Lo que indica que la series coinciden en un 40.35% en su comportamiento en los mismos instantes.

Ventas vs Importaciones

Etiqueta_VentasGlobal * Etiqueta_ImportacionesGlobal Crosstabulation

Count		Etiqueta_ImportacionesGlobal					Total	
		1,00	2,00	3,00	4,00	5,00		6,00
Etiqueta_VentasGlobal	1,00	1	4	0	0	0	0	5
	2,00	1	5	2	2	2	0	12
	3,00	0	2	2	1	3	0	8
	4,00	0	3	2	1	4	0	10
	5,00	0	1	4	2	8	2	17
	6,00	0	0	1	0	2	2	5
Total		2	15	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+5+2+1+8+2}{57} = 0.3333$$

Lo que indica que la series coinciden en un 33.33% en su comportamiento en los mismos instantes.

Empleo con Inversiones

Etiqueta_EmpleoGlobal * Etiqueta_InversionesGlobal Crosstabulation

Count		Etiqueta_InversionesGlobal					Total
		2,00	3,00	4,00	5,00	6,00	
Etiqueta_EmpleoGlobal	1,00	0	1	0	0	0	1
	2,00	10	2	1	3	0	16
	3,00	4	1	3	3	0	11
	4,00	2	2	2	1	0	7
	5,00	3	2	5	9	0	19
	6,00	0	1	0	1	1	3
Total		19	9	11	17	1	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{10+1+2+9+1}{57} = 0.4035$$

Lo que indica que la series coinciden en un 40.35% en su comportamiento en los mismos instantes.

Empleo con Importaciones

Etiqueta_EmpleoGlobal * Etiqueta_ImportacionesGlobal Crosstabulation

Count		Etiqueta_ImportacionesGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_EmpleoGlobal	1,00	0	1	0	0	0	0	1
	2,00	1	7	2	1	4	1	16
	3,00	1	4	2	1	2	1	11
	4,00	0	1	3	2	1	0	7
	5,00	0	2	4	1	11	1	19
	6,00	0	0	0	1	1	1	3
Total		2	15	11	6	19	4	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+7+2+2+11+1}{57} = 0.4035$$

Lo que indica que la series coinciden en un 40.35% en su comportamiento en los mismos instantes.

Inversiones con Importaciones

Etiqueta_InversionesGlobal * Etiqueta_ImportacionesGlobal Crosstabulation

Count		Etiqueta_ImportacionesGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	2,00	1	9	5	1	3	0	19
InversionesGlobal	3,00	0	2	1	4	2	0	9
	4,00	1	3	3	0	3	1	11
	5,00	0	1	2	1	11	2	17
	6,00	0	0	0	0	0	1	1
Total		2	15	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{9+1+0+11+1}{57} = 0.3859$$

Lo que indica que la series coinciden en un 38.59% en su comportamiento en los mismos instantes.

Resumen

En la tabla siguiente se resume la información del sector Industria, indicando la significación de cada una de las relaciones

	<i>Ventas</i>	<i>Empleo</i>	<i>Inversiones</i>	<i>Importaciones</i>
<i>Ventas</i>		0.3333(0.0005)	0.4035(0.0000)	0.3333(0.0005)
<i>Empleo</i>	0.3333(0.0005)		0.4035(0.0000)	0.4035(0.0000)
<i>Inversiones</i>	0.4035(0.0000)	0.4035(0.0000)		0.3859(0.0000)
<i>Importaciones</i>	0.3333(0.0005)	0.4035(0.0000)	0.3859(0.0000)	

Podemos observar que todas las relaciones de las variables globales son muy significativas, con p-valores por debajo de 0.0005. Se debe destacar que existe una mayor relación entre las variables a nivel global de la existe entre las variables sector por sector.

6.2. Comparación por Variables.

En esta sección compararemos las variables de cada sector.

6.2.1. Variable *Ventas*.

Servicios vs Comercio

Etiqueta_VentasServicios * Etiqueta_VentasComercio Crosstabulation

Count		Etiqueta_VentasComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasServicios	1,00	3	2	0	2	1	2	10
	2,00	1	1	2	2	5	0	11
	3,00	0	4	0	1	4	0	9
	4,00	0	1	0	1	1	0	3
	5,00	0	2	7	0	1	1	11
	6,00	0	3	2	1	5	2	13
Total		4	13	11	7	17	5	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{3+1+0+1+1+2}{57} = 0.1403$$

Lo que indica que la series coinciden en un 14.03% en su comportamiento en los mismos instantes. En otras palabras las Ventas del sector Servicios con las Ventas del sector comercio no fueron muy similares en el periodo de estudio.

Servicios vs Construcción

Etiqueta_VentasServicios * Etiqueta_VentasConstruccion Crosstabulation

Count		Etiqueta_VentasConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasServicios	1,00	2	2	0	1	4	1	10
	2,00	1	7	1	0	1	1	11
	3,00	1	2	2	1	1	2	9
	4,00	0	1	0	1	1	0	3
	5,00	0	2	2	4	2	1	11
	6,00	1	3	2	1	3	3	13
Total		5	17	7	8	12	8	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{2+7+2+1+2+3}{57} = 0.2982$$

Lo que indica que la series coinciden en un 29.82% en su comportamiento en los mismos instantes.

Servicios vs Industria

Etiqueta_VentasServicios * Etiqueta_VentasIndustria Crosstabulation

Count		Etiqueta_VentasIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasServicios	1,00	2	4	0	2	2	0	10
	2,00	1	1	1	2	5	1	11
	3,00	0	2	3	0	4	0	9
	4,00	0	0	2	0	1	0	3
	5,00	0	4	2	1	3	1	11
	6,00	1	2	3	1	4	2	13
Total		4	13	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{2+1+3+0+3+2}{57} = 0.1929$$

Lo que indica que la series coinciden en un 19.29% en su comportamiento en los mismos instantes.

