

FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN

UNIVERSIDAD DE LA REPÚBLICA



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

CONSTRUCCIÓN DE MODELOS DE SERIES TEMPORALES PARA PRONOSTICAR VENTAS EN ISUSA

Informe final de pasantía para obtener título
de Licenciado en Estadística

Romina Gonella Furtado
2013

Índice general

Índice general	1
1. Introducción	5
1.1. Objetivos	7
1.2. Antecedentes	7
2. Marco teórico	8
3. Marco metodológico	11
3.1. Procesos integrados	11
3.1.1. Procesos autorregresivos integrados de medias móviles (<i>ARIMA</i>)	11
3.1.2. Procesos <i>ARIMA</i> estacionales (<i>SARIMA</i>)	12
3.2. Modelos de Vectores Autorregresivos (<i>VAR</i>)	14
3.3. Cointegración y Modelos de Corrección de Error (<i>VECM</i>)	17
3.4. Etapas en la modelización univariada	18
3.4.1. Identificación	19
3.4.2. Estimación y selección de modelos <i>ARIMA</i>	19
3.4.3. Validación y Diagnóstico	23
3.4.4. Predicción	27
3.5. Etapas en la modelización multivariada	29
3.5.1. Definición de las variables a incluir en el sistema y determinación del orden de integración individual de las mismas.	29
3.5.2. Especificación del modelo - Prueba de Cointegración.	30
3.5.3. Estimación	31
3.5.4. Diagnóstico	32
3.5.5. Predicción y evaluación del modelo	34

4. Ácido sulfúrico	35
4.1. Identificación	35
4.2. Estimación	36
4.3. Validación y Diagnóstico	38
4.4. Predicción	38
5. Sulfato de aluminio	43
5.1. Identificación	43
5.2. Estimación	45
5.3. Validación y Diagnóstico	48
5.4. Predicción	48
6. Fertilizantes	52
6.1. Modelo univariado	52
6.1.1. Identificación	52
6.1.2. Estimación	56
6.1.3. Validación y Diagnóstico	59
6.1.4. Predicción	60
6.2. Modelo multivariado	64
6.2.1. Variables vinculadas a la serie objeto de estudio	64
6.2.2. Modelización	67
6.2.2.1. Prueba de cointegración y estimación	69
6.2.2.2. Validación y Diagnóstico	71
6.2.2.3. Predicción	71
7. Conclusiones	76
Bibliografía	78
A. Anexo de figuras	80
A.1. Ácido sulfúrico	80
A.2. Sulfato de aluminio	86
A.3. Fertilizantes	91
A.3.1. Modelo univariado	91

A.3.2. Modelo multivariado	100
B. Pruebas de existencia de raíz unitaria de Dickey Fuller	105
C. Prueba de cointegración de Johansen	109
D. Anexo de Código	112

Agradecimientos

Agradezco a todo el personal de ISUSA, especialmente a Virginia D´Agosto quien me abrió las puertas de esa casa, al contador Gerardo Martinez que me guió durante todo el período de la pasantía y a Rodney Larrea, compañero del día a día.

A mi tutora Laura Nalbarte por acompañarme durante todas las etapas de la pasantía y a Natalia Da Silva quien dedicó tanto tiempo y realizó innumerables aportes para que este trabajo pudiera concretarse.

A docentes y compañeros del Instituto de Estadística por el apoyo brindado, en especial a Silvia Rodríguez.

A mi familia, tanto de sangre como política, que me apoyó y supo comprender mis ausencias.

Finalmente agradezco a Derby, mi compañero, quien supo darme apoyo y contención en los momentos difíciles.

Capítulo 1

Introducción

ISUSA nace en 1947 como una empresa dedicada a la fabricación y venta de ácido sulfúrico y superfosfato¹, los cuales produce en su planta ubicada en el kilómetro 24 de la ruta 1 nacional, departamento de San José. Actualmente se dedica a la producción y comercialización de fertilizantes y productos químicos para la industria. Su producción se ha extendido mucho con respecto al emprendimiento inicial, hoy en día cuenta con plantas en Nueva Palmira (Colonia) y Agraciada (Soriano).

La empresa importa materias primas para procesarlas en las distintas plantas y elaborar sus productos, básicamente fosforita para fabricar superfosfato de calcio, azufre para hacer ácido sulfúrico, fosfato de amonio, fosfato monoamónico, cloruro de potasio y urea, los que se utilizan para formular o vender como tales en el mercado local. A partir de estas materias primas se elaboran distintos tipos de fertilizantes y productos químicos para usos tan diversos como la agricultura, la ganadería, la lechería, la forestación y la industria (fábricas de papel, laboratorios, curtiembres entre otros).

Las importaciones insumen aproximadamente dos meses entre que se inicia el trámite hasta que llega el producto a la fábrica, salvo casos excepcionales donde el plazo puede extenderse aún más. Por otro lado, el proceso productivo requiere cierto tiempo, que varía dependiendo del producto. Las materias primas y los productos terminados que aún no serán entregados se guardan en depósitos hasta el momento de ser utilizados o entregados, lo que genera el problema de contar con capacidad limitada y la necesidad de distribuir el espacio de forma eficiente a los efectos de cubrir los requerimientos de insumos de los distintos productos.

Dada esta situación, se plantea la necesidad de contar con un mecanismo que permita determinar con cierta anticipación el volumen aproximado de las ventas para poder organizar y planificar la importación y producción de manera de contar con el producto demandado en el momento oportuno. De esta manera se podría optimizar el uso del espacio disponible para depósito y reducir el tiempo de espera que se produce entre que el cliente solicita el producto y lo obtiene en su establecimiento. A su vez, se reduce el costo de oportunidad de tener almacenado un producto o materia prima que no será demandado o utilizado.

El propósito de este trabajo consiste en encontrar modelos para las ventas de **ácido sulfúrico**, **sulfato de aluminio** y **fertilizantes** en general, que sean de utilidad para que la empresa pueda realizar pronósticos eficientes de sus ventas y mejorar la planificación de su producción.

El **ácido sulfúrico** es un producto industrial fundamental, sus aplicaciones son muy numerosas

¹tipo de fertilizante fosfatado

y su consumo es extraordinario por su facilidad de reacción con otras materias. Es un producto que directa o indirectamente está presente en toda industria.

El **sulfato de aluminio** es una sal ampliamente usada en la industria, comúnmente en la purificación de agua potable y en la industria del papel.

Los **fertilizantes** son sustancias que aportan a las plantas uno o varios de los elementos nutritivos indispensables para su desarrollo vegetativo normal. Éstos se encargan de entregar y devolver a la tierra los nutrientes necesarios para el adecuado crecimiento de plantas, árboles, prados y arbustos.

A los efectos de simplificar el análisis, se consideran únicamente las ventas en plaza, excluyendo las exportaciones.

Se propone implementar en principio modelos univariados *ARIMA*² siguiendo la metodología propuesta por Box y Jenkins ([1]). Luego, se incluye información externa para construir modelos multivariados generalizando la metodología aplicada a los univariados.

El informe se divide en siete capítulos. El primero corresponde a la introducción, objetivos y antecedentes. En el segundo se desarrolla el marco teórico, donde se detallan los usos de los tres productos a modelizar. El tercero comprende el marco metodológico estadístico, los siguientes tres (capítulos cuatro a seis) están dedicados a la modelización de cada uno de los productos mencionados anteriormente. Por último se presentan las conclusiones donde se sintetizan los resultados obtenidos en cada modelización.

²Autorregresivos integrados de medias móviles, del inglés: Autorregresive Integrated Moving Average.

1.1. Objetivos

Objetivo general

El objetivo de esta investigación consiste en ajustar modelos de series temporales a las ventas en plaza de ácido sulfúrico, sulfato de aluminio y fertilizantes, producidos por ISUSA.

Se pretende encontrar modelos que contribuyan a la mejora de la gestión comercial de la empresa al hacer más eficientes algunos procesos como la compra de materia prima, la planificación de su producción y la organización y distribución del espacio de almacenamiento.

Objetivos específicos

- Profundizar el análisis de la serie de ventas de fertilizantes mediante la inclusión de variables vinculadas en modelos multivariados de tipo *VECM*³.
- Implementar de forma sencilla la utilización de los modelos resultantes para generar pronósticos mediante el uso del paquete estadístico R, a través de un programa hecho a medida que cubra la estimación y predicción y que sea capaz de actualizar las series a medida que se producen nuevos datos.

1.2. Antecedentes

Se revisó la bibliografía existente y no se encontraron antecedentes similares de aplicación de modelos *ARIMA* para la predicción de estos productos⁴, tanto a nivel nacional como en el exterior.

Se utilizó como referencia el informe final de pasantía de Magela Negro el cual desarrolla el análisis de series de tiempo univariadas y multivariadas para la demanda de combustibles en ANCAP. En este trabajo se emplean modelos *ARIMA* y *SARIMA* para la modelización univariada y modelos *VAR* para la modelización multivariada.

También se consideró un estudio realizado por el Instituto de Estadística (IESTA) en el año 2010 en el marco de un convenio con ANCAP, para realizar predicciones a largo plazo de la demanda de combustibles. En este estudio se utilizaron modelos *VAR* y *VECM*.

³Modelos de corrección de error o vector error correction models por las siglas en inglés.

⁴Ácido sulfúrico, sulfato de aluminio y fertilizantes.

Capítulo 2

Marco teórico

Ácido sulfúrico

El ácido sulfúrico es un producto industrial fundamental. Sus aplicaciones son muy numerosas y su consumo es extraordinario, por su facilidad de reacción con otras materias, eliminando metales, oxígeno, agua y otras sustancias no deseadas. Los usos del ácido sulfúrico son tan variados que el volumen de su producción proporciona un índice aproximado de la actividad general industrial. Este ácido es un producto que directa o indirectamente está presente en toda industria. En ISUSA el ácido sulfúrico se vende principalmente a granel.

Usos del ácido sulfúrico

- Producción de superfosfato de calcio (fertilizantes).
- Producción de sulfato de aluminio a partir de bauxita o hidróxido de aluminio para potabilización de agua y otros usos.
- Detergentes: en la sulfonación de dodecil-benceno, que es la materia prima básica para la mayoría de los detergentes utilizados en el hogar y la industria.
- Fábricas de Papel: En el proceso de producción de la pulpa de papel, e indirectamente con el uso de sulfato de aluminio.
- Agro-Fito Sanitario: en la producción de sulfato de cobre.
- Refinación de Petróleo: para las calderas y procesos químicos.
- Generación térmica de energía: para el tratamiento de las calderas.
- Metalurgia: para el decapado de metales.
- Producción de ácido para acumuladores eléctricos.
- Producción de sulfato de cromo: se lo utiliza en reacción con dicromato de potasio y un agente reductor. El sulfato de cromo se utiliza principalmente para el curtido de cueros (curtido al cromo).
- Fabricación de productos orgánicos, pinturas, pigmentos y rayón.
- Explosivos: por su gran capacidad de reacción.

- Producción de dióxido de cloro.

Sulfato de aluminio

El sulfato de aluminio es una sal que se puede presentar sólida y de color blanco o líquida, ampliamente usada en la industria. Puede obtenerse disolviendo hidróxido de aluminio en ácido sulfúrico.

Este producto es fabricado en ISUSA y se vende en polvo en bolsas de 25 kg, molido grueso en bolsas de 40 kg, en lajas o líquido, estos últimos se comercializan a granel.

Usos del sulfato de aluminio

- Floculante en la purificación de agua potable.
- Industria del papel, una parte se integra al papel como “mordiente”.
- Tratamiento de aguas residuales.
- Producción de otras sales de aluminio.
- Producción de jabones de aluminio y grasas para usos industriales.
- Industria del petróleo: manufactura de catalizadores sintéticos.
- Industria farmacéutica: como astringente en la preparación de drogas, cosméticos y anti-sudorales.
- Curtido de pieles finas, curtido al alumbre.

Fertilizantes

Se llama fertilizante a cualquier sustancia que aporte a las plantas uno o varios de los elementos nutritivos indispensables para su desarrollo vegetativo normal.

Las plantas extraen su alimento del agua del suelo y el oxígeno, hidrógeno y carbono del aire, pero para completar su alimentación necesitan utilizar ciertas sustancias químicas simples del suelo, son los llamados nutrientes vegetales.

Los fertilizantes se encargan de entregar y devolver a la tierra los nutrientes necesarios para el adecuado crecimiento de plantas, árboles, prados y arbustos.

Todos los suelos poseen una cierta cantidad de nutrientes vegetales provenientes de la parte mineral del suelo, (arena, arcilla, etc.) y del humus generado por el reciclaje de materias vegetales y animales caídas sobre la superficie (hojas, flores, raíces muertas, etc.).

Cuando se cultivan las plantas intensivamente, el equilibrio se altera, porque el proceso de reciclaje natural de los elementos esenciales del suelo es más lento de lo que demora la planta en utilizarlos.

Un fertilizante químico es un producto que contiene, por los menos, un elemento químico que la planta necesita para su ciclo de vida. La característica más importante de cualquier fertilizante es que debe tener una solubilidad mínima en agua, para que, de este modo pueda disolverse en el agua de riego, ya que la mayoría de los nutrientes entran en forma pasiva en la planta, a través del flujo del agua.

Estos elementos químicos o nutrientes pueden clasificarse en: macroelementos y microelementos.

Los macroelementos son aquéllos que se expresan como porcentaje de nutriente. Los principales son: nitrógeno, fósforo, potasio, calcio, magnesio y azufre.

Los microelementos se expresan como: ppm (parte por millón)¹, algunos de ellos son: hierro, zinc, cobre, manganeso, molibdeno, boro y cloro.

Existen distintos tipos de fertilizantes, según la composición de estos elementos, a modo de ejemplo existen los puros que poseen sólo un tipo de macroelemento o los compuestos, que poseen una combinación de ellos.

A su vez, estos se comercializan en distintas presentaciones, principalmente ISUSA los comercializa en tres: a granel, embolsado en bolsones o en bolsas. La diferencia entre bolsón y bolsa es la capacidad, los bolsones pueden almacenar hasta una tonelada mientras que las bolsas no superan los 50 kilogramos.

El fertilizante a granel es aquel que se vende sin envase², en estado sólido o líquido

¹Equivale a $\text{mg/kg} = \text{mg} / 1000\text{g}$.

²Es colocado directamente en un camión apropiado para transportarlo.

Capítulo 3

Marco metodológico

En el desarrollo de este capítulo se sigue a Peña ¹, Enders ², Pfaff, ³ y Juselius ⁴. Muchas partes se transcriben textualmente de dichos autores, aunque no siempre figuren entre comillas.

3.1. Procesos integrados

“Los procesos no estacionarios más importantes son los procesos integrados, que tienen la propiedad fundamental de que al diferenciarlos⁵ se obtienen procesos estacionarios”.

“Una propiedad importante que diferencia a los procesos integrados de los estacionarios es la forma en que desaparece la dependencia con el tiempo. En los procesos estacionarios *ARMA*⁶ las autocorrelaciones disminuyen geoméricamente, y se hacen prácticamente cero a los pocos retardos. En los procesos integrados las autocorrelaciones disminuyen linealmente con el tiempo y es posible encontrar coeficientes de autocorrelación distintos de cero hasta para retardos muy altos”.

“Se dice que un proceso es integrado de orden $h \geq 0$, y se representa por $I(h)$, cuando al diferenciarlo h veces se obtiene un proceso estacionario. Un proceso estacionario es, por tanto, siempre $I(0)$. En la práctica la mayoría de las series no estacionarias que son integradas tienen un orden $h \leq 3$ ”.

3.1.1. Procesos autorregresivos integrados de medias móviles (*ARIMA*⁷)

La idea de la sección anterior puede generalizarse para cualquier proceso *ARMA*, permitiendo que una o varias raíces del operador *AR* sean la unidad. Se obtienen entonces procesos del tipo:

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d y_t = c + (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t \quad (3.1)$$

llamados *ARIMA*(p, d, q). En esta notación p es el orden de la parte autorregresiva estacionaria, d es el número de raíces unitarias (orden de integración del proceso) y q es el orden de la parte

¹Peña, D. (2005), *Análisis de series temporales* ([11]).

²Enders, W. (1995), *Applied Econometric Time Series* ([7]).

³Pfaff, B. (2008), *Analysis of Integrated and Cointegrated Time Series with R* ([12]).

⁴Juselius, K. (2006), *The Cointegrated VAR model, methodology and applications* ([13]).

⁵En este contexto equivale a aplicar la primera diferencia regular: $\Delta = 1 - L$.

⁶*AR* de autorregresivo y *MA* de medias móviles.

⁷Del inglés autorregresive integrated moving average.

media móvil. Utilizando el operador diferencia, $\Delta = 1 - L$, el proceso anterior suele escribirse:

$$\phi_p(L)\Delta^d y_t = c + \theta_q(L)\varepsilon_t \quad (3.2)$$

3.1.2. Procesos *ARIMA* estacionales (*SARIMA*⁸)

Una serie es estacional cuando su valor esperado no es constante, pero varía con una pauta cíclica. En concreto si:

$$E(z_t) = E(z_{t+s})$$

se dice que la serie tiene una *estacionalidad de período s*. El período estacional, s , se define como el número de observaciones que forman el ciclo estacional.

Se puede convertir una serie no estacionaria con estacionalidad en estacionaria mediante la transformación:

$$z_t = \Delta_s^D \Delta^d y_t$$

donde D es el número de diferencias estacionales (si hay estacionalidad casi siempre $D = 1$, si no hay estacionalidad $D = 0$) y d el número de diferencias regulares (por lo general $d \leq 3$).

“Cuando exista dependencia estacional se puede generalizar el modelo *ARMA* para series estacionarias incorporando además de la dependencia *regular*, que es la asociada a los intervalos naturales de medida de la serie, la dependencia *estacional*, que es la asociada a observaciones separadas por s periodos”.