Servicios vs Global

Etiqueta_VentasServicios * Etiqueta_VentasGlobal Crosstabulation

Count		Etiqueta_VentasGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasServicios	1,00	3	4	2	0	1	0	10
	2,00	1	1	2	2	5	0	11
	3,00	1	4	1	1	2	0	9
	4,00	0	1	0	1	1	0	3
	5,00	0	2	2	3	4	0	11
	6,00	0	0	1	3	4	5	13
Total		5	12	8	10	17	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{3+1+1+1+4+5}{57} = 0.2631$$

Lo que indica que la series coinciden en un 26.31% en su comportamiento en los mismos instantes.

Comercio vs Construcción

Etiqueta_VentasComercio * Etiqueta_VentasConstruccion Crosstabulation

Count		Etiqueta_VentasConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasComercio	1,00	1	1	1	0	1	0	4
	2,00	1	5	2	1	3	1	13
	3,00	1	3	2	3	1	1	11
	4,00	1	3	0	1	1	1	7
	5,00	1	5	2	3	3	3	17
	6,00	0	0	0	0	3	2	5
Total		5	17	7	8	12	8	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+5+2+1+3+2}{57} = 0.2456$$

Lo que indica que la series coinciden en un 24.56% en su comportamiento en los mismos instantes.

Comercio vs Industria

Etiqueta_VentasComercio * Etiqueta_VentasIndustria Crosstabulation

Count		Etiqueta_VentasIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasComercio	1,00	1	2	0	1	0	0	4
	2,00	1	6	3	1	2	0	13
	3,00	0	2	4	1	4	0	11
	4,00	0	1	1	1	4	0	7
	5,00	2	1	3	2	6	3	17
	6,00	0	1	0	0	3	1	5
Total		4	13	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+6+4+1+6+1}{57} = 0.3333$$

Lo que indica que la series coinciden en un 33.33% en su comportamiento en los mismos instantes.

Comercio vs Global

Etiqueta_VentasComercio * Etiqueta_VentasGlobal Crosstabulation

Count		Etiqueta_VentasGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	3	1	0	0	0	0	4
VentasComercio	2,00	1	8	2	1	1	0	13
	3,00	0	0	3	4	4	0	11
	4,00	0	2	2	1	2	0	7
	5,00	1	1	0	4	8	3	17
	6,00	0	0	1	0	2	2	5
Total		5	12	8	10	17	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{3+8+3+1+8+2}{57} = 0.4385$$

Lo que indica que la series coinciden en un 43.85% en su comportamiento en los mismos instantes.

Construcción vs Industria

Etiqueta_VentasConstruccion * Etiqueta_VentasIndustria Crosstabulation

Count		Etiqueta_VentasIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	2	1	1	1	0	5
VentasConstruccion	2,00	2	4	2	2	7	0	17
	3,00	1	2	2	0	1	1	7
	4,00	0	2	2	0	3	1	8
	5,00	1	2	2	2	5	0	12
	6,00	0	1	2	1	2	2	8
Total		4	13	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+4+2+0+5+2}{57} = 0.2280$$

Lo que indica que la series coinciden en un 22.80% en su comportamiento en los mismos instantes.

Construcción vs Global

Etiqueta_VentasConstruccion * Etiqueta_VentasGlobal Crosstabulation

Count		Etiqueta_VentasGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasConstruccion	1,00	0	3	1	0	1	0	5
	2,00	3	2	3	4	5	0	17
	3,00	1	2	1	1	0	2	7
	4,00	0	1	2	1	3	1	8
	5,00	1	2	1	2	6	0	12
	6,00	0	2	0	2	2	2	8
Total		5	12	8	10	17	5	57

Por lo que la cota para el QSI es:

$$QSI = \frac{0+2+1+1+6+2}{57} = 0.2105$$

Lo que indica que la series coinciden en un 21.05% en su comportamiento en los mismos instantes.

Industria vs Global

Etiqueta_VentasIndustria * Etiqueta_VentasGlobal Crosstabulation

Count		Etiqueta_VentasGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_VentasIndustria	1,00	2	1	0	1	0	0	4
	2,00	2	6	3	1	0	1	13
	3,00	0	3	2	4	2	0	11
	4,00	0	2	0	3	1	0	6
	5,00	1	0	3	1	12	2	19
	6,00	0	0	0	0	2	2	4
Total		5	12	8	10	17	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{2+6+2+3+12+2}{57} = 0.4736$$

Lo que indica que la series coinciden en un 47.36% en su comportamiento en los mismos instantes.

Resumen

En la tabla siguiente se resume la información de la variable *Ventas*, indicando la significación de cada una de las relaciones

	Servicios	Comercio	Construcción	Industria	Global
Servicios		0.1403(0.6252)	0.2982(0.0041)	0.1929(0.2325)	0.2631(0.0213)
Comercio	0.1403(0.6252)		0.2456(0.0435)	0.3333(0.0005)	0.4385(0.0000)
Construcción	0.2982(0.0041)	0.2456(0.0435)		0.2280(0.0821)	0.2105(0.1436)
Industria	0.1929(0.2325)	0.3333(0.0005)	0.2280(0.0821)		0.4736(0.0000)
Global	0.2631(0.0213)	0.4385(0.0000)	0.2105(0.1436)	0.4736(0.0000)	

Las relaciones de la variable *Ventas* que podemos asegurar que son significativas con distintos niveles de significación son todas menos: Servicios-Comercio, Servicios-Industria, Construcción-Global. Llamativo es esta última relación, dado que las ventas en el sector construcción formar parte de las ventas globales (seguramente esta no significación se deba a que el valor de n no es alto).

6.2.2. Variable *Empleo*.

Servicios vs Comercio

Etiqueta_EmpleoServicios * Etiqueta_EmpleoComercio Crosstabulation

Count		Etiqueta_EmpleoComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	2	2	0	0	2	2	8
EmpleoServicios	2,00	2	5	1	1	3	0	12
	3,00	0	3	1	1	1	1	7
	4,00	1	2	4	1	2	2	12
	5,00	0	3	0	1	4	1	9
	6,00	1	1	3	1	1	2	9
Total		6	16	9	5	13	8	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{2+5+1+1+4+2}{57} = 0.2631$$

Lo que indica que la series coinciden en un 26.31% en su comportamiento en los mismos instantes.