“Un enfoque más simple para modelar estos procesos, y que funciona bien en la práctica, consiste es modelar de forma separada la dependencia regular y la estacional, y construir el modelo incorporando ambas. Se obtiene así el modelo *SARIMA* que tiene la forma”:

$$\Phi_P(L^s)\phi_p(L)\Delta_s^D \Delta^d y_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t \quad (3.3)$$

donde:

$\Phi_P(L^s) = (1 - \Phi_1 L^s - \dots - \Phi_P L^{sP})$ es el operador *AR* estacional de orden P ,

$\phi_p(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$ es el operador *AR* regular de orden p ,

$\Delta_s^D = (1 - L^s)^D$ representa las diferencias estacionales y $\Delta^d = (1 - L)^d$ las regulares,

$\Theta_Q(L^s) = (1 - \Theta_1 L^s - \dots - \Theta_Q L^{sQ})$ es el operador media móvil estacional de orden Q ,

$\theta_q(L) = (1 - \theta_1 L - \dots - \theta_q L^q)$ es el operador media móvil regular de orden q y

ε_t es un proceso de ruido blanco.

⁸Seasonal autorregresive integrated moving average por sus siglas en inglés.

Esta clase de modelos, introducidos por Box y Jenkins, representan bien muchas series estacionales que se encuentran en la práctica y se escriben de forma simplificada como el modelo $SARIMA(p, d, q)(P, D, Q)_s$.

“El modelo $SARIMA$ se basa en la hipótesis central de que la relación de dependencia estacional es la misma para todos los períodos”.

Función de autocorrelación simple de un proceso estacional

Sea $z_t = \Delta^d \Delta_s^D y_t$ el proceso estacionario obtenido al diferenciar la serie d veces de forma regular y D veces de forma estacional. Entonces, z_t sigue el proceso $ARMA$ estacional multiplicativo:

$$\Phi_P(L^s)\phi_p(L)z_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t \quad (3.4)$$

La función de autocorrelación simple de este proceso es una mezcla de las funciones de autocorrelación correspondientes a las partes regular y estacional.

1. En los retardos bajos ($j=1, \dots, 6$) se observará únicamente la parte regular.
2. En los retardos estacionales se observará básicamente la parte estacional.
3. Alrededor de los retardos estacionales se observará la “interacción” entre la parte regular y estacional que se manifestará en la repetición a ambos lados de cada retardo estacional de la función de autocorrelación simple de la parte regular.

En la práctica lo recomendable es fijarse en los retardos iniciales $1, 2, 3, 4, \dots$ para identificar la estructura regular y en los retardos estacionales, $s, 2s, 3s, \dots$ para identificar la estructura estacional. La interacción alrededor de los coeficientes estacionales puede entonces utilizarse como confirmación de la identificación realizada.

Función de autocorrelación parcial de un proceso estacional

La función de autocorrelación parcial de un proceso estacional multiplicativo es compleja, porque es una superposición no sólo de las funciones de autocorrelación parciales de las partes regular y estacional, sino que en la interacción alrededor de los retardos estacionales aparece a un lado la función de autocorrelación parcial de la parte regular y al otro la simple. En concreto, puede demostrarse que la función de autocorrelación parcial de un proceso estacional tiene la estructura siguiente:

1. En los primeros retardos aparecerá la PACF⁹ de la estructura regular y en los estacionales la PACF de la estacional.
2. A la derecha de cada coeficiente estacional (retardos $js+1, js+2, \dots$) aparecerá la PACF de la parte regular. Si el coeficiente estacional es positivo la PACF regular aparece invertida, mientras que si éste es negativo la PACF aparece con su signo.
3. A la izquierda de los coeficientes estacionales (retardos $js-1, js-2, \dots$) se observará la función de autocorrelación simple de la parte regular.

⁹Función de autocorrelación parcial.

Se comprueba que en modelos estacionales los coeficientes de correlación parcial para los retardos estacionales suelen ser de magnitud mucho menor que la que correspondería a una estructura estacional pura.

Esta complejidad recomienda fijarse sólo en los retardos iniciales $1, 2, 3, 4, \dots$ para identificar la estructura regular, y en los retardos estacionales, $s, 2s, 3s, \dots$ para identificar la estructura estacional.

3.2. Modelos de Vectores Autorregresivos (VAR^{10})

En el caso de dos variables, puede ocurrir que el patrón temporal de y_t sea afectado por presente y pasadas realizaciones de z_t y que a su vez el patrón temporal de z_t sea afectado por presente y pasadas realizaciones de y_t . Se considera un sistema bivariado simple:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{y_t} \quad (3.5)$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{z_t} \quad (3.6)$$

donde se asume:

1. y_t y z_t son procesos estacionarios
2. ε_{y_t} y ε_{z_t} son procesos de ruido blanco con desviaciones estándar σ_y y σ_z respectivamente
3. ε_{y_t} y ε_{z_t} son ruidos blancos incorrelacionados.

Las ecuaciones anteriores constituyen un vector autorregresivo de primer orden ($VAR(1)$) dado que el rezago más alto es uno. La estructura del sistema incorpora retroalimentación ya que y_t y z_t son capaces de afectarse el uno al otro.

Los términos ε_{y_t} y ε_{z_t} son shocks o innovaciones puras en y_t y z_t respectivamente. Si b_{21} es distinto de cero, ε_{y_t} tiene un efecto contemporáneo indirecto en z_t y si b_{12} es distinto de cero ε_{z_t} tiene un efecto contemporáneo indirecto en y_t .

Las ecuaciones anteriores no están expresadas en forma reducida ya que y_t tiene un efecto contemporáneo en z_t y z_t tiene un efecto contemporáneo en y_t . Es posible transformar el sistema de ecuaciones en una forma más fácil de usar, usando álgebra se puede escribir el sistema en forma más compacta.

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y_t} \\ \varepsilon_{z_t} \end{bmatrix}$$

o

$$Bx_t = \Gamma_0 + \Gamma_1x_{t-1} + \varepsilon_t$$

donde

¹⁰Del inglés vector autorregressive models.

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$

$$x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{y_t} \\ \varepsilon_{z_t} \end{bmatrix}$$

Al premultiplicar por B^{-1} se obtiene el vector autorregresivo (*VAR*) en *forma estándar*:

$$x_t = A_0 + A_1 x_{t-1} + e_t \quad (3.7)$$

donde:

$$A_0 = B^{-1}\Gamma_0$$

$$A_1 = B^{-1}\Gamma_1$$

$$e_t = B^{-1}\varepsilon_t$$

Se definen a su vez: a_{i0} como el elemento i del vector A_0 , a_{ij} como el elemento en fila i y columna j de la matrix A_1 y e_{it} como el elemento i del vector e_t . Con esta notación se puede escribir 3.7 en la forma:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t} \quad (3.8)$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \quad (3.9)$$

El primer sistema es llamado *VAR estructural* o *sistema primitivo* mientras que el segundo es llamado *VAR en forma estándar*. Es importante notar que los términos de error e_{1t} y e_{2t} están formados por los shocks ε_{y_t} y ε_{z_t} . Dado que $e_t = B^{-1}\varepsilon_t$ se puede escribir e_{1t} y e_{2t} como:

$$e_{1t} = (\varepsilon_{y_t} - b_{12}\varepsilon_{z_t})/(1 - b_{12}b_{21}) \quad (3.10)$$

$$e_{2t} = (\varepsilon_{z_t} - b_{21}\varepsilon_{y_t})/(1 - b_{12}b_{21}) \quad (3.11)$$

Como ε_{y_t} y ε_{z_t} son procesos ruido blanco, ambos tienen media cero, varianza constante y están (individualmente) serialmente incorrelacionados:

$$E(e_{1t}) = E[(\varepsilon_{y_t} - b_{12}\varepsilon_{z_t})/(1 - b_{12}b_{21})] = 0$$

$$Var(e_{1t}) = E(e_{1t}^2) = E[(\varepsilon_{y_t} - b_{12}\varepsilon_{z_t})^2/(1 - b_{12}b_{21})^2] = (\sigma_y^2 + b_{12}^2\sigma_z^2)/(1 - b_{12}b_{21})^2$$

$$E(e_{1t}e_{1t-i}) = E[(\varepsilon_{y_t} - b_{12}\varepsilon_{z_t})(\varepsilon_{y_{t-i}} - b_{12}\varepsilon_{z_{t-i}})]/(1 - b_{12}b_{21})^2 = 0$$

La varianza no depende del tiempo. De forma similar se puede demostrar lo anterior para e_{2t} . Sin embargo e_{1t} y e_{2t} están correlacionados. La covarianza entre los dos términos es:

$$E(e_{1t}e_{2t}) = E[(\varepsilon_{y_t} - b_{12}\varepsilon_{z_t})(\varepsilon_{z_t} - b_{21}\varepsilon_{y_t})]/(1 - b_{12}b_{21})^2$$

$$= -(b_{21}\sigma_y^2 + b_{12}\sigma_z^2)/(1 - b_{12}b_{21})^2$$

En general la ecuación anterior no será cero, por lo que los dos shocks estarán correlacionados. En el caso especial donde $b_{12} = b_{21} = 0$ (no hay efecto contemporáneo de y_t en z_t ni de z_t en y_t), los shocks estarán incorrelacionados.

Se define la matriz de varianzas y covarianzas de e_{1t} y e_{2t} como:

$$\Sigma = \begin{bmatrix} Var(e_{1t}) & Cov(e_{1t}, e_{2t}) \\ Cov(e_{1t}, e_{2t}) & Var(e_{2t}) \end{bmatrix}$$

Ya que todos los elementos de Σ son independientes del tiempo, se puede usar la forma más compacta

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

donde

$$Var(e_{it}) = \sigma_i^2$$

$$\sigma_{12} = \sigma_{21} = Cov(e_{1t}, e_{2t})$$

VAR(p)

Se considera la siguiente generalización multivariada de (3.7).

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + e_t \quad (3.12)$$

donde:

x_t es un vector de dimensión $n \times 1$ que contiene cada una de las n variables incluidas en el VAR.

A_0 es un vector de n constantes

A_i son matrices de coeficientes de dimensión $n \times n$

e_t es un vector de n términos de error

La expresión 3.12 cual puede expresarse en forma matricial como:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot \\ x_{nt} \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{20} \\ \cdot \\ A_{n0} \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) & \cdot & A_{1n}(L) \\ A_{21}(L) & A_{22}(L) & \cdot & A_{2n}(L) \\ \cdot & \cdot & \cdot & \cdot \\ A_{n1}(L) & A_{n2}(L) & \cdot & A_{nn}(L) \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ \cdot \\ x_{nt-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ \cdot \\ e_{nt} \end{bmatrix} \quad (3.13)$$

donde

A_{i0} son parámetros que representan las constantes

$A_{ij}(L)$ son polinomios en el operador de rezagos L

Los coeficientes individuales de $A_{ij}(L)$ son llamados $a_{ij}(1), a_{ij}(2), \dots$. Dado que todas las ecuaciones utilizan la misma cantidad de rezagos, todos los polinomios $A_{ij}(L)$ son del mismo grado. Los términos e_{it} son ruidos blancos que pueden estar correlacionados. Nuevamente se denomina Σ a la matriz de varianzas y covarianzas de dimensión $n \times n$.

Las variables a ser incluidas en el VAR son seleccionadas de acuerdo al modelo económico relevante. Los tests sobre la cantidad de rezagos seleccionan el largo apropiado.

3.3. Cointegración y Modelos de Corrección de Error (*VECM*¹¹)

El concepto de *cointegración* fue introducido por Granger en 1987.

La definición de cointegración es la siguiente:

Los componentes del vector $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$ son *cointegrados* de orden d, b , denotados por $x_t \sim CI(d, b)$ si

1. todos los componentes de x_t son integrados de orden d
2. existe un vector $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ tal que la combinación lineal $\beta x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt}$ es integrada de orden $(d - b)$, donde $b > 0$

El vector β es llamado *vector de cointegración*. Este vector no es único ya que para cualquier λ , $\lambda(\beta_1, \beta_2, \dots, \beta_n)$ es también un vector de cointegración. Típicamente se utiliza una de las variables para normalizar el vector de cointegración haciendo que su coeficiente valga uno. Si se quiere normalizar β respecto a x_{1t} , se selecciona $\lambda = 1/\beta_1$.

Si x_t tiene n componentes, pueden existir hasta $n - 1$ vectores de cointegración linealmente independientes. El número de vectores de cointegración se denomina *rango de cointegración* de x_t .

Corrección de error

El vector $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$ tiene una representación de corrección de error si puede ser expresado como

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \pi_1 \Delta x_{t-1} + \pi_2 \Delta x_{t-2} + \dots + \pi_p \Delta x_{t-p} + \varepsilon_t \quad (3.14)$$

donde

π_0 es un vector de dimensión $n \times 1$ de términos constantes con elementos π_{i0}

π_i una matriz de coeficientes de dimensión $n \times n$ con elementos $\pi_{jk}(i)$

π es una matriz con elementos π_{jk} tal que uno o más de los $\pi_{jk} \neq 0$

ε_t un vector de dimensión $n \times 1$ con elementos ε_{it} , tal que ε_{it} puede estar correlacionado con ε_{jt}

Sean todas las variables en x_t $I(1)$, si existe una representación de corrección de error de esas variables como (3.14), necesariamente existe una combinación lineal de las variables $I(1)$ que es estacionaria. Despejando πx_{t-1} de (3.14) se obtiene

$$\pi x_{t-1} = \Delta x_t - \pi_0 - \sum_{i=1}^p \pi_i \Delta x_{t-i} - \varepsilon_t$$

¹¹Del inglés vector error correction models.

Como cada expresión del lado derecho es estacionaria, πx_{t-1} debe también ser estacionaria. Como π contiene sólo constantes, cada fila de π es un vector de cointegración de x_t ; la primera fila puede escribirse como $(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \dots + \pi_{1n}x_{nt-1})$. Dado que cada serie x_{it-1} es $I(1)$, $(\pi_{11}, \pi_{12}, \dots, \pi_{1n})$ debe ser un vector de cointegración para x_t . La característica clave en (3.14) es la presencia de la matriz π . Hay dos puntos importantes al respecto:

1. Si todos los elementos de π son iguales a cero, (3.14) es un VAR tradicional en primeras diferencias, no existe corrección de error dado que Δx_t no responde a la desviación del período previo del equilibrio de largo plazo
2. Si una o más de las π_{jl} difiere de cero, Δx_t responde a la desviación del período previo del equilibrio de largo plazo. Entonces, estimar un VAR en primeras diferencias es inapropiado. La omisión de la expresión πx_{t-1} implica un error de especificación si x_t tiene una representación de corrección de error como (3.14).

En el caso de n variables, pueden haber múltiples vectores de cointegración.

Hay varias formas de testear la cointegración. En este trabajo se utiliza la metodología de Johansen.

El procedimiento se basa en determinar el rango de la matriz π , que equivale al número de vectores de cointegración existentes. Existen dos variantes del modelo según se utilice el estadístico del máximo autovalor o el de la traza.

3.4. Etapas en la modelización univariada

La metodología seguida en este trabajo es la propuesta por Box y Jenkins en 1976 para modelización de series univariadas. Estos modelos son univariados en el sentido de que se construyen en función de los valores pasados de la serie sin utilizar otras variables aleatorias (sí pueden contener variables determinísticas). Las etapas que constituyen este procedimiento son ¹²:

1. Identificación de modelos posibles.
2. Estimación de parámetros.
3. Validación y diagnóstico del o los modelos candidatos.
4. Predicción.

La primera etapa consiste en *identificar* el posible modelo *ARIMA* que sigue la serie, lo que requiere decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria y determinar un modelo *ARMA* para la serie estacionaria, es decir, los órdenes p y q de su estructura autorregresiva y de media móvil y, si el proceso es estacional, los órdenes P y Q de la estructura *ARMA* estacional.

Una vez seleccionado provisoriamente un modelo para la serie estacionaria se pasa a la segunda etapa de estimación, donde los parámetros *AR* y *MA* del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo. La tercera etapa es la de diagnóstico, donde se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos no contienen información, se acepta el modelo

¹²En el desarrollo de esta sección se sigue a Peña, D. (2005), *Análisis de Series Temporales* ([11]).

como adecuado y se lo utiliza para la previsión o la toma de decisiones. Si los residuos muestran estructura, se modifica el modelo para incorporar la y se repiten las tres etapas anteriores hasta obtener un modelo adecuado.

3.4.1. Identificación

En esta etapa se determina si es necesario transformar la serie para que tenga varianza constante y/o diferenciarla para que sea estacionaria.

Cuando la variabilidad de la serie aumente linealmente con el nivel de la serie, tomando logaritmos se consigue una serie con variabilidad constante.

Para estabilizar la media de la serie puede ser necesario aplicar diferencias regulares y estacionales. La decisión de aplicar estas diferencias puede basarse en el gráfico de la serie y en las función de autocorrelación muestral, pero también se pueden realizar contrastes formales.

La necesidad de aplicar diferencias regulares se puede identificar en el gráfico de la serie si esta presenta tendencia o muestra cambios de nivel en la media. También se podrá identificar en autocorrelograma simple por el decrecimiento lento de las autocorrelaciones de forma lineal a medida que aumenta el rezago.

Por otro lado la necesidad de aplicar diferencias estacionales se manifiesta a través del gráfico de la serie que presentará una pauta repetida de período s (el período estacional) y también se manifestará a través del autocorrelograma simple que presentará coeficientes positivos que decrecen lentamente en los rezagos $s, 2s, 3s, \dots$

Contrastes de raíces unitarias

Si se desea decidir si un proceso es o no estacionario se pueden realizar contrastes de raíces unitarias. Estos contrastes indican si se debe o no tomar una diferencia adicional en una serie para convertirla en estacionaria. En este trabajo se utiliza la prueba de Dickey Fuller aumentado¹³.

Dickey y Fuller consideran tres modelos diferentes que pueden ser usados para testear la presencia de raíz unitaria:

$$\begin{aligned}\Delta y_t &= \gamma y_{t-1} + \varepsilon_t \\ \Delta y_t &= a_0 + \gamma y_{t-1} + \varepsilon_t \\ \Delta y_t &= a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t\end{aligned}$$

En todos los modelos el parámetro de interés es γ ; si $\gamma = 0$ el proceso posee una raíz unitaria.