Servicios vs Construcción

Etiqueta_EmpleoServicios * Etiqueta_EmpleoConstruccion Crosstabulation

Count		Etiqueta_EmpleoConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	3	4	1	0	0	8
EmpleoServicios	2,00	1	4	2	1	4	0	12
	3,00	0	1	2	1	1	2	7
	4,00	1	4	1	3	3	0	12
	5,00	1	0	1	0	6	1	9
	6,00	0	4	1	0	2	2	9
Total		3	16	11	6	16	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+4+2+3+6+2}{57} = 0.2982$$

Lo que indica que la series coinciden en un 29.82% en su comportamiento en los mismos instantes.

Servicios vs Industria

Etiqueta_EmpleoServicios * Etiqueta_EmpleoIndustria Crosstabulation

Count		Etiqueta_EmpleoIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	2	1	0	4	1	8
EmpleoServicios	2,00	1	3	0	3	4	1	12
	3,00	0	3	2	1	1	0	7
	4,00	0	6	2	1	2	1	12
	5,00	0	0	1	2	4	2	9
	6,00	0	5	1	2	1	0	9
Total		1	19	7	9	16	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+3+2+1+4+0}{57} = 0.1754$$

Lo que indica que la series coinciden en un 17.54% en su comportamiento en los mismos instantes.

Servicios vs Global

Etiqueta_EmpleoServicios * Etiqueta_EmpleoGlobal Crosstabulation

Count		Etiqueta_EmpleoGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	5	1	0	1	0	8
EmpleoServicios	2,00	0	7	2	2	1	0	12
	3,00	0	1	4	1	1	0	7
	4,00	0	3	4	3	2	0	12
	5,00	0	0	0	0	8	1	9
	6,00	0	0	0	1	6	2	9
Total		1	16	11	7	19	3	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+7+4+3+8+2}{57} = 0.4385$$

Lo que indica que la series coinciden en un 43.85% en su comportamiento en los mismos instantes

Comercio vs Construcción

Etiqueta_EmpleoComercio * Etiqueta_EmpleoConstruccion Crosstabulation

Count		Etiqueta_EmpleoConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	3	1	0	1	0	6
EmpleoComercio	2,00	1	3	5	1	3	3	16
	3,00	0	5	0	2	1	1	9
	4,00	0	0	2	1	2	0	5
	5,00	1	4	2	0	5	1	13
	6,00	0	1	1	2	4	0	8
Total		3	16	11	6	16	5	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+3+0+1+5+0}{57} = 0.1754$$

Lo que indica que la series coinciden en un 17.54% en su comportamiento en los mismos instantes

Comercio vs Industria

Etiqueta_EmpleoComercio * Etiqueta_EmpleoIndustria Crosstabulation

Count		Etiqueta_EmpleoIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	2	0	2	2	0	6
EmpleoComercio	2,00	1	5	4	2	3	1	16
	3,00	0	6	0	0	3	0	9
	4,00	0	1	2	1	1	0	5
	5,00	0	2	1	4	5	1	13
	6,00	0	3	0	0	2	3	8
Total		1	19	7	9	16	5	57

Por lo que la cota para el QSI es:

$$QSI = \frac{0+5+0+1+5+3}{57} = 0.2456$$

Lo que indica que la series coinciden en un 24.56% en su comportamiento en los mismos instantes.

Comercio vs Global

Etiqueta_EmpleoComercio * Etiqueta_EmpleoGlobal Crosstabulation

Count		Etiqueta_EmpleoGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	5	0	1	0	0	6
EmpleoComercio	2,00	1	8	3	0	3	1	16
	3,00	0	1	4	1	3	0	9
	4,00	0	0	2	1	2	0	5
	5,00	0	2	1	3	7	0	13
	6,00	0	0	1	1	4	2	8
Total		1	16	11	7	19	3	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+8+4+1+7+2}{57} = 0.3859$$

Lo que indica que la series coinciden en un 38.59% en su comportamiento en los mismos instante.

Construcción vs Industria

Etiqueta_EmpleoConstruccion * Etiqueta_EmpleoIndustria Crosstabulation

Count		Etiqueta_EmpleoIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	2	0	1	0	0	3
EmpleoConstruccion	2,00	0	6	1	2	7	0	16
	3,00	1	3	2	1	3	1	11
	4,00	0	2	1	1	1	1	6
	5,00	0	4	1	3	5	3	16
	6,00	0	2	2	1	0	0	5
Total		1	19	7	9	16	5	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+6+2+1+5+0}{57} = 0.2456$$

Lo que indica que la series coinciden en un 24.56% en su comportamiento en los mismos instante.

Construcción vs Global

Etiqueta_EmpleoConstruccion * Etiqueta_EmpleoGlobal Crosstabulation

Count		Etiqueta_EmpleoGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	2	0	0	1	0	3
EmpleoConstruccion	2,00	0	7	1	5	3	0	16
	3,00	1	4	4	0	2	0	11
	4,00	0	1	2	1	2	0	6
	5,00	0	2	3	1	8	2	16
	6,00	0	0	1	0	3	1	5
Total		1	16	11	7	19	3	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+7+4+1+8+1}{57} = 0.3684$$

Lo que indica que la series coinciden en un 36.84% en su comportamiento en los mismos instante.

Industria vs Global

Etiqueta_EmpleoIndustria * Etiqueta_EmpleoGlobal Crosstabulation

Count		Etiqueta_EmpleoGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	0	0	0	1
EmpleoIndustria	2,00	0	8	5	1	4	1	19
	3,00	1	1	1	2	1	1	7
	4,00	0	2	2	1	4	0	9
	5,00	0	4	3	3	6	0	16
	6,00	0	0	0	0	4	1	5
Total		1	16	11	7	19	3	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+8+1+1+6+1}{57} = 0.2982$$

Lo que indica que la series coinciden en un 29.82% en su comportamiento en los mismos instante.

Resumen

En la tabla siguiente se resume la información de la variable *Empleo*, indicando la significación de cada una de las relaciones

	Servicios	Comercio	Construcción	Industria	Global
Servicios		0.2631(0.0213)	0.2982(0.0041)	0.1754(0.3484)	0.4385(0.0000)
Comercio	0.2631(0.0213)		0.1754(0.3484)	0.2456(0.0435)	0.3859(0.0000)
Construcción	0.2982(0.0041)	0.1754(0.3484)		0.2456(0.0435)	0.3684(0.0000)
Industria	0.1754(0.3484)	0.2456(0.0435)	0.2456(0.0435)		0.2982(0.0041)
Global	0.4385(0.0000)	0.3859(0.0000)	0.3684(0.0000)	0.2982(0.0041)	

Podemos observar que la mayoría son significativas a nivel 0.05. Tan solo Servicios-Industria y Comercio-Construcción no lo son. Ahora si, todas las variables tienen una significación con la variable global como era de esperar.

6.2.3. Variable *Inversiones*

Servicios vs Comercio

Etiqueta_InversionesServicios * Etiqueta_InversionesComercio Crosstabulation

Count		Etiqueta_InversionesComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_InversionesServicios	1,00	0	1	1	0	1	0	3
	2,00	1	6	3	4	3	1	18
	3,00	0	6	1	0	1	0	8
	4,00	0	3	4	2	1	0	10
	5,00	0	1	1	1	6	2	11
	6,00	0	2	1	0	3	1	7
Total		1	19	11	7	15	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+6+1+2+6+1}{57} = 0.2807$$

Lo que indica que la series coinciden en un 28.07% en su comportamiento en los mismos instante.

Servicios vs Inversiones

Etiqueta_InversionesServicios * Etiqueta_InversionesConstruccion Crosstabulation

Count		Etiqueta_InversionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	0	2	0	3
InversionesServicios	2,00	0	7	3	0	5	3	18
	3,00	0	3	1	2	2	0	8
	4,00	0	1	4	2	3	0	10
	5,00	2	2	3	1	3	0	11
	6,00	0	2	1	0	3	1	7
Total		2	16	12	5	18	4	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+7+1+2+3+1}{57} = 0.2456$$

Lo que indica que la series coinciden en un 24.56% en su comportamiento en los mismos instante.

Servicios vs Industria

Etiqueta_InversionesServicios * Etiqueta_InversionesIndustria Crosstabulation

Count		Etiqueta_InversionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	2	0	0	0	0	3
InversionesServicios	2,00	2	7	2	3	3	1	18
	3,00	0	4	1	2	1	0	8
	4,00	0	1	1	3	5	0	10
	5,00	1	1	2	2	2	3	11
	6,00	0	2	0	0	3	2	7
Total		4	17	6	10	14	6	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+7+1+3+2+2}{57} = 0.2807$$

Lo que indica que la series coinciden en un 28.07% en su comportamiento en los mismos instante.

Servicios vs Global

Etiqueta_InversionesServicios * Etiqueta_InversionesGlobal Crosstabulation

Count		Etiqueta_InversionesGlobal					Total
		2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	3	0	0	0	0	3
InversionesServicios	2,00	10	5	2	1	0	18
	3,00	5	1	1	1	0	8
	4,00	1	1	7	1	0	10
	5,00	0	2	1	8	0	11
	6,00	0	0	0	6	1	7
Total		19	9	11	17	1	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{10+1+7+8+1}{57} = 0.4736$$

Lo que indica que la series coinciden en un 47.35% en su comportamiento en los mismos instante.

Comercio vs Construcción

Etiqueta_InversionesComercio * Etiqueta_InversionesConstruccion Crosstabulation

Count		Etiqueta_InversionesConstruccion					Total
		1,00	2,00	3,00	4,00	5,00	
Etiqueta_	1,00	0	1	0	0	0	1
InversionesComercio	2,00	0	6	4	2	7	19
	3,00	0	3	2	2	4	11
	4,00	0	1	1	0	4	7
	5,00	1	5	5	0	3	15
	6,00	1	0	0	1	0	4
Total		2	16	12	5	18	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+6+2+0+3+2}{57} = 0.2280$$

Lo que indica que la series coinciden en un 22.80% en su comportamiento en los mismos instante.

Comercio vs Industria

Etiqueta_InversionesComercio * Etiqueta_InversionesIndustria Crosstabulation

Count		Etiqueta_InversionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	0	0	1	0	0	1
InversionesComercio	2,00	1	8	3	2	3	2	19
	3,00	0	4	2	2	2	1	11
	4,00	0	2	0	2	2	1	7
	5,00	3	3	0	3	4	2	15
	6,00	0	0	1	0	3	0	4
Total		4	17	6	10	14	6	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+8+2+2+4+0}{57} = 0.2807$$

Lo que indica que la series coinciden en un 28.07% en su comportamiento en los mismos instante.

Comercio vs Global

Etiqueta_InversionesComercio * Etiqueta_InversionesGlobal Crosstabulation

Count		Etiqueta_InversionesGlobal					Total
		2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	0	0	0	0	1
InversionesComercio	2,00	10	6	1	2	0	19
	3,00	4	0	6	1	0	11
	4,00	2	2	2	1	0	7
	5,00	2	1	2	10	0	15
	6,00	0	0	0	3	1	4
Total		19	9	11	17	1	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{10+0+2+10+1}{57} = 0.4035$$

Lo que indica que la series coinciden en un 40.35% en su comportamiento en los mismos instante.