3.4.2. Estimación y selección de modelos ARIMA

Se dispone de una serie estacionaria (original o transformada) de T elementos $\{z_t\} \ t = 1, \dots, T$, y se quiere estimar por máxima verosimilitud los parámetros de un modelo *ARMA* especificado. La función de verosimilitud exacta de un proceso *ARMA*, asumiendo normalidad, es:

¹³Ver anexo B.

$$L(\beta) = -\frac{T}{2}\log(\sigma^2) - \frac{1}{2}\sum_{t=1}^T \log(v_{t|t-1}) - \frac{1}{2\sigma^2}\sum_{t=1}^T \frac{e_t^2}{v_{t|t-1}} \quad (3.15)$$

donde tanto las varianzas condicionales $v_{t|t-1}$ como los errores e_t dependen de los parámetros. La maximización de la función de verosimilitud exacta se efectúa con un algoritmo de optimización no lineal.

La verosimilitud condicional se obtiene haciendo que el o los valores iniciales de la serie y eventualmente el o los valores iniciales de los residuos sean dados y no estocásticos. Llamando $r = \max(p, q)$ y $\beta = (\mu, \phi_1, \dots, \theta_q, \sigma^2)$ al vector de parámetros, la función de verosimilitud condicional es:

$$L_C(\beta|\mathbf{e}_0, \mathbf{z}_p) = \frac{-(T-r)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=r+1}^T e_t^2 \quad (3.16)$$

donde se tiene que $e_t^2 = e_t^2(\beta|\mathbf{e}_0, \mathbf{z}_p)$, para resaltar que las innovaciones se calculan a partir de unos vectores \mathbf{e}_0 y \mathbf{z}_p de valores iniciales.

La maximización de (3.16) requiere un valor inicial de los parámetros que puede obtenerse con algún algoritmo como el de Hannan y Rissanen¹⁴.

Criterios de selección de modelos

Una vez que se ha estimado un conjunto de modelos M_1, \dots, M_m y que se ha analizado la significación de sus parámetros, se procede a seleccionar el que mejor explica la serie observada. Para ello puede utilizarse el criterio *AIC* de Akaike que considera el ajuste y la cantidad de parámetros. El indicador *AIC* se define como:

$$AIC = E(-2L(\beta)) = T\log(\hat{\sigma}_{MV}^2) + 2k \quad (3.17)$$

donde $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ es el vector de parámetros, T es el tamaño muestral utilizado para estimar el modelo, $\hat{\sigma}_{MV}^2$ el estimador MV de la varianza de las innovaciones y k el número de parámetros estimados para calcular las predicciones a un paso. Por tanto, seleccionar el modelo con verosimilitud esperada máxima equivale a escoger el que minimiza la verosimilitud con signo negativo dado por la ecuación 3.17. Esto indica que entre un conjunto de modelos estimados el que se ajuste más a la serie observada será el que presente el menor valor en el *AIC*.

El problema con el *AIC* es que tiende a sobreestimar el número de parámetros en el modelo y este efecto puede ser muy grande en pequeñas muestras. Una alternativa que corrige esta sobrestimación es el criterio *AIC* corregido, *AICC*, dado por:

$$AICC = T\log(\hat{\sigma}_{MV}^2) + T\frac{(1+k/T)}{1-(k+2)/T} \quad (3.18)$$

Si se utiliza el criterio *AIC* para comparar modelos ajustados a una serie es importante que T , el número efectivo de observaciones utilizado para estimar el modelo sea el mismo para todos ellos. El criterio *AIC* tiene validez si los parámetros no son conocidos y se estiman con los datos y esperanza de la verosimilitud se calcula respecto a futuras observaciones.

¹⁴Por explicación del algoritmo, ver Peña ([11]).

Análisis de intervención

Box y Tiao (1975) denominaron análisis de intervención a la inclusión en un modelo de series temporales de sucesos específicos que producen efectos deterministas mediante variables ficticias.

El proceso de estimación por máxima verosimilitud equivale a las tres etapas siguientes:

1. Se parte de una estimación inicial del efecto de la intervención, \hat{w}_0 , se construye la serie corregida del efecto de la intervención:

$$z_t = y_t - \hat{w}_0 I_h(t)$$

2. Se estiman los parámetros del modelo *ARIMA* en la serie corregida, z_t , de la forma habitual, maximizando la verosimilitud.
3. Se calculan los residuos del modelo con los parámetros estimados, y se utilizan estos residuos para estimar el efecto de la intervención escribiendo los residuos como función de estos efectos.

De esta manera se obtiene una ecuación de regresión, donde los residuos, $\hat{\varepsilon}_t$, son la variable dependiente, y la intervención modificada, ξ_t , es la variable independiente o regresor. El parámetro a estimar en esta regresión es el tamaño del efecto, w_0 . Escribiendo la ecuación como:

$$\hat{\varepsilon}_t = w_0 \xi_t + \varepsilon_t$$

la estimación del parámetro será $\hat{w}_0 = \sum \hat{\varepsilon}_t \xi_t / \sum \xi_t^2$ Con este nuevo valor de w_0 se vuelve a (1) iterando estas tres etapas hasta obtener convergencia.

Con mucha frecuencia ocurren en las series reales hechos puntuales que se desconocen. Es importante ser capaz de identificar estas situaciones desconocidas porque:

1. Si sus efectos son grandes pueden sesgar la estimación de los parámetros, lo que producirá malas predicciones futuras.
2. Si el suceso ha ocurrido en la última parte de la serie y alguna observación afectada se utiliza para generar predicciones estas no serán buenas, incluso aunque los parámetros estén bien estimados.
3. Si estos sucesos atípicos pueden volver a aparecer en el futuro y se identifican y estiman sus efectos, se puede incorporar esta información en las predicciones y obtener intervalos de predicción más realistas.

Existen muchos tipos de atípicos, los más habituales son los aditivos, innovativos, cambios de nivel, cambios transitorios y efecto rampa. A continuación se detallan los tipos de atípicos que fueron detectados en las series objeto de estudio en el presente trabajo.

Atípicos aditivos (AO^{15}): Se dice que ha ocurrido un AO sobre una serie temporal en el instante h si el valor de la serie se genera en ese instante de manera distinta al resto. El modelo que seguirá la serie observada, y_t si ha sido afectada por un AO en t será:

$$y_t = \begin{cases} x_t & \text{si } t \neq h \\ x_t + w_A & t = h \end{cases}$$

donde x_t sigue un modelo $ARIMA$

$$x_t = \psi(L)\varepsilon_t$$

Entonces, el modelo que sigue la serie observada, y_t , es

$$y_t = w_A I_h(t) + \psi(L)\varepsilon_t \quad (3.19)$$

donde $I_h(t) = 0, t \neq h; I_h(h) = 1$. La huella que un atípico aditivo deja en la serie es una alteración del valor en un punto.

Atípicos innovativos (IO^{16}): Se dice que ha ocurrido un IO en una serie temporal en $t = h$ cuando la innovación en ese punto está directamente afectada por una cantidad desconocida debida a un suceso atípico.

La innovación de una serie temporal representa el efecto agregado de todas las variables que influyen sobre la serie observada de forma aleatoria.

El modelo para la serie que sufre un IO de magnitud w_I en el instante h es

$$y_t = \psi(L)(w_I I_h(t) + \varepsilon_t) \quad (3.20)$$

Cambios transitorios (TC^{17}): El TC está definido por

$$y_t = \frac{w_{TC}}{1 - \delta L} I_h(t) + \psi(L)\varepsilon_t$$

En este tipo de atípicos se produce una alteración de la serie en un punto que también afecta a los siguientes pero el efecto va desapareciendo a medida que los puntos se alejan del momento en que ocurrió el atípico, hasta que en cierto punto desaparece dicho efecto.

Atípicos estacionales (SO^{18}): El SO está definido por

$$y_t = \begin{cases} x_t & \text{si } t < h \\ x_t + w_S & t = h, sh, 2sh, \dots \\ x_t - \frac{1}{11}w_S & \text{si } t > h \text{ y } t \neq ksh, k \in N \end{cases}$$

¹⁵ Additive outlier por sus siglas en inglés.

¹⁶ Innovative outlier por sus siglas en inglés.

¹⁷ Transitory change por sus siglas en inglés.

¹⁸ Seasonal outlier por sus siglas en inglés.

Cambios de nivel estacionales (SLS^{19}):

$$y_t = \frac{1}{1 - L^s} w_{SLS} I_t^{(\alpha)} + x_t$$

donde $I_t^{(\alpha)}$ es una indicatriz que vale uno en $t = \alpha$ y cero en el resto.

Procedimientos de estimación de atípicos

En la práctica la posición y la naturaleza de los atípicos que pueden aparecer sobre una serie es desconocida y se precisa un procedimiento para identificarlos y detectar sus efectos. En concreto se requiere, para cada dato atípico:

- Detectar el momento de aparición.
- Identificar el tipo de atípico.
- Estimar su tamaño.

Una vez conocidos estos aspectos es posible eliminar su efecto sobre la serie. Existe un procedimiento desarrollado por Chang, Liu y Tiao ([3]) diseñado para encontrar atípicos del tipo AO , IO , TC y LS^{20} . En los TC se supone que el denominador es fijo, igual a 0,7, con lo que el modelo para este efecto es $w/(1 - 0,7L)$. En todos los cuatro casos el atípico queda definido por un único parámetro w .

Dada una serie temporal que sigue un modelo estimado, $\hat{\pi}(L)y_t = \hat{e}_t$, la detección de los posibles valores atípicos se realiza detectando el instante y tipo más probable y contrastando si es significativo. Si lo es, se “limpia” de la serie, es decir, se elimina su efecto y se repite el proceso hasta que el más probable no sea significativo, con lo que el procedimiento termina.

3.4.3. Validación y Diagnóstico

El diagnóstico del modelo requiere comprobar que se cumplen las hipótesis realizadas respecto a las innovaciones.

Estas deben tener:

1. Media marginal igual a cero.
2. Varianza marginal constante.
3. Falta de correlación para cualquier retardo.
4. Distribución normal.

Además, estas propiedades deben verificarse no sólo respecto a las distribuciones marginales sino también a las distribuciones condicionadas a cualquier conjunto de información de valores pasados de la serie.

¹⁹Seasonal level shift por sus siglas en inglés.

²⁰Cambio de nivel, level shift por sus siglas en inglés.

Contrastes de autocorrelación

El primer contraste a realizar es si las innovaciones están incorrelacionadas. Para ello se estima su función de autocorrelación simple mediante:

$$\hat{r}_k = \frac{\sum_{t=1}^{T-k} (e_t - \bar{e})(e_{t+k} - \bar{e})}{\sum_{t=1}^T (e_t - \bar{e})^2} \quad (3.21)$$

donde \bar{e} es la media de los T residuos. Si los residuos son independientes, los coeficientes, \hat{r}_k , para k no muy pequeño, serán aproximadamente variables aleatorias con media cero, varianza asintótica $1/T$ y distribución normal. La varianza asintótica es válida para k grande, pero no para los primeros retardos.

En consecuencia, el valor $1/\sqrt{T}$ debe considerarse como un límite máximo de la desviación típica de las autocorrelaciones residuales.

Prueba de Box-Pierce: Con el objetivo de realizar una prueba conjunta con respecto a las autocorrelaciones, Box y Pierce (1970) desarrollaron el siguiente contraste:

$$\tilde{Q}(r) = T(T+2) \sum_{j=1}^h \frac{r_j^2}{T-j} \quad (3.22)$$

con

$$r_j = \frac{\sum_{t=j+1}^T \varepsilon_t \varepsilon_{t-j}}{\sum_{t=1}^T \varepsilon_t^2}$$

Si el modelo ajustado era el correcto y los parámetros conocidos, la ecuación 3.22 tiene, para T grande, distribución χ_h^2 ya que la distribución límite para $r = (r_1, \dots, r_h)'$ es normal multivariada con vector de medias cero, $Var(r_j) = \frac{T-j}{T(T+2)}$ y $Cov(r_l, r_m) = 0$ ($l \neq m$).

Usando la aproximación $Var(r_j) = \frac{1}{T}$, los autores sugieren que la distribución de

$$Q(r) = T \sum_{j=1}^h r_j^2$$

podría ser aproximada por una χ_{h-p-q}^2 . Además, mostraron que cuando los $p+q$ parámetros de un modelo apropiado son estimados y los \hat{r}_j reemplazan a los r_j , entonces

$$Q(\hat{r}) = T \sum_{j=1}^h \hat{r}_j^2$$

tiene, para T grande, distribución χ_{h-p-q}^2 .

Estudios posteriores demostraron que la distribución de $Q(\hat{r})$ puede desviarse de χ_{h-p-q}^2 , lo cual puede deberse a la aproximación de la varianza, por lo que Ljung y Box propusieron la variante que se detalla a continuación.

Prueba de Ljung-Box: Otro contraste global de que los primeros h coeficientes son cero (h debe ser grande) es el contraste de Ljung-Box. Si las innovaciones se comportan como ruido blanco, los coeficientes de autocorrelación estimados son asintóticamente normales, con media cero y varianza $(T - k)/T(T + 2)$. Por tanto, el estadístico

$$Q(\hat{r}) = T(T + 2) \sum_{j=1}^h \frac{\hat{r}_j^2}{T - j} \quad (3.23)$$

se distribuye, asintóticamente, como una χ^2 con grados de libertad igual al número de coeficientes en la suma (h) menos el número de parámetros estimados, K . Para modelos no estacionales $K = p + q + 1$, o $K = p + q$, según que el modelo tenga o no constante, y para los estacionales, que habitualmente no tienen constante, $K = P + p + Q + q$.

Se concluye que el modelo es inadecuado si el valor de $Q(\hat{r})$ obtenido con la ecuación 3.23 es mayor que el percentil $1 - \alpha$ de la distribución χ^2 con $h - K$ grados de libertad, que se denota por $\chi_{1-\alpha}^2(h - K)$.

También se suelen analizar los correlogramas de los residuos al cuadrado, ya que pueden mostrar dependencia entre los residuos que no se advierte en los anteriores. A su vez es posible aplicar las pruebas de Box-Pierce y Ljung-Box a esta serie de residuos al cuadrado.

Contraste de media cero

Para contrastar la hipótesis de que la perturbación tiene esperanza nula en el caso general, suponiendo T residuos y $p + q$ parámetros, se calcula su media:

$$\bar{e} = \frac{\sum e_t}{T}$$

y su varianza

$$\hat{s}_e^2 = \frac{\sum_{t=1}^T (e_t - \bar{e})^2}{T - p - q}$$

y se concluye que $E(\bar{e}_t) \neq 0$, si

$$\frac{\bar{e}}{\hat{s}_e/\sqrt{T}}$$

es mayor al valor crítico dado por la distribución $N(0, 1)$. Este contraste debe de aplicarse después de comprobar que los residuos están incorrelacionados, para asegurar que \hat{s}_e es un estimador razonable de la varianza.

Contraste de homoscedasticidad

La estabilidad de la varianza marginal de las innovaciones se comprueba estudiando el gráfico de los residuos a lo largo del tiempo. Si a la vista de los residuos estimados parece existir un cambio de varianza a partir de un punto, $t = n_1$, se puede dividir el intervalo muestral en dos partes y aplicar un contraste de varianzas. En la hipótesis de que ambos tramos tienen la misma varianza, el estadístico:

$$F = \frac{\sum_{t=1}^{n_1} \hat{\varepsilon}_t^2 / n_1}{\sum_{t=n_1+1}^T \hat{\varepsilon}_t^2 / (T - n_1)} = \frac{s_1^2}{s_2^2}$$

se distribuirá aproximadamente como una F con n_1 y $T - n_1$ grados de libertad. Para aplicar este contraste conviene que, al menos, se disponga de 10 observaciones en cada tramo. De la misma forma, si se sospechan h cambios de varianza en los períodos n_1, \dots, n_h el contraste de igualdad de varianzas es:

$$\lambda = T \log(\hat{\sigma}^2) - \sum_{i=1}^h n_i \log(s_i^2)$$

donde $\hat{\sigma}^2$ es la varianza de los residuos en toda la muestra y s_i^2 la varianza en el tramo i de longitud n_i observaciones. Bajo la hipótesis de que la varianza es la misma en todos los tramos se este estadístico será, asintóticamente, una χ^2 con $h - 1$ grados de libertad. se rechaza la hipótesis nula cuando el estadístico F es mayor al valor crítico $\chi_{1-\alpha}^2(h - 1)$.

Contrastes de normalidad

En las pruebas que se explican a continuación, la hipótesis nula corresponde al cumplimiento de la normalidad, mientras que la alternativa es el no cumplimiento.

Prueba de Jarque-Bera: La prueba se basa en comparar los coeficientes de simetría y curtosis de los residuos estimados con los de la distribución normal estándar.

El coeficiente de simetría y el de kurtosis de una distribución cualquiera se definen como el tercer y cuarto momento respectivamente y pueden expresarse como:

$$cs_x = E(x^3) \tag{3.24}$$

$$ck_x = E(x^4) \tag{3.25}$$

En particular, para la distribución normal estándar estos coeficientes valen cero (simetría) y tres (kurtosis).

El estadístico de la prueba es:

$$X = \frac{Tcs^2}{6} + \frac{T(ck - 3)^2}{24}$$

el cual se distribuye χ^2 con dos grados de libertad. Se rechaza la hipótesis nula cuando el estadístico X es mayor al valor crítico correspondiente, $\chi_{1-\alpha}^2(2)$.

Prueba de Shapiro-Wilks: El estadístico de la prueba se define como:

$$W = \frac{(\sum_{i=1}^T a_i x_{(i)})^2}{\sum_{i=1}^T (x_i - \bar{x})^2}$$

donde $x_{(i)}$ es el estadístico de orden i , el valor de la variable que ocupa el i -ésimo lugar en la muestra ordenada.

$\bar{x} = \frac{\sum_{i=1}^T x_i}{T}$ es la media muestral
las constantes a_i se calculan:

$$(a_1, a_2, \dots, a_T) = \frac{m^{tr} V^{-1}}{(m^{tr} V^{-1} V^{-1} m)^{1/2}}$$

donde $m = (m_1, \dots, m_T)^{tr}$ son los valores medios de los estadísticos de orden, de variables aleatorias independientes e idénticamente distribuidas, simuladas de distribuciones normales. V es la matriz de covarianzas de ese estadístico de orden.

Se rechaza la hipótesis nula cuando W es menor al valor crítico correspondiente.

3.4.4. Predicción

Los predictores óptimos son aquellos que minimizan, en promedio, los errores de predicción al cuadrado. Estos predictores se obtienen calculando los valores esperados de las observaciones futuras, condicionadas a los datos observados.