Construcción vs Industria

Etiqueta_InversionesConstruccion * Etiqueta_InversionesIndustria Crosstabulation

Count		Etiqueta_InversionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	0	0	0	1	0	2
InversionesConstruccion	2,00	0	7	0	4	3	2	16
	3,00	1	1	1	3	4	2	12
	4,00	0	1	1	1	2	0	5
	5,00	2	6	4	2	2	2	18
	6,00	0	2	0	0	2	0	4
Total		4	17	6	10	14	6	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+7+1+1+2+0}{57} = 0.2105$$

Lo que indica que la series coinciden en un 21.05% en su comportamiento en los mismos instante.

Construcción vs Global

Etiqueta_InversionesConstruccion * Etiqueta_InversionesGlobal Crosstabulation

Count		Etiqueta_InversionesGlobal					Total
		2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	1	0	2
InversionesConstruccion	2,00	8	2	1	5	0	16
	3,00	2	3	2	5	0	12
	4,00	1	0	3	1	0	5
	5,00	8	2	4	4	0	18
	6,00	0	1	1	1	1	4
Total		19	9	11	17	1	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{8+3+3+4+1}{57} = 0.3333$$

Lo que indica que la series coinciden en un 33.33% en su comportamiento en los mismos instante.

Industria vs Global

Etiqueta_InversionesIndustria * Etiqueta_InversionesGlobal Crosstabulation

Count		Etiqueta_InversionesGlobal					Total
		2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	3	1	0	0	0	4
InversionesIndustria	2,00	12	1	2	2	0	17
	3,00	1	3	1	1	0	6
	4,00	3	1	4	2	0	10
	5,00	0	2	4	7	1	14
	6,00	0	1	0	5	0	6
Total		19	9	11	17	1	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{12+3+4+7+0}{57} = 0.4561$$

Lo que indica que la series coinciden en un 45.61% en su comportamiento en los mismos instante.

Resumen

En la tabla siguiente se resume la información de la variable *Inversiones*, indicando la significación de cada una de las relaciones

	Servicios	Comercio	Construcción	Industria	Global
Servicios		0.2807(0.0097)	0.2456(0.0435)	0.2807(0.0097)	0.4736(0.0000)
Comercio	0.2807(0.0097)		0.2280(0.0821)	0.2807(0.0097)	0.4035(0.0000)
Construcción	0.2456(0.0435)	0.2280(0.0821)		0.2105(0.1436)	0.3333(0.0005)
Industria	0.2807(0.0097)	0.2807(0.0097)	0.2105(0.1436)		0.4561(0.0000)
Global	0.4736(0.0000)	0.4035(0.0000)	0.3333(0.0005)	0.4561(0.0000)	

De todas las relaciones son significativas a distintos niveles menos Industria-Construcción. Nuevamente todas las variables tienen una muy alta significación con la variable global como era de esperar.

6.2.4. Variable *Importaciones*.

Servicios vs Comercio

Etiqueta_ImportacionesServicios * Etiqueta_ImportacionesComercio Crosstabulation

Count		Etiqueta_ImportacionesComercio						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	3	1	0	0	5
ImportacionesServ	2,00	0	8	3	2	3	3	19
	3,00	0	1	1	1	1	0	4
	4,00	0	0	0	0	5	1	6
	5,00	1	3	3	4	2	2	15
	6,00	1	4	0	0	2	1	8
Total		2	17	10	8	13	7	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+8+1+0+2+1}{57} = 0.2105$$

Lo que indica que las series coinciden en un 21.05% en su comportamiento en los mismos instantes.

Servicios vs Construcción

Etiqueta_ImportacionesServicios * Etiqueta_ImportacionesConstruccion Crosstabulation

Count		Etiqueta_ImportacionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	0	2	2	5
ImportacionesServ	2,00	3	3	4	2	3	4	19
	3,00	2	1	1	0	0	0	4
	4,00	0	2	0	1	1	2	6
	5,00	5	2	0	1	5	2	15
	6,00	2	2	0	1	2	1	8
Total		12	11	5	5	13	11	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+3+1+1+5+1}{57} = 0.1929$$

Lo que indica que la series coinciden en un 19.29% en su comportamiento en los mismos instante.

Servicios vs Industria

Etiqueta_ImportacionesServicios * Etiqueta_ImportacionesIndustria Crosstabulation

Count		Etiqueta_ImportacionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	2	1	1	0	5
ImportacionesServicios	2,00	1	3	4	5	4	2	19
	3,00	0	2	2	0	0	0	4
	4,00	0	1	1	0	4	0	6
	5,00	1	5	0	3	5	1	15
	6,00	0	0	3	2	3	0	8
Total		2	12	12	11	17	3	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+3+2+0+5+0}{57} = 0.1754$$

Lo que indica que la series coinciden en un 17.54% en su comportamiento en los mismos instante.

Servicios vs Global

Etiqueta_ImportacionesServicios * Etiqueta_ImportacionesGlobal Crosstabulation

Count		Etiqueta_ImportacionesGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	2	3	0	0	0	0	5
ImportacionesServicios	2,00	0	7	8	2	2	0	19
	3,00	0	3	1	0	0	0	4
	4,00	0	0	0	1	5	0	6
	5,00	0	1	2	3	7	2	15
	6,00	0	1	0	0	5	2	8
Total		2	15	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{2+7+1+1+7+2}{57} = 0.3508$$

Lo que indica que la series coinciden en un 35.08% en su comportamiento en los mismos instante.

Comercio vs Construcción

Etiqueta_ImportacionesComercio * Etiqueta_ImportacionesConstruccion Crosstabulation

Count		Etiqueta_ImportacionesConstruccion						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	1	0	0	0	0	2
ImportacionesComercio	2,00	3	2	3	3	3	3	17
	3,00	2	2	0	0	4	2	10
	4,00	3	1	1	0	2	1	8
	5,00	2	2	1	2	3	3	13
	6,00	1	3	0	0	1	2	7
Total		12	11	5	5	13	11	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{1+2+0+0+3+2}{57} = 0.1403$$

Lo que indica que la series coinciden en un 14.03% en su comportamiento en los mismos instante.