Se supone que se ha observado una realización de longitud T , de una serie temporal $\mathbf{y}_T = (y_1, \dots, y_T)$, y se desea prever un valor futuro $k > 0$ períodos adelante, y_{T+k} . Se llama $\hat{y}_{T+k|T} = f(\mathbf{y}_T)$ a un predictor de y_{T+k} obtenido como función de los valores observados

$$\hat{y}_{T+k|T} = \alpha_1 y_T + \alpha_2 y_{T-1} + \dots + \alpha_T y_1$$

Un predictor quedara definido al indicar los parámetros $\alpha_1, \dots, \alpha_T$ utilizadas para construirlo. Se quiere que la predicción este lo más cerca posible del verdadero valor que se quiere prever, de manera que el error de predicción

$$e_{T+k|T} = y_{T+k} - \hat{y}_{T+k|T}$$

sea lo más pequeño posible.

Si se quiere minimizar el error cuadrático medio de predicción (ECMP) de y_{T+k} dada la información \mathbf{y}_T . Entonces, se tiene que minimizar:

$$ECMP(y_{T+k}|\mathbf{y}_T) = E[(y_{T+k} - \hat{y}_{T+k})^2|\mathbf{y}_T] = E[e_{T+k}^2|\mathbf{y}_T]$$

donde la esperanza se toma respecto a la distribución de la variable y_{T+k} condicionada a los valores observados y_T . Se puede demostrar que el predictor que minimiza este error cuadrático medio es la esperanza de la variable y_{T+k} condicionada a la información disponible

$$\hat{y}_{T+k|T} = E[y_{T+k}|\mathbf{y}_T] = \mu_{T+k|T}$$

Se definen

$$\begin{aligned}\hat{y}_T(j) &= E[y_{T+j}|\mathbf{y}_T] \quad j = 1, 2, \dots \\ \hat{\varepsilon}_T(j) &= E[\varepsilon_{T+j}|\mathbf{y}_T] \quad j = 1, 2, \dots\end{aligned}$$

donde el subíndice T representa el origen de la predicción, que se supone fijo, y j el horizonte de la misma, que irá cambiando para generar predicciones de distintas variables futuras desde el origen T .

Sea $y_t = \psi(L)\varepsilon_t$ la representación $MA(\infty)$ del proceso. Entonces:

$$y_{T+k} = \sum_0^{\infty} \psi_i \varepsilon_{T+k-i} \quad (\psi_0 = 1) \quad (3.26)$$

La predicción óptima será, tomando esperanzas condicionadas a las primeras T observaciones

$$\hat{y}_t(k) = \sum_0^{\infty} \psi_{k+j} \varepsilon_{T-j} \quad (3.27)$$

y restando (3.27) de (3.26) se obtiene el error de predicción

$$e_T(k) = y_{T+k} - \hat{y}_T(k) = \varepsilon_{T+k} + \psi_1 \varepsilon_{T+k-1} + \dots + \psi_{k-1} \varepsilon_{T+1}$$

cuya varianza será:

$$Var(e_T(k)) = \sigma^2(1 + \psi_1^2 + \dots + \psi_{k-1}^2) \quad (3.28)$$

Esta ecuación muestra que la incertidumbre de la predicción es muy diferente para modelos estacionarios y no estacionarios. En un modelo estacionario $\psi_k \rightarrow 0$ si $k \rightarrow \infty$, y la varianza de la predicción a largo plazo converge a un valor constante, la varianza marginal del proceso. Esto es consecuencia de que la predicción a largo plazo es la media del proceso.

Para modelos no estacionarios la serie $\sum \psi_i^2$ no es convergente y la incertidumbre de la predicción a largo plazo crece sin límite hasta infinito. A largo plazo no se puede prever el comportamiento de un proceso no estacionario. Si se supone que la distribución de las innovaciones es normal, entonces $\hat{y}_T(k)$ será una variable normal de media y_{T+k} y varianza dada por (3.28), con lo que se pueden calcular intervalos de confianza para la predicción de la siguiente forma:

$$y_{T+k} \in (\hat{y}_T(k) \pm \lambda_{1-\alpha} \sqrt{\text{Var}(e_T(k))}) \quad (3.29)$$

donde $\lambda_{1-\alpha}$ es el percentil $1 - \alpha$ de la distribución normal estándar.

Los intervalos de predicción mostrados anteriormente se calculan suponiendo los parámetros conocidos y teniendo sólo en cuenta la incertidumbre por el desconocimiento de las innovaciones futuras. En general, cuando el modelo se estima con los datos, se tienen cuatro tipos de incertidumbre en la predicción, asociadas al desconocimiento de:

1. Las innovaciones futuras.
2. La distribución de las innovaciones.
3. Los verdaderos valores de los parámetros.
4. El modelo que ha generado los datos.

3.5. Etapas en la modelización multivariada

A grandes rasgos los pasos a llevar adelante para obtener un modelo multivariado son:

1. Definición de las variables a incluir en el sistema y determinación del orden de integración individual de las mismas.
2. Especificación del modelo - Prueba de Cointegración.
3. Estimación.
4. Diagnóstico
5. Predicción y evaluación del modelo.

3.5.1. Definición de las variables a incluir en el sistema y determinación del orden de integración individual de las mismas.

En primer lugar se identifican las variables a incluir en el sistema. El criterio para decidir cuales variables incluir depende del contexto del que se trate. Se espera que exista cierta correlación entre las variables de forma que expliquen en mayor o menor medida el comportamiento unas de otras.

La elección de las variables está sujeta también a su disponibilidad y posibilidad de recolección, ya que en algunos casos puede tratarse de fenómenos difíciles o imposibles de registrar.

La tarea de elegir posibles variables para integrar un sistema debe sustentarse con teoría, evidencia empírica o experiencia de expertos en el tema.

Luego de identificar las posibles variables a incluir, se realizan análisis exploratorios para conocer su comportamiento individual e intentar descubrir si hay interrelación entre ellas. Se realizan gráficos de las series individuales y se comparan sus trayectorias. Es muy importante en esta etapa realizar también gráficos de las correlaciones cruzadas entre las series.

Para cada una de las variables consideradas se efectúan pruebas de raíces unitarias de Dickey Fuller para determinar el orden de integración.

3.5.2. Especificación del modelo - Prueba de Cointegración.

Si todas las variables a incluir tienen individualmente el mismo orden de integración y este es mayor a cero, es necesario realizar una prueba de cointegración para determinar si están cointegradas, ya que en ese caso el modelo adecuado es un *VECM*, mientras que si no existe cointegración el modelo adecuado es un *VAR* en diferencias.

En este trabajo se utiliza la prueba de Johansen para testear la cointegración.

La especificación del modelo está muy relacionada con la prueba de cointegración dado que la especificación depende del resultado que se obtenga en la prueba de cointegración y a su vez para aplicar la prueba de cointegración es necesario especificar el tipo de modelo. Ambas etapas se llevan a cabo de forma simultánea ya que para cada especificación que se considera se obtiene un resultado en la prueba de cointegración.

Las variables pueden incluirse en el modelo como endógenas o exógenas. Las endógenas se determinan dentro del sistema y aparecen como variables a explicar en una ecuación y como variables explicativas en las otras ecuaciones. Las exógenas se determinan fuera del sistema y por lo tanto aparecen siempre como variables explicativas.

Causalidad de Granger

Un test de causalidad se utiliza para determinar si los rezagos de una variable entran en la ecuación de otra variable. Para entenderlo mejor, se explicita a continuación el modelo VAR(p) general con dos ecuaciones:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix} + \begin{pmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \quad (3.30)$$

donde:

$$A_{ij}(L) = a_{ij}(1) + a_{ij}(2)L + a_{ij}(3)L^2 + \dots + a_{ij}(p)L^{(p-1)}.$$

En este contexto se dice que $\{z_t\}$ no causa en el sentido de Granger a $\{y_t\}$ si los valores rezagados de $\{z_t\}$, z_{t-i} , no entran en la ecuación de y_t , es decir, si todos los coeficientes de $A_{21}(L)$ son iguales a cero.

La forma directa de determinar la causalidad de Granger es usar una prueba F estándar a la restricción:

$$a_{21}(1) = a_{21}(2) = a_{21}(3) = \dots = 0$$

En el caso de n variables en donde $A_{ij}(L)$ representa los coeficientes de valores rezagados de la variable j en la variable i , la variable j no causa en el sentido de Granger a la variable i si todos los coeficientes del polinomio $A_{ij}(L)$ son cero.

Las cinco variantes en la especificación del modelo

Partiendo del siguiente VAR(1) simplificado (sin efectos de corto plazo):

$$\Delta x_t = \alpha \beta' x_{t-1} + \mu_0 + \mu_1 t + e_t \quad (3.31)$$

donde

$$\mu_0 = \alpha \beta_0 + \gamma_0$$

y

$$\mu_1 = \alpha \beta_1 + \gamma_1$$

se consideran 5 variantes que surgen de imponer distintas restricciones a los componentes determinísticos.

- **Caso 1:** $\mu_0 = \mathbf{0}$ y $\mu_1 = \mathbf{0}$. Este caso corresponde a un modelo sin componentes determinísticos. $E(\Delta x_t) = 0$ y $E(\beta' x_t) = 0$.
- **Caso 2:** $\mu_1 = \mathbf{0}$ y $\gamma_0 = \mathbf{0}$ pero $\beta_0 \neq \mathbf{0}$. La constante está restringida a las relaciones de cointegración. En este caso no hay tendencias lineales en los datos, $E(\Delta x_t) = 0$. El único componente determinístico es la constante en las relaciones de cointegración, lo que implica que la media de equilibrio es diferente de cero.
- **Caso 3:** $\mu_1 = \mathbf{0}$ y μ_0 no está restringida. No hay tendencia lineal en el VAR pero sí puede haber en las variables en niveles. Si $\mu_0 \neq 0$ hay tendencia lineal en las variables en niveles y presencia de constante en la relación de cointegración. Si $\gamma_0 \neq 0$ y $\beta_0 = 0$ hay tendencia lineal en las variables en niveles pero la tendencia se cancela en las relaciones de cointegración.
- **Caso 4:** $\gamma_1 = \mathbf{0}$, pero γ_0 , β_0 y β_1 no están restringidos. La tendencia está restringida únicamente a las relaciones de cointegración, pero la constante no está restringida en el modelo. Cuando $\gamma_1 = 0$ hay tendencia lineal en los datos. Si además $\beta_1 \neq 0$ la tendencia no se cancela en las relaciones de cointegración.
- **Caso 5:** μ_0 y μ_1 no restringidas. Tendencia y constante no restringidas en el modelo.

La elección entre los distintos casos es muy importante ya que la distribución asintótica de los estadísticos de las pruebas de hipótesis dependen de las restricciones impuestas a los componentes determinísticos.

3.5.3. Estimación

En el caso de un VAR(n) como 3.12 el vector A_0 contiene n términos y cada matriz A_i contiene n^2 coeficientes, entonces necesitan ser estimados $n + pn^2$ términos. Un VAR estará sobreparametrizado si alguno de esos coeficientes pudiera ser excluido del modelo. Dado que el lado derecho de (3.12) contiene únicamente variables predeterminadas y que los términos de error se asumen serialmente incorrelacionados con varianza constante, cada ecuación en el sistema puede estimarse usando MCO²¹.

²¹Mínimos Cuadrados Ordinarios.

3.5.4. Diagnóstico

En esta sección se sigue a Pfaff, B. (2006) *Analysis of Integrated and Cointegrated Time series with R* ([12]).

Pruebas de incorrelación

Se utiliza la prueba de Portmanteau y el test LM propuesto por Breusch y Godfrey. El estadístico de Portmanteau se define como

$$Q_h = T \sum_{j=1}^h \text{tr}(\hat{C}'_j \hat{C}_o^{-1} \hat{C}_j \hat{C}_0^{-1})$$

donde

$$\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^T \hat{e}_t \hat{e}'_{t-i}$$

Este estadístico se distribuye aproximadamente $\chi^2(K^2h - n^*)$, donde K es la cantidad de variables endógenas y n^* es el número de coeficientes de un VAR(p) excluyendo términos determinísticos. Para muestras pequeñas puede tener mejores propiedades el siguiente estadístico

$$Q_h^* = T^2 \sum_{j=1}^h \frac{1}{T-j} \text{tr}(\hat{C}'_j \hat{C}_o^{-1} \hat{C}_j \hat{C}_0^{-1})$$

El estadístico LM de Breusch-Godfrey se basa en las siguientes regresiones auxiliares

$$\hat{e}_t = A_1 x_{t-1} + \dots + A_p x_{t-p} + CD_t + B_1 \hat{e}_{t-1} + \dots + B_h \hat{e}_{t-h} + \varepsilon_t$$

La hipótesis nula es $H_0 : B_1 = \dots = B_h = 0$, y la alternativa $H_1 : \exists B_i \neq 0$ para $i = 1, 2, \dots, h$. El test se define como

$$LM_h = T(K - \text{tr}(\tilde{\Sigma}_R^{-1} \tilde{\Sigma}_E))$$

donde $\tilde{\Sigma}_R$ y $\tilde{\Sigma}_E$ representan las matrices de varianzas y covarianzas de los residuos de los modelos restringido y no restringido, respectivamente. El estadístico LM_h se distribuye $\chi^2(hK^2)$. Edgerton y Shukur propusieron una corrección a este estadístico, que se define como:

$$LMF_h = \frac{1 - (1 - R_r^2)^{1/r} N r - q}{(1 - R_r^2)^{1/r} K m}$$

donde:

$$R_r^2 = 1 - |\tilde{\Sigma}_E|/|\tilde{\Sigma}_R|$$

$$r = ((K^2 m^2 - 4)/(K^2 + m^2 - 5))^{1/2}, q = 1/2 K m - 1 \text{ y}$$

$$N = T - K - m - 1/2(K - m + 1),$$

con n igual al número de regresores en el sistema original y $m = Kh$. Este estadístico se distribuye

$F(hK^2, int(Nr - q))$.

Se rechaza la hipótesis nula cuando el estadístico LMF_h es mayor a $F_{1-\alpha}(hK^2, int(Nr - q))$, donde α es el nivel de significación de la prueba.

Prueba de homoscedasticidad

Se utilizan los test ARCH tanto univariados como multivariados.

El test multivariado se basa en la siguiente regresión

$$vech(\hat{e}_t \hat{e}_t') = \beta_0 + B_1 vech(\hat{e}_{t-1} \hat{e}_{t-1}') + \dots + B_q vech(\hat{e}_{t-q} \hat{e}_{t-q}') + v_t$$

donde v_t representa un proceso de error esférico²² y $vech$ es el operador de apilamiento de columnas para matrices simétricas que apila las columnas de la diagonal principal hacia abajo. La dimensión de β_0 es $\frac{1}{2}K(K+1)$ y la de las matrices B_i $\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1)$. La hipótesis nula es $H_0 : B_1 = B_2 = \dots = B_q = 0$ y la alternativa es $H_1 : B_1 \neq 0 \cap B_2 \neq 0 \cap \dots \cap B_q \neq 0$. El test estadístico se define como

$$VARCH_{LM}(q) = \frac{1}{2}TK(K+1)R_m^2$$

con

$$R_m^2 = 1 - \frac{2}{K(K+1)}tr(\hat{\Omega}\hat{\Omega}_0^{-1})$$

donde $\hat{\Omega}$ representa la matriz de varianzas y covarianzas del modelo de regresión definido más arriba. Este estadístico se distribuye $\chi^2(qK^2(K+1)^2/4)$.

La hipótesis nula es rechazada cuando el estadístico $VARCH_{LM}(q)$ es superior a $\chi_{1-\alpha}^2(qK^2(K+1)^2/4)$.

Prueba de normalidad

Se utiliza la prueba de Jarque-Bera para series multivariadas, la cual se aplica a los residuos de un $VAR(p)$. El estadístico se define como

$$JB_{mv} = s_3^2 + s_4^2$$

donde s_3^2 y s_4^2 se definen como

$$s_3^2 = Tb_1'b_1/6$$

$$s_4^2 = T(b_2 - 3K)'(b_2 - 3K)/24$$

y b_1 y b_2 son los vectores que representan el tercer y cuarto momento de los residuos estandarizados $\hat{e}_t^s = \tilde{P}^-(\hat{e}_t - \bar{\hat{e}}_t)$ y \tilde{P} es una matriz triangular inferior con diagonal positiva tal que $\tilde{P}\tilde{P}' = \tilde{\Sigma}_e$, esto es, la descomposición de Choleski de la matriz de varianzas y covarianzas de los residuos. El estadístico JB_{mv} se distribuye $\chi^2(2K)$, mientras que tanto la simetría como la kurtosis multivariada se distribuyen $\chi^2(K)$.

Se rechaza la hipótesis nula cuando el estadístico JB_{mv} es mayor al valor crítico $\chi_{1-\alpha}^2(2K)$.

²²Homoscedástico y no autocorrelacionado.

3.5.5. Predicción y evaluación del modelo

Una vez que el modelo *VAR* fue estimado y pasó las pruebas diagnósticas, puede usarse para predicción. Para un *VAR* dado, se pueden realizar predicciones calculando recursivamente la siguiente ecuación:

$$y_{T+h|T} = A_1 y_{T+h-1|T} + \dots + A_p y_{T+h-p|T} + CD_{T+h}$$

para $h = 1, 2, \dots, n$. La matriz de varianzas y covarianzas de los errores de predicción está dada por:

$$Cov \left(\begin{bmatrix} y_{T+1} - y_{T+1|T} \\ \vdots \\ y_{T+h} - y_{T+h|T} \end{bmatrix} \right) = \begin{bmatrix} I & 0 & \dots & 0 \\ \Phi_1 & I & & 0 \\ \vdots & & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \dots & I \end{bmatrix} (\Sigma_e \otimes I_h) \begin{bmatrix} I & 0 & \dots & 0 \\ \Phi_1 & I & & 0 \\ \vdots & & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \dots & I \end{bmatrix}'$$

donde Φ_i son las matrices de coeficientes de la representación de medias móviles de Wold de un proceso *VAR*(p) estable y \otimes es el producto de Kronecker²³. Los intervalos de confianza para la predicción pueden calcularse como

$$[y_{k,T+h|T} - c_{1-\frac{\alpha}{2}} \sigma_k(h), y_{k,T+h|T} + c_{1-\frac{\alpha}{2}} \sigma_k(h)]$$

donde $c_{1-\frac{\alpha}{2}}$ representa el percentil $(1 - \frac{\alpha}{2})$ de la distribución normal y $\sigma_k(h)$ el desvío estándar de la k -ésima variable en el paso h .