Comercio vs Industria

Etiqueta_ImportacionesComercio * Etiqueta_ImportacionesIndustria Crosstabulation

Count		Etiqueta_ImportacionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	1	0	0	2
ImportacionesComercio	2,00	1	2	2	6	6	0	17
	3,00	1	2	1	2	3	1	10
	4,00	0	2	3	1	2	0	8
	5,00	0	4	5	1	3	0	13
	6,00	0	1	1	0	3	2	7
Total		2	12	12	11	17	3	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{0+2+1+1+3+2}{57} = 0.1578$$

Lo que indica que la series coinciden en un 15.78% en su comportamiento en los mismos instante.

Comercio vs Global

Etiqueta_VentasComercio * Etiqueta_ImportacionesGlobal Crosstabulation

Count		Etiqueta_ImportacionesGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	3	0	0	0	0	4
VentasComercio	2,00	1	5	3	2	2	0	13
	3,00	0	2	3	3	3	0	11
	4,00	0	1	1	0	5	0	7
	5,00	0	4	4	0	7	2	17
	6,00	0	0	0	1	2	2	5
Total		2	15	11	6	19	4	57

Por lo que la cota para el QSI es:

$$c_{QSI} = \frac{1+5+3+0+7+2}{57} = 0.3157$$

Lo que indica que la series coinciden en un 31.57% en su comportamiento en los mismos instante.

Construcción vs Industria

Etiqueta_ImportacionesConstruccion * Etiqueta_ImportacionesIndustria Crosstabulation

Count		Etiqueta_ImportacionesIndustria						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	1	2	2	3	3	1	12
Importaciones	2,00	0	6	1	1	3	0	11
Construccion	3,00	0	0	3	1	1	0	5
	4,00	0	1	0	3	1	0	5
	5,00	0	1	4	2	6	0	13
	6,00	1	2	2	1	3	2	11
Total		2	12	12	11	17	3	57

Por lo que la cota para el QSI es:

$$QSI = \frac{1+6+3+3+6+2}{57} = 0.3684$$

Lo que indica que la series coinciden en un 36.84% en su comportamiento en los mismos instante.

Construcción vs Global

Etiqueta_ImportacionesConstruccion * Etiqueta_ImportacionesGlobal Crosstabulation

Count		Etiqueta_ImportacionesGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	3	2	2	4	1	12
Importaciones	2,00	1	3	2	2	3	0	11
Construccion	3,00	0	3	2	0	0	0	5
	4,00	0	1	1	0	3	0	5
	5,00	1	2	2	1	6	1	13
	6,00	0	3	2	1	3	2	11
Total		2	15	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+3+2+0+6+2}{57} = 0.2280$$

Lo que indica que la series coinciden en un 22.80% en su comportamiento en los mismos instante.

Industria vs Global

Etiqueta_ImportacionesIndustria * Etiqueta_ImportacionesGlobal Crosstabulation

Count		Etiqueta_ImportacionesGlobal						Total
		1,00	2,00	3,00	4,00	5,00	6,00	
Etiqueta_	1,00	0	1	0	1	0	0	2
ImportacionesIndustria	2,00	1	4	3	3	1	0	12
	3,00	1	3	4	0	3	1	12
	4,00	0	5	2	0	4	0	11
	5,00	0	2	2	1	10	2	17
	6,00	0	0	0	1	1	1	3
Total		2	15	11	6	19	4	57

Por lo que la cota para el QSI es:

$$cQSI = \frac{0+4+4+0+10+1}{57} = 0.3333$$

Lo que indica que la series coinciden en un 33.33% en su comportamiento en los mismos instante.

Resumen

En la tabla siguiente se resume la información de la variable *Importaciones*, indicando la significación de cada una de las relaciones.

	Servicios	Comercio	Construcción	Industria	Global
Servicios		0.2105(0.1436)	0.1929(0.2325)	0.1754(0.6515)	0.3508(0.0000)
Comercio	0.2105(0.1436)		0.1403(0.6252)	0.1578(0.4840)	0.3157(0.0016)
Construcción	0.1929(0.2325)	0.1403(0.6252)		0.3684(0.0000)	0.2280(0.0821)
Industria	0.1754(0.6515)	0.1578(0.4840)	0.3684(0.0000)		0.3333(0.0005)
Global	0.3508(0.0000)	0.3157(0.0016)	0.2280(0.0821)	0.3333(0.0005)	

En este caso existen muy pocas relaciones significativas, pero todas las relaciones con las variables globales son significativas como era de esperar.

7. CONCLUSIONES.

En el presente estudio se desarrollo un análisis de las cuatro variables macroeconómicas importantes de El Salvador: Ventas, Empleo, Inversiones e Importaciones de manera global y por sectores: Servicios, Comercio, Construcción e Industria. Dichas variables fueron medidas desde el primer trimestre del año 1995 hasta el segundo trimestre de 2009.

El estudio se dividió en varias fases de las cuales podemos resaltar:

1. Análisis descriptivo de los datos.
2. Modelación de los datos globales.
3. Comparación de datos por sectores.

1. Análisis descriptivo de los datos.

Primero se desarrolló un estudio descriptivo de las 20 variables en estudio. Donde para cada variable se presentó su respectivo grafico de secuencia, y sus estadísticos principales, media, varianza, mínimo, máximo y rango. De aquí se obtuvo que la mayoría de las series son estacionarias en media, excepto la variable *Inversiones* para el sector Construcción que presenta tendencia. En general el comportamiento de las variables es muy similar en cada sector, aunque en el sector construcción se refleja una leve diferencia negativa respecto a los demás sectores.

2. Modelación de los datos globales.

Como uno de los objetivos principales del estudio es tratar de explicar el comportamiento de las variables macroeconómicas en función de ellas mismas y de las otras variables, se ha utilizado dos herramientas de modelización: modelos ARIMA Y Regresión Dinámica.