²³Sea A una matriz de $m \times n$ y B una matriz de $p \times q$, el producto $A \otimes B$ da como resultado una matriz de dimensión $mp \times nq$. A cada elemento de A se lo multiplica por la matriz B .

Capítulo 4

Ácido sulfúrico

Se llevan a cabo para la serie objeto de estudio las etapas explicadas en la metodología:

- Identificación.
- Estimación.
- Validación y Diagnóstico.
- Predicción.

4.1. Identificación

La serie que se muestra en la figura 4.1 corresponde a las ventas mensuales en plaza de ácido sulfúrico producido por ISUSA, entre enero de 1998 y diciembre de 2010.

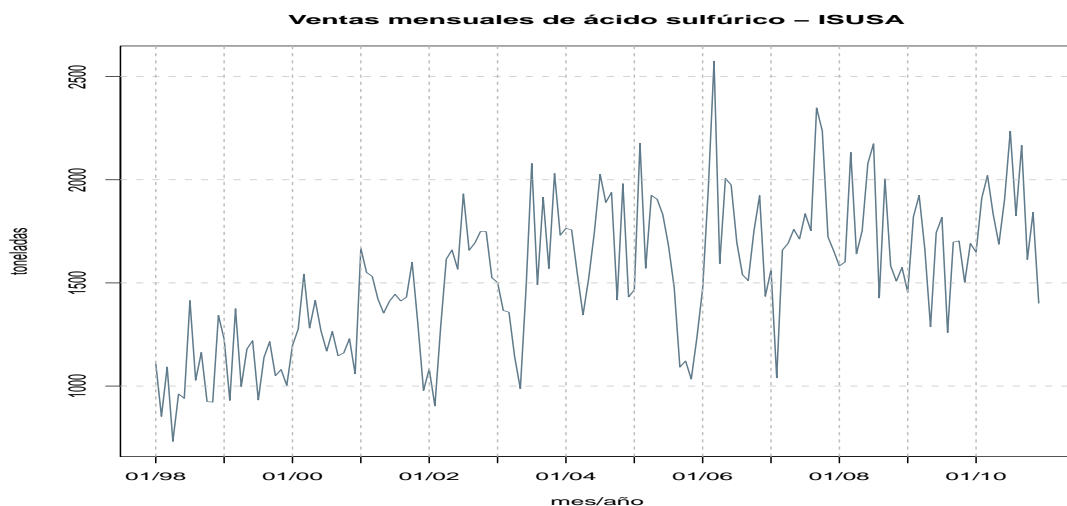


Figura 4.1: Gráfico de la serie de ventas mensuales de ácido sulfúrico.

Del análisis del gráfico se observa una serie con tendencia y posible componente estacional. En primer lugar se analiza la transformación logarítmica de la serie, los resultados determinan

que no es conveniente la transformación de la misma.

A continuación se realizan las pruebas de Dickey Fuller de raíz unitaria a la serie original. Se obtiene como resultado que se rechaza la hipótesis de existencia de raíz unitaria, por lo que no se considera la diferenciación de la serie pero se analiza la inclusión de una tendencia determinística. En el análisis de los autocorrelogramas (figura 4.2) se observan muchos rezagos significativos que siguen una pauta estacional dada por las autocorrelaciones que crecen y decrecen con el tiempo. Esto sugiere que el proceso pueda tener estacionalidad.

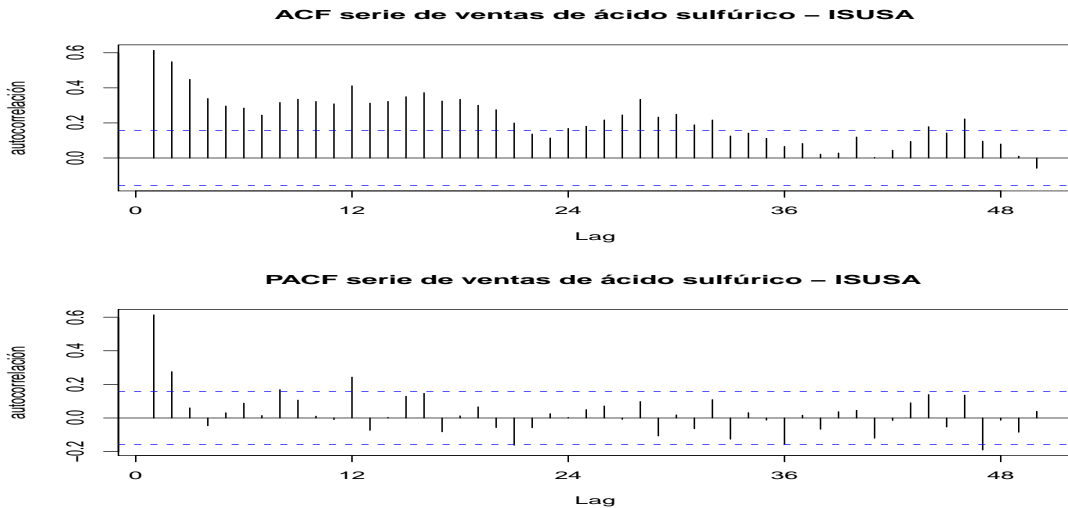


Figura 4.2: Autocorrelogramas simple y parcial de la serie de ventas de ácido sulfúrico.

La estructura del correlograma no es fácil de identificar ya que si bien se observa un comportamiento de los primeros rezagos que concuerda con una estructura autorregresiva, hay muchos rezagos significativos cuando deberían tender a cero para confirmar dicha estructura.

Se observa estacionalidad por las autocorrelaciones que decrecen y luego vuelven a crecer.

En principio se prueba incluir en los modelos una estacionalidad determinística mediante 11 variables que indican los meses de enero a noviembre. Estas variables resultan no significativas, por lo que se descarta esa posibilidad. Se decide incorporar la estacionalidad en el modelo *ARIMA*. Ante esta situación, se sugieren los siguientes modelos:

- $SARIMA(1, 0, 0)(1, 0, 0)$
- $SARIMA(2, 0, 0)(1, 0, 0)$
- $SARIMA(0, 0, 1)(1, 0, 0)$
- $SARIMA(0, 0, 2)(1, 0, 0)$

4.2. Estimación

Se estiman los modelos identificados en la etapa anterior a los que se incluye una tendencia determinística y el efecto calendario. Este último fue analizado considerando distintas configuraciones, en cada modelo se elige la forma de modelar este efecto mediante el *AICC*; la que presente el menor valor será la elegida. Finalmente se incluye para cada mes la cantidad de días de lunes a

viernes corregida por la cantidad de sábados y domingos.

El parámetro ϕ_2 resulta no significativo para el modelo $SARIMA(2, 0, 0)(1, 0, 0)$, por lo cual este modelo es descartado.

Para el resto se realiza una detección automática de atípicos¹, que arroja los siguientes resultados:

- Para $SARIMA(1, 0, 0)(1, 0, 0)$: LS en julio de 2003, AO en marzo de 2006.
- Para $SARIMA(0, 0, 1)(1, 0, 0)$: LS en julio de 2002.
- Para $SARIMA(0, 0, 2)(1, 0, 0)$: LS en julio de 2003, AO en marzo de 2006.

Al estimar los modelos con esas intervenciones se obtiene que los parámetros del cambio de nivel (LS) para el primer y tercer modelo no son significativos. Se quitan esas intervenciones y se vuelven a estimar los modelos. Al observar los residuos, se advierte que hay posibles atípicos que no fueron detectados (figura A.1 en anexo A).

Se realizan nuevas intervenciones obteniéndose los modelos cuya significación de parámetros puede apreciarse en la tabla siguiente:

SARIMA(1,0,0)(1,0,0)					SARIMA(0,0,1)(1,0,0)				
parámetro	valor	desvío	t-valor	p-valor	parámetro	valor	desvío	t-valor	p-valor
ϕ_1	0.40	0.08	5.11	9.41e-07	θ_1	0.26	0.07	3.57	4.79e-04
Φ_1	0.40	0.08	5.20	6.13e-07	Φ_2	0.34	0.08	4.08	7.27e-05
cte	1129.58	81.53	13.86	0.00e+00	cte	1134.88	58.50	19.40	0.00e+00
$tend$	4.99	0.88	5.70	5.96e-08	$tend$	3.38	0.99	3.41	8.18e-04
DC	20.03	5.30	3.78	2.21e-04	DC	20.28	5.89	3.44	7.42e-04
$AO_{set/05}$	-402.47	179.97	-2.24	2.68e-02	$LS_{jul/02}$	196.13	89.51	2.19	2.99e-02
$AO_{nov/05}$	-594.62	182.36	-3.26	1.37e-03	$IO_{set/05}$	-471.87	195.55	-2.41	1.70e-02
$AO_{mar/06}$	896.92	180.38	4.97	1.74e-06	$AO_{nov/05}$	-648.43	197.13	-3.29	1.24e-03
$AO_{feb/07}$	-624.78	178.71	-3.50	6.17e-04	$AO_{mar/06}$	888.21	193.44	4.59	9.07e-06
$AO_{set/07}$	484.51	184.44	2.63	9.49e-03	$AO_{feb/07}$	-643.92	191.67	-3.36	9.84e-04
					$AO_{set/07}$	544.26	200.99	2.71	7.54e-03
SARIMA(0,0,2)(1,0,0)									
parámetro	valor	desvío	t-valor	p-valor					
θ_1	0.32	0.08	3.92	1.36e-04					
θ_2	0.29	0.09	3.34	1.05e-03					
Φ_1	0.38	0.08	4.78	4.06e-06					
cte	1128.03	74.58	15.13	0.00e+00					
$tend$	4.99	0.80	6.22	4.42e-09					
DC	19.70	4.97	3.96	1.14e-04					
$AO_{feb/05}$	415.25	182.99	2.27	2.46e-02					
$AO_{set/05}$	-465.02	180.99	-2.57	1.11e-02					
$AO_{nov/05}$	-584.59	188.01	-3.11	2.24e-03					
$AO_{mar/06}$	847.25	176.37	4.80	3.68e-06					
$AO_{feb/07}$	-623.33	176.32	-3.54	5.39e-04					
$AO_{set/07}$	562.48	181.97	3.09	2.37e-03					

Cuadro 4.1: Significación de parámetros, modelos con intervenciones.

¹Se utiliza una función creada por Fernando Massa, ver código en anexo D.

Los tres modelos candidatos son sometidos al análisis del cumplimiento de supuestos así como a la performance que presentan en predicción.

4.3. Validación y Diagnóstico

Se realizan pruebas de incorrelación, media cero, homoscedasticidad y normalidad a los residuos de los modelos candidatos².

Para testear la incorrelación de los residuos se analiza el gráfico de los autocorrelogramas (simple y parcial) y se hacen pruebas de Ljung-Box y Box-Pierce. En el modelo $SARIMA(0, 0, 1)(1, 0, 0)$ el segundo rezago sobrasale de las líneas punteadas lo cual indica que no se ha captado toda la regularidad de la serie por lo cual este modelo es descartado. Los otros dos modelos dos verifican la incorrelación de los residuos por lo que se prueba para ellos el cumplimiento de los otros supuestos.

Para contrastar si la media es cero se efectúa la prueba explicitada en la metodología. No se rechaza para ninguno de los dos modelos la hipótesis de media cero.

Para determinar si los residuos poseen varianza constante, es decir si se cumple el supuesto de homoscedasticidad, se analiza el gráfico de los residuos de los modelos. No se observa aumento o disminución de la variabilidad con el tiempo por lo que se cumple este supuesto para ambos modelos.

Las dos pruebas de normalidad aplicadas a los modelos candidatos permiten concluir que se cumple el supuesto de normalidad de los residuos. A su vez los gráficos A.10 y A.11³ aportan evidencia para justificar la normalidad. El Q-Q plot (quantile-quantile plot) compara los cuantiles de la distribución empírica de los residuos con los teóricos de una distribución normal, por lo que si los residuos son normales se espera que los puntos se encuentren sobre la línea. La desviación de muchos puntos de la recta sugiere que los residuos no tienen distribución normal. El histograma de los residuos se utiliza para comparar su forma con la función de distribución de una variable normal. Ninguno de los gráficos anteriores aporta evidencia para descartar el supuesto de normalidad de los residuos para los modelos candidatos.

En función de los anterior los modelos candidatos verifican todos los supuestos respecto a los residuos, por lo que se los prueba en predicción.

4.4. Predicción

Para los modelos que verifican todos los supuestos se realizan predicciones a varios horizontes dentro de la muestra, con particiones de la serie adecuadas de forma de cubrir todo el año 2010. Se calculan errores cuadráticos medios (ECM), errores acumulados absolutos (EAA) y errores acumulados relativos (EAR) de la siguiente forma:

$$ECM_H(y) = \frac{\sum_{i=1}^H (y_i - \hat{y}_i)^2}{H}$$

$$EAA_H(y) = \sum_{i=1}^H (y_i - \hat{y}_i)$$

$$EAR_H(y) = \frac{\sum_{i=1}^H (y_i - \hat{y}_i)}{\sum_{i=1}^H y_i}$$

²Por resultados ver anexo A

³Ver anexo A.

donde H es el horizonte de predicción

La tabla (4.2) muestra un resumen de las predicciones.

horizonte	período	SARIMA(1,0,0)(1,0,0)			SARIMA(0,0,2)(1,0,0)		
		ECM	EAA	EAR(%)	ECM	EAA	EAR(%)
2 pasos	Ene/Feb-2010	3629.48	119.92	3.37	5309.73	144.41	4.06
2 pasos	Mar/Abr-2010	413.27	19.26	0.50	301.90	9.16	0.24
2 pasos	May/Jun-2010	3372.97	108.67	3.02	2820.17	102.16	2.84
2 pasos	Jul/Ago-2010	78570.51	508.37	12.52	74743.93	492.07	12.12
2 pasos	Set/Oct-2010	47074.83	37.63	1.00	35273.21	24.25	0.64
2 pasos	Nov/Dic-2010	112162.35	-391.84	-12.08	96532.65	-413.07	-12.74
3 pasos	Ene-Mar/2010	3485.62	176.47	3.16	4507.39	198.29	3.55
3 pasos	Abr-Jun/2010	1865.03	75.20	1.39	1569.29	69.26	1.28
3 pasos	Jul-Set/2010	82158.52	807.26	12.96	79406.64	789.95	12.68
3 pasos	Oct-Dic/2010	122929.14	-876.02	-18.04	106879.79	-808.06	-16.64
4 pasos	Ene/Abr-2010	2624.01	182.73	2.46	3380.58	198.70	2.68
4 pasos	May/Ago-2010	44759.52	643.31	8.40	43576.22	624.92	8.16
4 pasos	Set/Dic-2010	85903.20	-463.38	-6.60	80121.10	-475.10	-6.76
6 pasos	Ene-Jun/2010	3057.40	300.97	2.73	3287.36	305.07	2.77
6 pasos	Jul-Dic/2010	84219.71	184.49	1.66	83507.99	155.94	1.41
12 pasos	Ene-Dic/2010	44927.87	557.64	2.52	44944.11	537.94	2.43
6 v. 1 paso	Ene-Jun/2010	1919.72	-	-	2193.20	-	-
6 v. 1 paso	Jul-Dic/2010	90136.64	-	-	72321.68	-	-
12 v. 1 paso	Ene-Dic/2010	46028.18	-	-	37257.44	-	-

Cuadro 4.2: Errores cuadráticos medios y acumulados de las predicciones.

Los valores de $AICC$ son 2128.11 para $SARIMA(1, 0, 0)(1, 0, 0)$ y 2124.05 para $SARIMA(0, 0, 2)(1, 0, 0)$, lo cual indica que el segundo ajusta mejor. A su vez este es el que tiene menor error de predicción en la mayoría de los casos evaluados, por lo que se lo elije como modelo para predecir las ventas de ácido sulfúrico.

El modelo estimado para la serie de ventas en plaza de ácido sulfúrico es:

$$\hat{y}_t = 1128,03 + 4,99t + 0,38y_{t-12} + 0,32e_{t-1} + 0,29e_{t-2} + 19,70DC(t) + 415,25AO_{feb/05} - 465,02AO_{set/05} - 584,59AO_{nov/05} + 847,25AO_{mar/06} - 623,33AO_{feb/07} + 562,48AO_{set/07}$$

donde:

\hat{y}_t representa las ventas de ácido sulfúrico estimadas en el mes t

y_{t-12} representa las ventas reales del mes $t - 12$

e_{t-1} y e_{t-2} representan los errores de predicción de los meses $t - 1$ y $t - 2$ respectivamente

$DC(t)$ modela la cantidad de días hábiles que hay en el mes: $DC(t) =$ suma de días de lunes a viernes en mes t menos $5/2 \times$ (suma de sábados y domingos en mes t)

$AO_{set/05}(t)$ a $AO_{set/07}(t)$ son atípicos aditivos, valen uno en el mes indicado y cero en el resto

El modelo anterior se puede describir de la siguiente forma: se parte de 1128.03 toneladas a las cuales se suman 4.99 por cada mes adicional, a este valor se le suma 0.38 multiplicado por las

toneladas vendidas en el mismo mes del año anterior, a esto se suma 0.32 multiplicado por el error de predicción cometido en el mes anterior y 0.29 multiplicado por el error de predicción cometido dos meses antes. Al valor resultante se le agrega el producto de 19.70 por el valor de la variable que modela el efecto calendario correspondiente a ese mes. Si el mes a predecir es setiembre o noviembre de 2005, marzo de 2006, febrero o setiembre de 2007 se le suma o resta, según corresponda, la cantidad de toneladas que figura en 4.1.

Predicción a doce pasos

Se corta la serie en diciembre de 2009 y se estima el modelo seleccionado anteriormente. Con este modelo se realizan predicciones a doce pasos y se comparan los resultados con la serie real para el año 2010.

fecha	ventas	predicción	error absoluto	error relativo (%)	lim inf IC	lim sup IC
Ene/2010	1648.92	1566.91	82.01	4.97	1173.73	1960.10
Feb/2010	1911.89	1849.49	62.41	3.26	1434.48	2264.49
Mar/2010	2020.87	1966.99	53.88	2.67	1541.05	2392.93
Abr/2010	1832.65	1832.24	0.41	0.02	1406.30	2258.17
May/2010	1687.57	1617.88	69.69	4.13	1191.94	2043.82
Jun/2010	1905.38	1868.71	36.68	1.93	1442.77	2294.65
Jul/2010	2235.49	1842.44	393.04	17.58	1416.50	2268.38
Ago/2010	1825.99	1686.06	139.94	7.66	1260.12	2112.00
Set/2010	2166.68	1860.26	306.42	14.14	1434.32	2286.20
Oct/2010	1613.71	1765.15	-151.43	-9.38	1339.21	2191.09
Nov/2010	1841.45	1818.85	22.61	1.23	1392.91	2244.79
Dic/2010	1401.64	1879.35	-477.71	-34.08	1453.41	2305.29
Ene-Dic/2010	22092.24	21554.33	537.95	2.44	-	-

ECM = 44944.11

Cuadro 4.3: Predicción año 2010 con información hasta diciembre 2009.

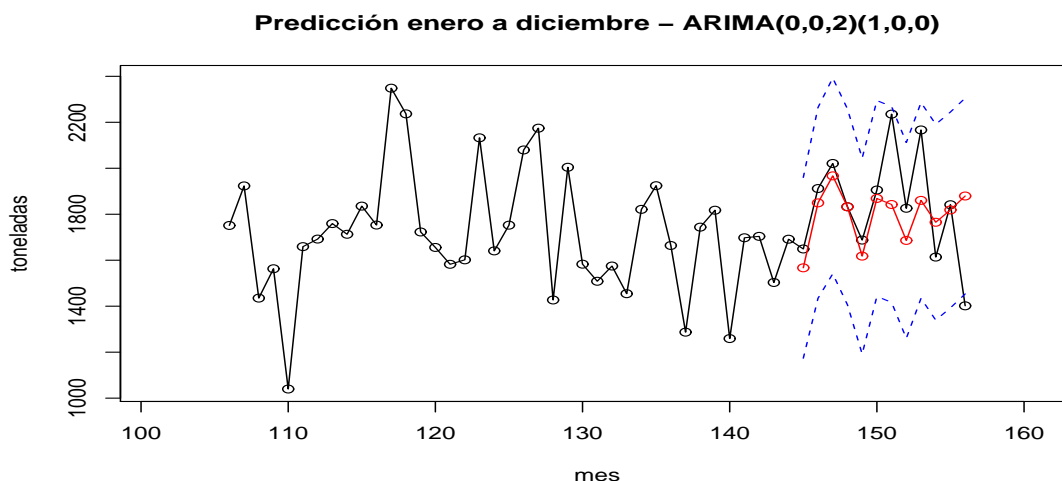


Figura 4.3: Predicción a doce pasos: año 2010 con información hasta diciembre 2009.

En general el modelo tiene buena performance ya que los errores de predicción en la mayoría de los meses son inferiores al 5%. En los meses de julio, setiembre y diciembre los errores de predicción son superiores al 10%.

Predicción doce veces a un paso

En este caso nuevamente se corta la serie en diciembre de 2009 y se estima el modelo con la serie reducida pero se realizan predicciones a un paso, luego se incorpora el valor observado en enero de 2010 y se realizan nuevamente predicciones a un paso. Así sucesivamente hasta cubrir las predicciones para todo el año 2010.

fecha	ventas	predicción	error absoluto	error relativo (%)	lim inf IC	lim sup IC
Ene/2010	1648.92	1566.91	82.01	4.97	1173.73	1960.10
Feb/2010	1911.89	1878.58	33.32	1.74	1486.48	2270.68
Mar/2010	2020.87	1999.52	21.34	1.06	1608.75	2390.30
Abr/2010	1832.65	1852.37	-19.72	-1.08	1462.92	2241.83
May/2010	1687.57	1621.97	65.60	3.89	1233.84	2010.09
Jun/2010	1905.38	1892.10	13.28	0.70	1505.09	2279.12
Jul/2010	2235.49	1870.24	365.24	16.34	1484.52	2255.97
Ago/2010	1825.99	1831.70	-5.71	-0.31	1442.98	2220.42
Set/2010	2166.68	1967.14	199.54	9.21	1579.70	2354.57
Oct/2010	1613.71	1857.95	-244.24	-15.14	1470.58	2245.32
Nov/2010	1841.45	1815.88	25.58	1.39	1427.92	2203.83
Dic/2010	1401.64	1849.27	-447.63	-31.94	1462.53	2236.00

$$ECM = 37257.44$$

Cuadro 4.4: Predicción doce veces a un paso: año 2010 con información hasta el paso anterior.

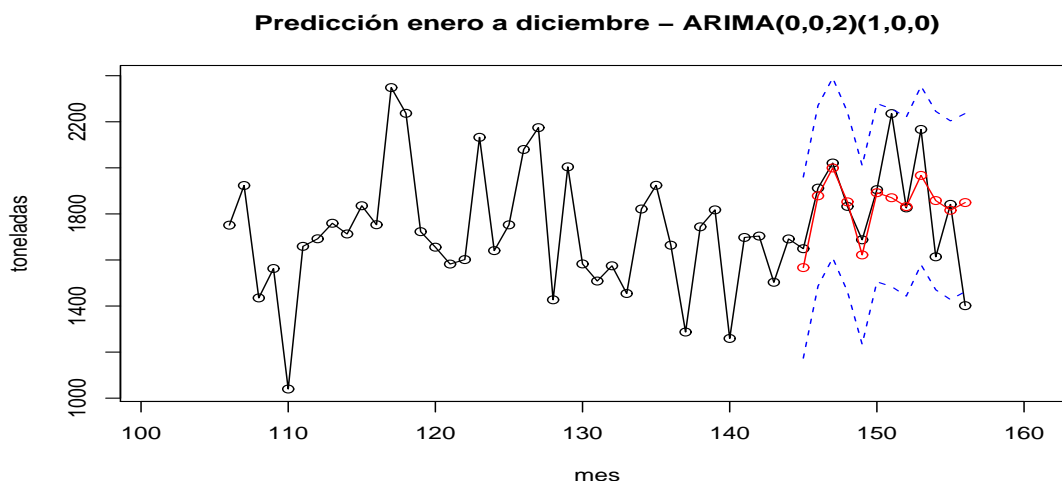


Figura 4.4: Predicción doce veces a un paso: año 2010 con información hasta el paso anterior.

Como puede apreciarse en los cuadros 4.3 y 4.4 al incorporar información del mes anterior se mejora la predicción en casi todos los meses, salvo en abril, octubre y noviembre.

Comentarios sobre el modelo

El modelo estimado para las ventas en plaza de ácido sulfúrico es eficiente para hacer pronósticos en la mayoría de los meses, según los resultados obtenidos en las predicciones de prueba; las realizadas con particiones de la muestra de forma de poder evaluarlas. Se eligió un modelo de tipo $SARIMA(0,0,2)(1,0,0)$, es decir, un modelo de medias móviles de orden dos para la parte regular y un autorregresivo de orden uno para la parte estacional. Esto indica que para la predicción de cierto mes se utiliza información de los dos meses anteriores y del mismo mes del año anterior, si estos datos no siguen la dinámica de la serie (son atípicos) es probable que las predicciones no sean precisas.

Los errores de predicción más grandes tienen lugar en los meses de julio, setiembre, octubre y diciembre de 2010, tanto en las predicciones a doce pasos como en las predicciones a un paso. Esto puede deberse a la presencia de atípicos en esos meses que no hayan sido modelizados o a una estructura estacional no incluida en el modelo.

Capítulo 5

Sulfato de aluminio

Al igual que en el análisis del capítulo anterior, este capítulo se divide en 4 secciones que abarcan las distintas etapas llevadas a cabo para obtener un modelo univariado, en esta oportunidad, para las ventas de sulfato de aluminio.

5.1. Identificación

La serie a modelizar corresponde a las ventas mensuales en plaza de sulfato de aluminio, medidas en toneladas y comprendidas entre enero de 1998 y diciembre de 2010. En la figura 5.1 se puede observar la representación gráfica de la misma.

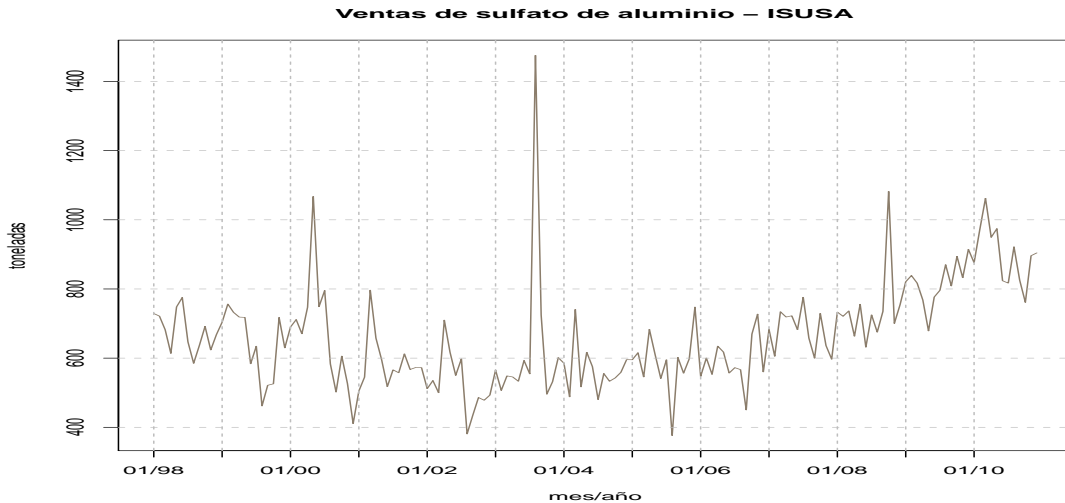


Figura 5.1: Gráfico de la serie de ventas mensuales en plaza de sulfato de aluminio.

Esta serie presenta algunas irregularidades y un aparente crecimiento a partir de 2007. Se aprecia la presencia de posibles atípicos, en particular un valor máximo que tiene lugar en 2003. En los autocorrelogramas que se muestran en la figura 5.2 se puede apreciar muchos rezagos por fuera de las bandas en el ACF y dos en el PACF. Esto indica que la serie tiene memoria larga y podría tener una raíz unitaria.

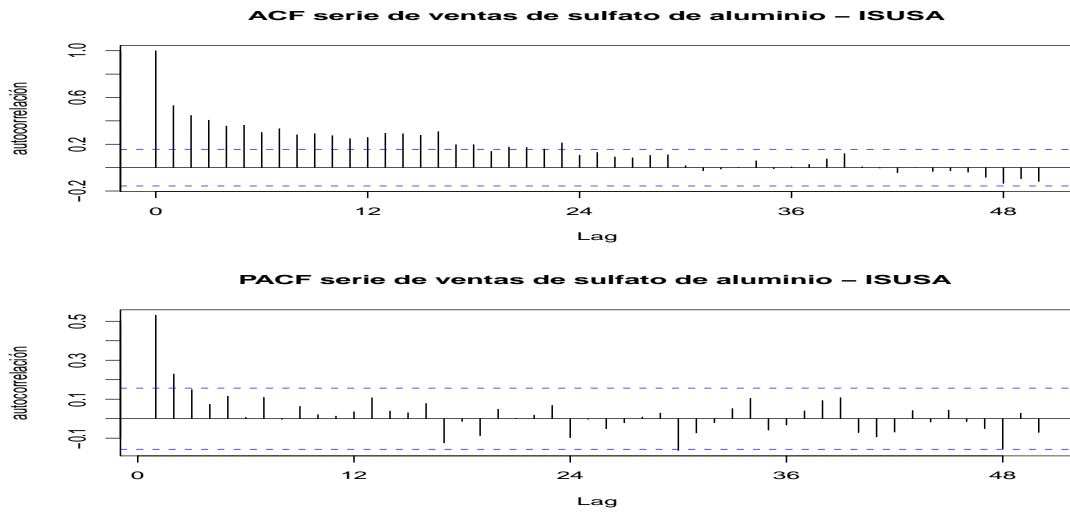


Figura 5.2: Autocorrelogramas de la serie de ventas mensuales de sulfato de aluminio.

Se investiga la conveniencia de transformar la serie en logaritmos obteniendo como resultado que no es necesario transformar la serie.

Se aplica la prueba de Dickey Fuller aumentado para determinar si la serie presenta raíces unitarias. No se rechaza la hipótesis nula de raíz unitaria al 5% de significación, por lo que no se descarta la presencia de una raíz unitaria. Luego se testea la existencia de una segunda raíz unitaria en el proceso, en este caso se rechaza la hipótesis nula por lo que se descarta la presencia de una segunda raíz unitaria. Lo anterior permite concluir que la serie de ventas en plaza de sulfato de aluminio es integrada de orden 1 ($I(1)$).

Se modelizará la serie en primeras diferencias, que se muestra en la siguiente figura.

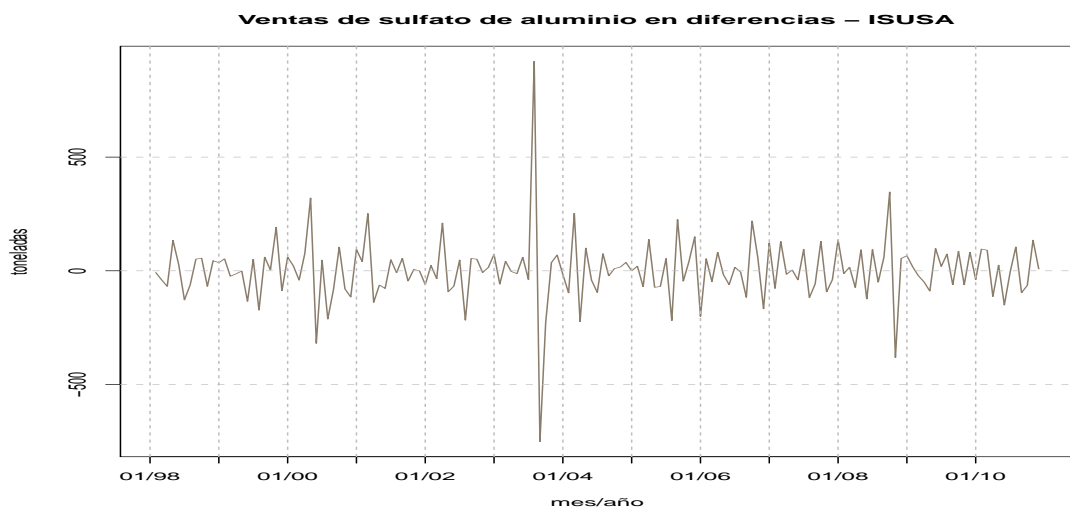


Figura 5.3: Gráfico de la serie de ventas mensuales en plaza de sulfato de aluminio en diferencias.

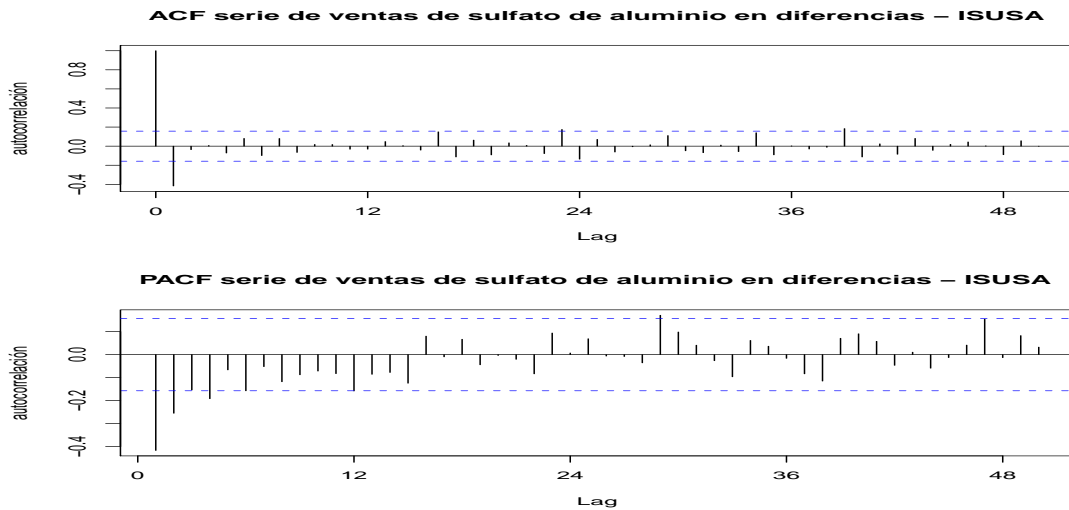


Figura 5.4: Autocorrelogramas simple y parcial de la serie de ventas de sulfato de aluminio en diferencias.

Al diferenciar la serie se obtiene otra serie con media constante, se elimina el efecto de crecimiento observado en la serie original. En los autocorrelogramas de la serie en diferencias (figura 5.4) se observa el primer rezago por fuera de las bandas en el *ACF* y el primero, segundo y cuarto fuera de las bandas en el *PACF*. Esta estructura sugiere que la serie podría incluir una parte *MA* de orden uno y una parte *AR* de orden dos o cuatro. En función de lo anterior se decide estimar inicialmente los siguientes modelos:

- *ARIMA*(0, 1, 1)
- *ARIMA*(2, 1, 1)
- *ARIMA*(4, 1, 1)

5.2. Estimación

Tal como se mencionó anteriormente la serie es integrada de orden 1 por lo que la tendencia es estocástica y no determinística. Teniendo en cuenta lo anterior se estiman los modelos identificados anteriormente, agregándole variables regresoras que modelicen eventos determinísticos tales como feriados no laborables, turismo y efecto calendario. Ninguna de estas variables resulta significativa, por lo que no se incluyen en los modelos.

Se descartan modelos que no tengan todos los parámetros significativos. Los modelos posibles son:

- *ARIMA*(0, 1, 1)
- *ARIMA*(1, 1, 1)
- *ARIMA*(0, 1, 2)

El último modelo, que tiene en cuenta un *MA*(2) no surgió de la primera visualización de los autocorrelogramas sino del análisis de los residuos de los otros modelos.