En los modelos ARIMA se obtuvieron los siguientes resultados:

- La variable *Ventas* es ajustada con el siguiente modelo:

$$Ventas_t = 7.518 + 0.651Ventas_{t-1} + a_t$$

Lo que indica que las ventas actuales dependen de las ventas del trimestre anterior. Con este modelo solo se explica el 40.90% de la variabilidad. Este bajo porcentaje nos indica que probablemente la variable Ventas no se explica por ella misma. Es razonable pensar que de las otras tres variables restantes es posible que haya una o varias que guarden relación con Ventas, y que dichas variables pueden explicar mejor a Ventas. Por eso se hace necesario realizar la regresión dinámica.

- La variable *Empleo* es ajustada con el modelo:

$$Empleo_t = 1.946 + 0.692Empleo_{t-1} + a_t$$

El modelo nos indica que el empleo actual sólo depende del empleo del trimestre anterior. El modelo explica el 39.2% de la variabilidad.

- La variable *Inversiones* es ajustada con el modelo:

$$Inversiones_t = 10.13 + 0.8Inversiones_{t-1} + 0.973Inversiones_{t-4} + 0.798a_{t-4} + a_t$$

Lo que nos indica que las inversiones actuales dependen de las inversiones del trimestres anterior y de la inversión de hace cuatro trimestres y de una perturbación aleatoria de hace cuatro trimestres. Con dicho modelo se ajustan el 53.7% de la variabilidad.

- Para la variable *Importaciones* el modelo ajustado es el siguiente:

$$Importaciones_t = 3.431 + 0.759Importaciones_{t-1} + a_t$$

Lo que nos indica que las importaciones actuales dependen únicamente de las importaciones del trimestre anterior. Con este modelo se ajustan el 48.7 % de la variabilidad.

Con estos modelos el porcentaje de ajuste no es muy bueno. Debido a que cada una de las variables está fuertemente correlada en los mismos instantes de tiempo con las otras variables, es necesario introducir las variables con las que guarda relación para mejorar los modelos. Un procedimiento para realizarlo es la denominada regresión dinámica.

Para ver que en efecto todas las variables están relacionadas se utilizan las correlaciones entre las variables en el instante t. Al realizar el procedimiento se comprobó que todas las variables están fuertemente relacionadas.

Con la regresión dinámica se pretende escribir la variable dependiente en función de los retardos de las variables independientes y los retardos de la misma.

Para obtener unos buenos modelos de partida para la regresión dinámica se utilizaron dos métodos: El método **FORWARD** (de menos a más) y el método **BACKWARD** (de más a menos).

El método **BACKWARD** consiste en iniciar con el modelo más básico. A partir de ese modelo básico se agregaran los factores faltantes en el modelo. Lo que le “hace falta” al modelo estará indicado por la función de autocorrelación de los residuos.

El método **FORWARD** consiste en iniciar con el modelo máximo. A partir de este modelo debemos ir eliminando los factores menos significativos hasta llegar al modelo final.

Con el valor de estos dos modelos iniciales y teniendo en cuenta la alta multicolinealidad existente entre todas las variables, se determino el modelo óptimo para cada una de las series. Siguiendo los procesos anteriores obtenemos los siguientes modelos:

- La variable *Ventas* es ajustada con el siguiente modelo:

$$Ventas_t = 4.490 + 0.535Ventas_{t-1} + 0.719Importaciones_t + a_t$$

Este modelo indica que las Ventas dependen de las Ventas del trimestre anterior y de las Importaciones actuales. Esta relación es muy clara ya que para vender se necesita materia prima que muchas veces es importada.

Con este modelo se obtiene que el 71.8% de la variabilidad es ajustada por el modelo, que es una mejora considerable comparándolo con el modelo ARIMA obtenido para ventas.

- La variable *Empleo* es ajustada con el modelo:

$$Empleo_t = -4.323 + 0.419Empleo_{t-1} + 0.602Inversiones_t + a_t$$

Es decir, el Empleo actual depende del Empleo del trimestre anterior y de las Inversiones actuales.

Con este modelo se logra un ajuste del 60.2% de la variabilidad es ajustada por el modelo, lo que es mejor que el conseguido con el modelo ARIMA.

- La variable *Importaciones* es ajustada con el siguiente modelo:

$$Importaciones_t = -7.58 + 0.392Ventas_t + 0.859Inversiones_t + a_t$$

El modelo nos indica que las Importaciones actuales dependen únicamente de las Ventas actuales y de la Inversión actual.

Con este modelo se tiene un ajuste del 82.7% de la variabilidad, que es notablemente superior al ajuste conseguido por el modelo ARIMA.

- La variable *Inversiones* es ajustada con el siguiente modelo:

$$Inversiones_t = 8.378 + 0.581Inversiones_{t-4} + 0.322Empleo_t + 0.421Importaciones_t + a_t$$

El modelo nos indica que la Inversiones actuales dependen únicamente de las Inversiones de hace cuatro trimestres, del Empleo y las Importaciones actuales.

El ajuste de este modelo es del 81.0% de la variabilidad, que también es un notablemente avance sobre el modelo ARIMA.

Con todo lo anterior podemos concluir que el modelo de regresión dinámica ajusta notablemente mejor que el modelo ARIMA. Se debe tener en cuenta que para poder utilizar el modelo de predicción dinámica, es necesario conocer el datos de las otras variables en el instante que deseamos predecir, lo que no siempre será posible.

3. Comparación de los datos.

La comparación de datos se realizó en dos partes:

- Comparación de variables por sectores. Comparar Las variables Ventas, Empleo, Importaciones e Inversiones para verificar si han tenido el mismo comportamiento entre ellas dentro de cada uno de los sectores. Dicho análisis se basa en el índice cualitativo de similitud (QSI), concretamente, en una cota de este índice (cQSI) que se interpreta como

el porcentaje de años coincidentes entre las dos series. Las conclusiones obtenidas son:

- a. En el sector Servicios no existen gran relación entre las distintas series. Solo se observa una relación entre las *Importaciones* y las *Ventas* y entre las *Importaciones* y las *Inversiones* y con p-valores superiores a 0.04.
 - b. En el sector Comercio las únicas variables que no se han comportado de manera similar son *Empleo* e *Importaciones*. Notar que todas la relaciones son significativas a nivel 0.01.
 - c. En el sector Construcción las variables presentan relaciones cruzadas, pero podemos concluir que las series *Empleo* e *Inversiones* y las variables *Ventas-Importaciones* son las que tienen parecido comportamiento dos a dos.
 - d. En el sector Industria todas las series presentan relaciones significativas a nivel 0.1, con lo que se puede decir que existen un comportamiento bastante similar de todas las series del sector Industria
 - e. De manera global todas las variables han tenido el mismo comportamiento. Podemos argumentar esto dado que todas las relaciones son significativas con un nivel de significación de 0.01.
- Comparación de variables. La comparación es de cada variable en cada uno de los sectores. Los resultados son:
 - a. Las *Ventas* en los distintos sectores se pueden agrupar de la siguiente forma. Las ventas del sector servicios y las del sector construcción por un lado, y las de Industria con las de comercio. Sobre la relación de las ventas de cada sector con las ventas globales, debemos destacar la n significación del sector de la construcción (seguramente esta no significación se deba al no alto valor de n).

- b. El *Empleo* presenta relaciones cruzadas. Lo más destacable que podemos obtener es la altísima relación con la variable empleo global (significativo con un nivel 0.004).
- c. Las *Inversiones* presentan relaciones significativas (a un nivel 0.1) entre todas las variables menos entre Construcción e Industria (con un p-valor 0.1436). Por lo que podemos suponer que todas las series tienen un comportamiento similar. Destacar nuevamente la altísima relación existente con la variable global de inversión.
- d. Las *Importaciones* solo se comportan de manera similar las series de construcción e industria y las series de cada sector con la serie global de importaciones.

De forma general podemos extraer varias conclusiones:

1. Si tenemos los datos de alguna de las series en estudio, es posible realizar un modelo que explique parte de su variabilidad (en torno al 40%). Pero no es posible precisar lo suficiente para realizar predicciones con un grado adecuado de precisión.
2. Si solo desconocemos el dato actual de la serie en estudio, (es decir, disponemos de las series independientes hasta la actualidad y de la dependiente hasta el instante inmediatamente anterior) podemos crear un modelo que es capaz de explicar mucho mejor la variabilidad (en torno al 70%). No se ha planteado la posibilidad de realizar predicciones por el escaso valor de n, pero se podría estudiar en un futuro.
3. Si disponemos de todas las series hasta el instante inmediatamente anterior, el resultado del modelo no mejora el obtenido con los datos de la propia serie.
4. No todos los sectores se comportan igual. Existen sectores donde se observa comportamientos similares entre las distintas variables y otros en los que esta igualdad es más limitada. Aunque de forma general se puede decir que existe relación entre las series de los distintos sectores.

5. Debemos destacar la gran relación existente entre todas las variables que miden el global de la economía Salvadoreña.
6. Con respecto al comportamiento de cada variable en los distintos sectores, indicar que nuevamente se observan discrepancias en función de la variable. Aunque lo que si que es común es la relación de cada variable del sector con su variable global. Esto se debe a que la variable global no es más que una agregación ponderada de las variables de los distintos sectores.

8. BIBLIOGRAFÍA.

1. Alfonso Novales: Estadística y Econometría, Editorial McGraw-Hill, 1997.
2. Álvaro Trigueros Argüello: Impacto del incremento del desempleo en Estados Unidos sobre las remesas familiares en El Salvador. FUSADES, 2009.
3. Antonio Aznar: Métodos de predicción en Economía I, Editorial Ariel, 1993.
4. Antonio Aznar: Métodos de predicción en Economía II, Editorial Ariel, 1993.
5. Ana M. Sánchez, Eugenio Fediani Martel: Redefiniendo la Interpolación para el Análisis de Series Temporales Irregulares. Universidad Pablo de Olavide, 2007.
6. Ana M. Martín, Valoración de la pobreza mediante técnicas de agregación de datos de diferente naturaleza. Tesis Doctoral, Universidad Pablo de Olavide, 2005.
7. Andrés M. Alonso, E. Ann Maharaj: Comparación de Series Temporales mediante Subsampling. Universidad Autónoma de Madrid/Universidad de Monash, 2006.
8. Daniel Peña: Análisis de series temporales, Alianza Editorial, S.A, 2005.
9. Gregory C. Reinsel. Elements of Multivariate Time Serie Analysis. Editorial Springer, 2001.
10. Helmut Lükepohl. New Introduction to Multiple Time Series Analysis. Editorial Springer, 2002.
11. James D. Hamilton: Time Series Analysis. Pricenton University Press, 1994.
12. Jaime Arnau Gras. Diseños de Series Temporales: técnicas de análisis. Edicions Universitat de Barcelona, 2001.
13. José Alberto Mauricio: Introducción al Análisis de Series Temporales, Universidad Complutense de Madrid, 1995.
14. José Hernández Alonso. Análisis de Series Temporales Económicas I. ESIC editorial, 2008.
15. José Maria Caridad. Econometría: modelos Econométricos y Series Temporales. Editorial Reverté. S.A, 1998.

16. Juan Antonio Ortega, Francisco Javier Cuberos. Comparación Cualitativa de Series Temporales. Índice Cualitativo de Similitud- QSI. . Instituto Politécnico Nacional de México, 2001.
17. Jeffrey M. Wooldridge. Introducción a la Econometría: un enfoque moderno. Thomson Learning, 2001.
18. Octavio Luís Pineda. Métodos y modelos Econométricos. Editorial Limusa, 1999.
19. Peter J. Brockwell, Richard A. Time Series: Theory and methods. Editorial Springer, 2002.
20. William W. Wei. Time Series Analysis Univariate and Multivariate Methods. Editorial Springer, 2006.
21. William W. S: Time Series Analysis. Addison-Wesley Publishing Company, 1995.