En primera instancia se identifican 3 atípicos para los modelos anteriores:

- Atípico atidivo (AO^1) en mayo de 2000.
- Atípico atidivo (AO) en agosto de 2003.
- Atípico aditivo (AO) en octubre de 2008.

Al incluirse los 3 atípicos, el modelo $ARIMA(0, 1, 2)$ presenta el parámetro θ_2 no significativo. Los otros dos modelos tienen todos los parámetros significativos.

Como resultado de estimar los modelos intervenidos con las variables que modelan los atípicos detectados, se obtienen residuos que no verifican todos los supuestos para ser considerados ruido blanco (a pesar de tener todos los parámetros significativos), en particular no verifican la incorrelación de los residuos al cuadrado. El análisis visual de los residuos y sus autocorrelogramas sugiere que no se han captado todos los atípicos presentes en la serie.

Frente a esta situación, se procede a realizar la detección y estimación de nuevos atípicos. Se analizan los residuos de cada modelo en busca de observaciones extremas (muy grandes o muy chicas). Se estudia el comportamiento de la serie de residuos a lo largo del tiempo, para detectar cambios de nivel o transitorios. Se encuentran picos que se intervienen como atípicos aditivos, incorporándose de a uno en los modelos y analizando como se modifican los residuos. El procedimiento termina cuando se logra que los residuos al cuadrado cumplan incorrelación, bajo la condición de que se cumplan también las otras propiedades para que sean ruido blanco y todos los parámetros sean significativos.

Se decide incorporar a los modelos las siguientes intervenciones:

- Atípico aditivo (AO) en marzo de 2001.
- Atípico aditivo (AO) en agosto de 2002.
- Atípico aditivo (AO) en agosto de 2005.

La significación de los modelos con estas nuevas intervenciones se muestra en el cuadro 5.1. La figura que se muestra a continuación indica la ubicación de los atípicos en la serie de ventas de sulfato de aluminio.

¹Atípico aditivo, del inglés: Additive Outlier.

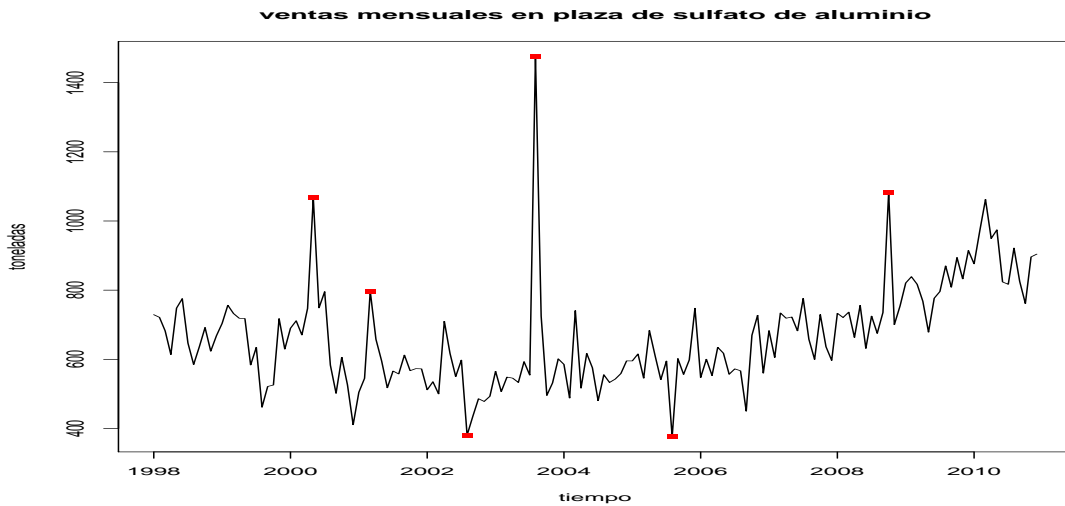


Figura 5.5: Observaciones intervenidas en la serie de ventas de sulfato de aluminio.

Se vuelve a probar la significación de las variables que modelan el efecto calendario, esta vez aplicadas a los modelos con seis intervenciones. No se obtiene ninguna variable significativa, por lo que no se incluye ese efecto en los modelos.

ARIMA(0,1,1)					ARIMA(1,1,1)				
parámetro	valor	desvío	t-valor	p-valor	parámetro	valor	desvío	t-valor	p-valor
θ_1	-0.64	0.08	-7.55	3.63e-12	ϕ_1	0.20	0.13	1.55	1.23e-01
$AO_{may/00}$	380.33	69.34	5.48	1.66e-07	θ_1	-0.79	0.08	-9.39	0.00e+00
$AO_{mar/01}$	233.77	66.98	3.49	6.30e-04	$AO_{may/00}$	375.75	68.62	5.48	1.75e-07
$AO_{ago/02}$	-152.40	67.10	-2.27	2.45e-02	$AO_{mar/01}$	213.53	67.85	3.15	1.98e-03
$AO_{ago/03}$	894.59	67.37	13.28	0.00e+00	$AO_{ago/02}$	-150.86	67.03	-2.25	2.58e-02
$AO_{ago/05}$	-218.75	66.99	-3.27	1.35e-03	$AO_{ago/03}$	877.55	68.26	12.86	0.00e+00
$AO_{oct/08}$	349.73	67.06	5.22	5.83e-07	$AO_{ago/05}$	-220.56	66.92	-3.30	1.22e-03
					$AO_{oct/08}$	351.25	67.00	5.24	5.19e-07
ARIMA(0,1,2)									
parámetro	valor	desvío	t-valor	p-valor					
θ_1	-0.61	0.08	-7.49	5.10e-12					
θ_2	-0.10	0.09	-1.12	2.66e-01					
$AO_{may/00}$	381.85	68.46	5.58	1.08e-07					
$AO_{mar/01}$	219.23	68.11	3.22	1.57e-03					
$AO_{ago/02}$	-152.36	67.25	-2.27	2.49e-02					
$AO_{ago/03}$	881.88	68.56	12.86	0.00e+00					
$AO_{ago/05}$	-222.27	67.25	-3.31	1.18e-03					
$AO_{oct/08}$	349.63	67.22	5.20	6.26e-07					

Cuadro 5.1: Significación de parámetros de modelos con seis intervenciones, sulfato de aluminio.

Al incorporar estas intervenciones, los parámetros ϕ_1 del modelo $ARIMA(1, 1, 1)$ y θ_2 del modelo $ARIMA(0, 1, 2)$ resultan no significativos con p -valor de 0.123 y 0.266 respectivamente. Se rechazan estos modelos siendo el único candidato el $ARIMA(0, 1, 1)$.

5.3. Validación y Diagnóstico

En esta etapa se verifica el cumplimiento de los supuestos respecto a los residuos del modelo $ARIMA(0, 1, 1)$: incorrelación, media cero, homoscedasticidad y normalidad.

En el gráfico de los autocorrelogramas de los residuos se puede observar que se captó la dinámica si bien todavía quedan algunos rezagos significativos (ver figura A.16 en A). Se obtiene en las pruebas de incorrelación de Ljung-Box y Box-Pierce que se cumple el supuesto de incorrelación. En la prueba de media cero tampoco se rechaza la hipótesis nula. Con respecto a la homoscedasticidad se observa el gráfico de los residuos para determinar si se perciben cambios de varianza. No se aprecian cambios en la variabilidad de los residuos, por lo que se verifica también este supuesto. El supuesto de normalidad se verifica tanto en la prueba de Jarque-Bera como en la de Shapiro-Wilks.

El modelo estimado para la serie de ventas de sulfato de aluminio es:

$$\hat{y}_t = y_{t-1} - 0,64e_{t-1} + 380,33AO_{may/00}(t) + 233,77AO_{mar/01}(t) - 152,40AO_{ago/02}(t) + 894,59AO_{ago/03}(t) - 218,75AO_{ago/05}(t) + 349,73AO_{oct/08}(t)$$

donde:

\hat{y}_t representa las ventas estimadas de sulfato de aluminio en el mes t

y_{t-1} representa las ventas reales de sulfato de aluminio en el mes $t - 1$

e_{t-1} es el error de predicción a un paso en el mes $t - 1$, $e_{t-1} = y_{t-1} - \hat{y}_{t-1}$

$AO_{may/00}(t)$ a $AO_{oct/08}(t)$ son atípicos aditivos, valen uno en el mes indicado y cero en el resto

5.4. Predicción

En esta etapa se prueba la performance del modelo candidato en predicción.

Se realizan predicciones dentro de la muestra con distintos cortes y se calculan errores cuadráticos medios, errores acumulados absolutos y errores acumulados relativos.

Predicción a doce pasos

fecha	ventas	predicción	error absoluto	error relativo (%)	lim inf IC	lim sup IC
Ene/2010	875.97	860.1	15.87	1.81	717.45	1002.74
Feb/2010	971.62	860.1	111.52	11.48	710.03	1010.16
Mar/2010	1062.19	860.1	202.09	19.03	702.95	1017.24
Abr/2010	949.38	860.1	89.28	9.40	696.18	1024.01
May/2010	974.36	860.1	114.26	11.73	689.68	1030.51
Jun/2010	823.69	860.1	-36.41	-4.42	683.42	1036.77
Jul/2010	816.98	860.1	-43.12	-5.28	677.37	1042.82
Ago/2010	921.84	860.1	61.74	6.70	671.51	1048.68
Set/2010	825.41	860.1	-34.68	-4.20	665.84	1054.35
Oct/2010	761.33	860.1	-98.77	-12.97	660.32	1059.87
Nov/2010	896.14	860.1	36.05	4.02	654.95	1065.24
Dic/2010	904.24	860.1	44.14	4.88	649.72	1070.47
Ene-Dic/2010	10783.15	10321.20	461.97	4.28 -	-	-

ECM = 7980.041

Cuadro 5.2: Predicción año 2010 con información hasta diciembre 2009.

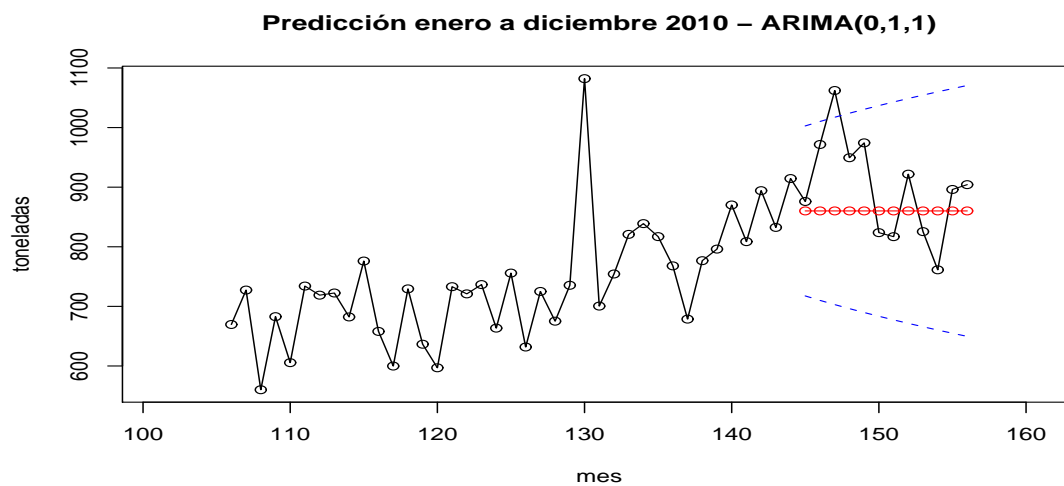


Figura 5.6: Predicción a doce pasos: año 2010 con información hasta diciembre 2010.

Predicción doce veces a un paso

fecha	ventas	predicción	error absoluto	error relativo (%)	lim inf IC	lim sup IC
Ene/2010	875.97	860.1	15.87	1.81	717.45	1002.74
Feb/2010	971.62	865.67	105.95	10.90	723.5	1007.84
Mar/2010	1062.19	903.52	158.67	14.94	760.79	1046.24
Abr/2010	949.38	967.34	-17.96	-1.89	822.84	1111.84
May/2010	974.36	959.75	14.61	1.5	815.72	1103.78
Jun/2010	823.69	965.47	-141.78	-17.21	821.90	1109.03
Jul/2010	816.98	913.25	-96.28	-11.78	768.41	1058.09
Ago/2010	921.84	877.38	44.46	4.82	732.20	1022.56
Set/2010	825.41	894.16	-68.75	-8.33	749.31	1039.02
Oct/2010	761.33	869.03	-107.70	-14.15	724.24	1013.83
Nov/2010	896.14	827.53	68.61	7.66	682.20	972.86
Dic/2010	904.24	854.01	50.23	5.55	708.78	999.24

ECM = 7674.498

Cuadro 5.3: Predicción doce veces a un paso: año 2010 con información hasta el paso anterior.

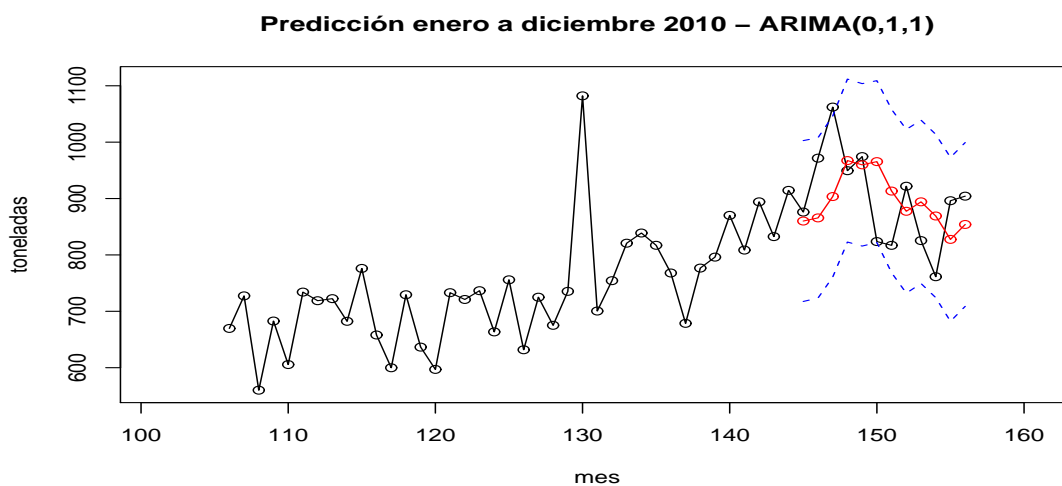


Figura 5.7: Predicción doce veces a un paso: año 2010 con información hasta el paso anterior.

Al igual que en las predicciones de ácido sulfúrico, en este caso la predicción mejora al incorporar información hasta el mes anterior, pero esto ocurre fundamentalmente entre los meses de enero a mayo, mientras que en el resto de los meses la predicción empeora al incorporar información del mes anterior.

Comentarios sobre el modelo

El modelo elegido para explicar las ventas en plaza de sulfato de aluminio en diferencias es de tipo $MA(1)^2$, esto quiere decir que las ventas de un mes se estiman únicamente utilizando información de las ventas del mes anterior. La limitación de este tipo de modelos es que si sólo se conocen los valores de la serie hasta el momento t , las predicciones para los momentos $t+1, t+2, \dots, t+k, \dots$ serán idénticas, ya que los valores reales de los períodos futuros se estiman por sus predicciones. Ante esta situación, se obtienen mejores predicciones tomando horizontes de corto plazo (1 o 2 meses) y actualizando la serie a medida que llegan nuevas observaciones.

A horizontes mayores a dos o tres pasos es posible que las predicciones puntuales no sean precisas, pero las pruebas realizadas dentro de la muestra indican que el modelo es más eficiente para predecir el total del volumen de ventas esperado de cierto período futuro, por ejemplo un semestre o un año.

²Medias móviles de orden uno, del inglés: Moving Average.

Capítulo 6

Fertilizantes

6.1. Modelo univariado

Se contemplan nuevamente las etapas seguidas en los dos capítulos anteriores.

Cabe señalar que esta serie es especial en el sentido de que incluye las ventas de todos los tipos de fertilizantes que se comercializan en plaza. Esto es, se consideran tanto fertilizantes sólidos como líquidos, embolsados o a granel, de una fórmula pura o combinada.

Esta heterogeneidad en la definición de la serie puede causar problemas al momento de encontrar un modelo adecuado para pronosticar las ventas de fertilizantes.

6.1.1. Identificación

La serie a modelizar corresponde a las ventas mensuales en plaza de fertilizantes, medidas en toneladas y comprendidas entre enero de 1998 y diciembre de 2010.

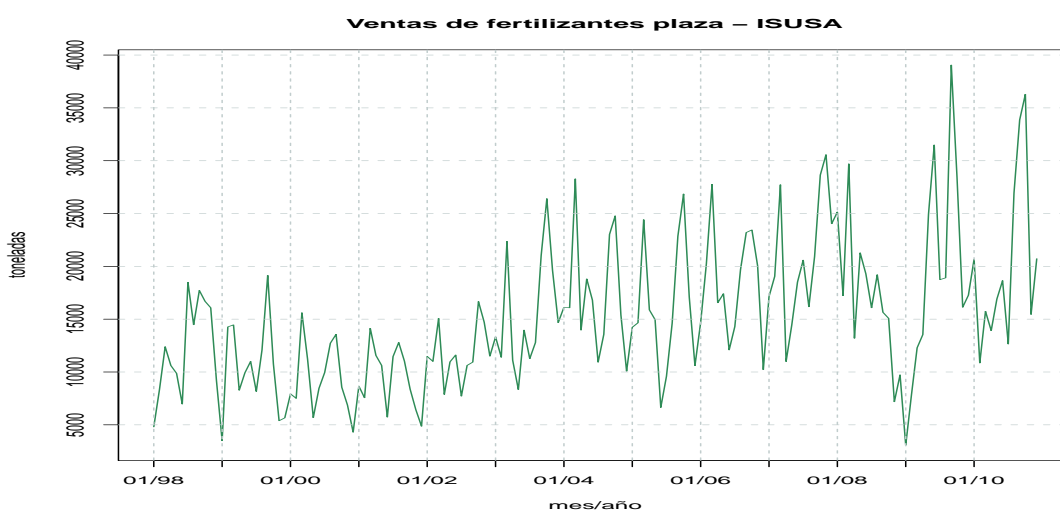


Figura 6.1: Gráfico de la serie de ventas mensuales de fertilizantes.

Esta serie no parece estable en media, se puede observar un cambio de pendiente a principios de 2003. A partir de 2008 se aprecia un comportamiento irregular con algunos picos de ventas.

Hasta el año 2003 las ventas de fertilizantes variaban entorno a las 10 mil toneladas, mientras

que a partir de 2003 la media se situó en 15 mil toneladas mensuales aproximadamente.

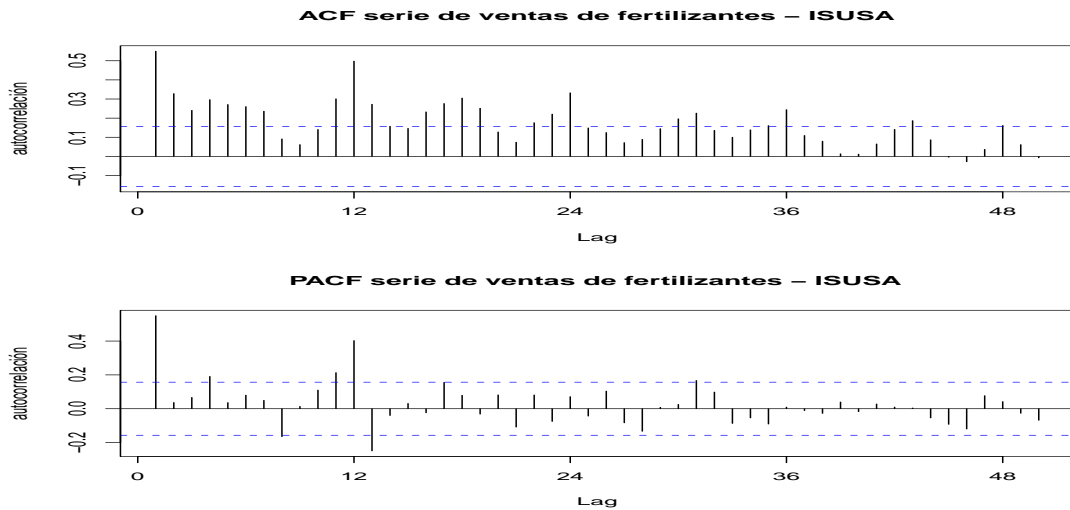


Figura 6.2: Autocorrelogramas de la serie de ventas mensuales de fertilizantes.

En el *ACF* se observan rezagos fuera de las bandas en los múltiplos de 12, lo cual indica la presencia de estacionalidad en la serie. Estos a su vez decrecen a medida que aumenta la cantidad de rezagos, esto podría indicar una parte *AR* estacional. Alrededor de los rezagos estacionales se observa un rápido decrecimiento de las autocorrelaciones que junto al decrecimiento de los primeros rezagos regulares sugiere una estructura *AR* regular.

En el *PACF* se observan fundamentalmente los rezagos 1, 12 y 13 fuera de las bandas lo cual podría indicar un *AR*(2) en la parte estacional y también un *AR*(1) en la parte regular.

Mediante la prueba para testear la conveniencia de transformar la serie en logaritmos se determina que no es necesario realizar dicha transformación.

En cuanto a la estacionariedad, se efectúa el contraste Dickey Fuller aumentado, obteniéndose que el *PGD*¹ de la serie es integrado de orden uno.

El gráfico de la serie de ventas de fertilizantes en diferencias y los autocorrelogramas se exhiben en las figuras (6.3) y (6.4) respectivamente.

¹Proceso generador de los datos.

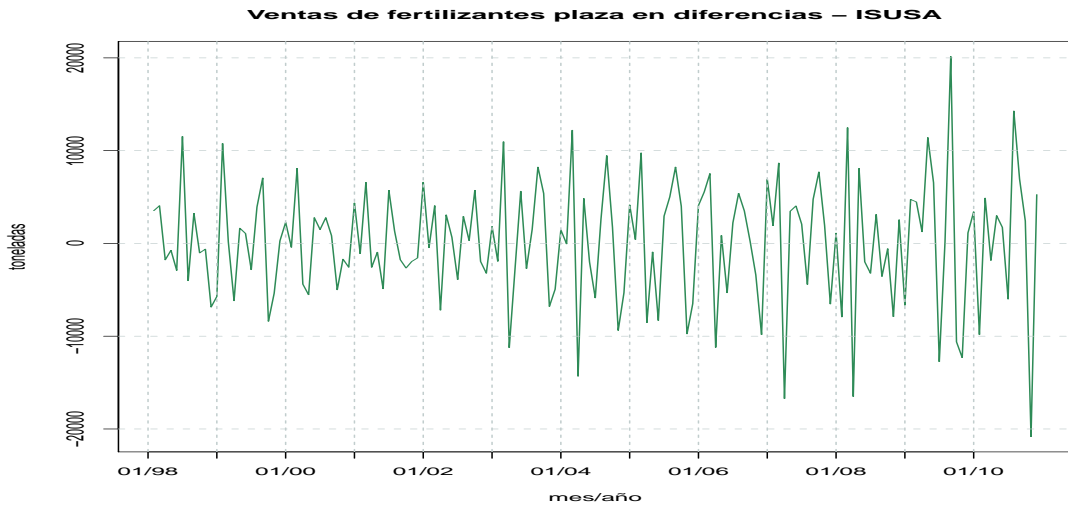


Figura 6.3: Gráfico de la serie de ventas mensuales de fertilizantes en diferencias.

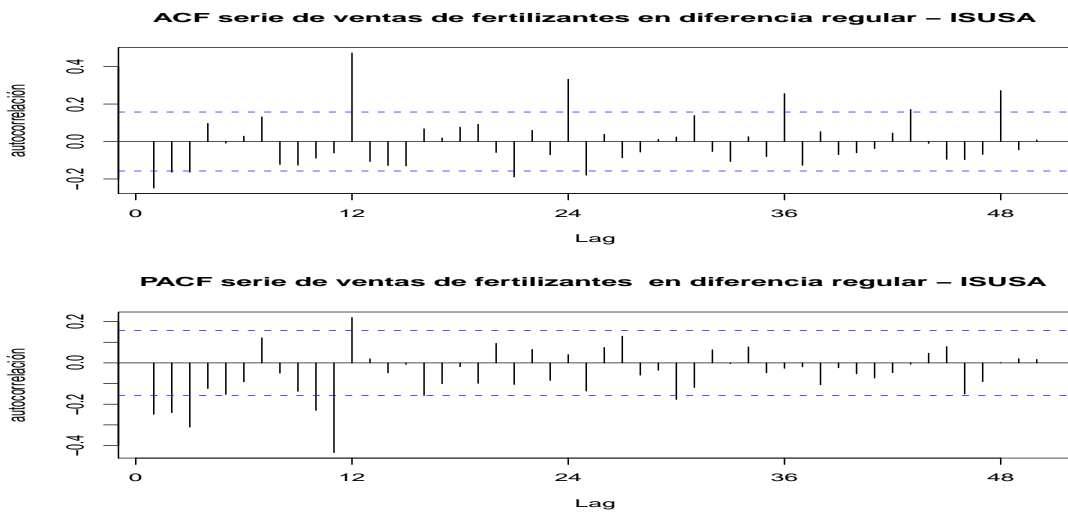


Figura 6.4: Autocorrelogramas de la serie de ventas mensuales de fertilizantes en diferencias.

Las autocorrelaciones altas en los rezagos múltiplo de 12 del autocorrelograma simple (figura 6.4) sugieren la necesidad de tomar la primera diferencia estacional para obtener una serie estacionaria. En los gráficos 6.5 y 6.6 se muestra el gráfico y los autocorrelogramas para esta serie, respectivamente.

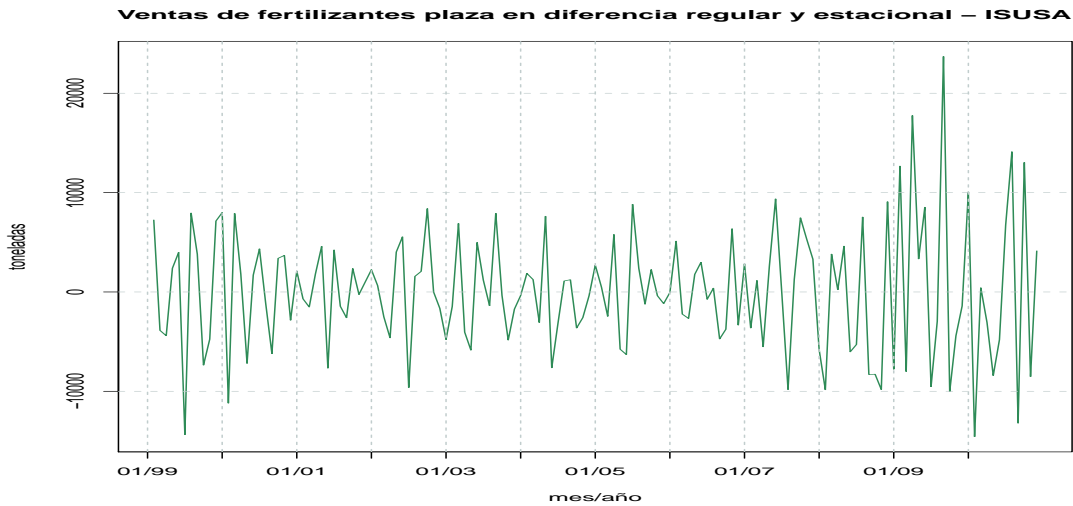


Figura 6.5: Gráfico de la primera diferencia estacional de la serie de ventas mensuales de fertilizantes en diferencias.

Esta serie es estable en media, pero puede no serlo en varianza. Sobre el final se observa que la variabilidad es mayor que al principio, esto puede deberse a presencia de atípicos que sea necesario intervenir.

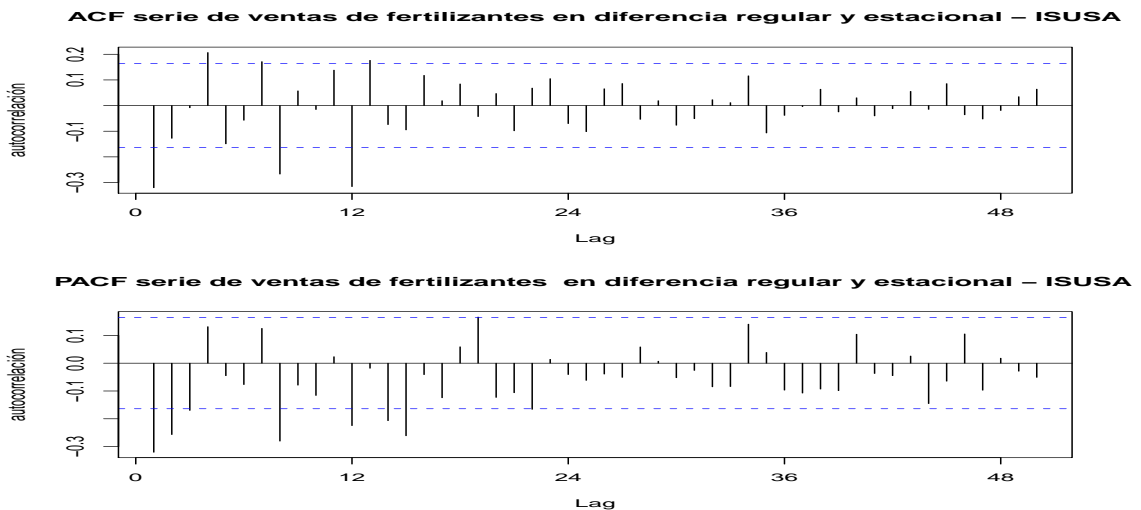


Figura 6.6: Autocorrelogramas de la primera diferencia estacional de la serie de ventas mensuales de fertilizantes en diferencias.

La estructura que muestran los autocorrelogramas es compleja, el primer rezago fuera de las bandas en al ACF y los dos primeros en el $PACF$ sugieren una estructura regular $MA(1)$, $ARMA(2, 1)$ o $ARMA(3, 1)$ (el tercer rezago del $PACF$ es levemente inferior al límite punteado). En lo que respecta a la estructura estacional, solamente el rezago 12 sobresale de las bandas tanto en ACF como en $PACF$, por lo que se consideran los siguientes modelos para la parte estacional: $AR(1)$, $MA(1)$ y $ARMA(1, 1)$.

Los modelos posibles son:

- $SARIMA(0, 1, 1)(0, 1, 1)$
- $SARIMA(0, 1, 1)(1, 1, 0)$
- $SARIMA(0, 1, 1)(1, 1, 1)$
- $SARIMA(3, 1, 1)(0, 1, 1)$
- $SARIMA(3, 1, 1)(1, 1, 0)$
- $SARIMA(3, 1, 1)(1, 1, 1)$

6.1.2. Estimación

Se estiman los modelos inicialmente identificados, luego se estiman otros modelos como resultado de extraer los parámetros no significativos en los modelos estimados al principio. Se obtienen los siguientes modelos posibles:

- $SARIMA(0, 1, 1)(0, 1, 1)$
- $SARIMA(0, 1, 1)(1, 1, 0)$
- $SARIMA(0, 1, 1)(1, 1, 1)$
- $SARIMA(0, 1, 2)(0, 1, 1)$
- $SARIMA(0, 1, 2)(1, 1, 0)$
- $SARIMA(1, 1, 1)(0, 1, 1)$
- $SARIMA(1, 1, 1)(1, 1, 0)$
- $SARIMA(3, 1, 0)(0, 1, 1)$
- $SARIMA(3, 1, 0)(1, 1, 0)$

En principio se realiza una detección automática de atípicos, la cual arroja los siguientes resultados:

- Para $SARIMA(0, 1, 1)(0, 1, 1)$: TC^2 en noviembre de 2007, LS en abril de 2009, SLS en setiembre de 2009.
- Para $SARIMA(0, 1, 1)(1, 1, 0)$: LS en noviembre de 2007, TC en abril de 2009.
- Para $SARIMA(0, 1, 1)(1, 1, 1)$: LS en noviembre de 2007, TC en abril de 2009, AO en setiembre de 2009.
- Para $SARIMA(0, 1, 2)(0, 1, 1)$: TC en noviembre de 2007, TC en setiembre de 2008, AO en enero de 2009 y SO en marzo de 2009.
- Para $SARIMA(0, 1, 2)(1, 1, 0)$: SO en noviembre de 2007, TC en setiembre de 2008, AO en enero de 2009, TC en marzo de 2009.

²Ver definiciones de clases de atípicos en capítulo 3.

- Para $SARIMA(1, 1, 1)(0, 1, 1)$: TC en noviembre de 2007, TC en setiembre de 2008.
- Para $SARIMA(1, 1, 1)(1, 1, 0)$: TC en noviembre de 2007, SO en octubre de 2008, AO en enero 2009, TC en marzo de 2009, AO en setiembre de 2009.
- Para $SARIMA(3, 1, 0)(0, 1, 1)$: LS en setiembre de 2008, LS en abril de 2009.
- Para $SARIMA(3, 1, 0)(1, 1, 0)$: LS en setiembre de 2008, AO en enero de 2009, TC en abril de 2009, AO en setiembre de 2009.

El gráfico de los residuos de los modelos con estas intervenciones figuran en A.21 (ver anexo A). Los residuos de todos los modelos anteriores presentan irregularidades, sobre todo hacia el final de la serie, que generan la duda sobre el cumplimiento del supuesto de homoscedasticidad y pueden provocar errores grandes al generar predicciones. Por lo anterior se cree conveniente realizar otras intervenciones además de las encontradas de forma automática.

Se analizan los residuos de forma de detectar puntos extremos o cambios en el nivel de la serie, luego de identificar la posición de los posibles atípicos se estiman modelos, determinando previamente el tipo de atípico o estimando para un modelo y una cierta ubicación varias posibilidades. Como resultado de mantener únicamente los modelos que presentan todos los parámetros significativos, se obtienen los siguientes candidatos:

- $SARIMA(0, 1, 1)(0, 1, 1)$: AO en junio 2005, TC noviembre 2007, LS abril 2009, AO junio 2009, SLS setiembre de 2009, AO agosto 2010, AO octubre 2010.
- $SARIMA(0, 1, 1)(1, 1, 0)$: AO en julio 1999, AO en junio 2005, LS en noviembre de 2007, TC en abril de 2009, AO en junio de 2009, AO en setiembre de 2009, TC en agosto de 2010.
- $SARIMA(0, 1, 1)(1, 1, 1)$: AO en junio 2005, LS en noviembre de 2007, TC en abril de 2009, AO en junio de 2009, AO en setiembre de 2009, TC en agosto 2010.
- $SARIMA(0, 1, 2)(0, 1, 1)$: AO en julio de 1999, AO en junio de 2005, TC en noviembre de 2007, TC en setiembre de 2008, AO en enero de 2009, SO en marzo de 2009, AO en junio de 2009, AO en setiembre de 2009, AO en agosto de 2010.
- $SARIMA(0, 1, 2)(1, 1, 0)$: AO en julio de 1999, AO en junio de 2005, SO en noviembre de 2007, TC en setiembre de 2008, AO en enero de 2009, TC en marzo de 2009, AO en junio de 2009, AO en setiembre de 2009, AO en agosto de 2010.
- $SARIMA(3, 1, 0)(0, 1, 1)$: IO en julio de 1999, AO en junio de 2005, LS en setiembre de 2008, LS en abril de 2009, AO en junio 2009, AO en setiembre de 2009, AO en agosto de 2010.
- $SARIMA(3, 1, 0)(1, 1, 0)$: IO en julio de 1999, AO junio de 2005, LS en setiembre de 2008, IO en noviembre de 2008, AO en enero de 2009, TC en abril de 2009, AO en junio de 2009, AO en setiembre de 2009, AO en agosto de 2010.

La significación de los parámetros de estos modelos puede verse en el cuadro (6.1)

SARIMA(0,1,1)(0,1,1)				SARIMA(0,1,1)(1,1,0)					
parámetro	valor	desvío	t-valor	p-valor	parámetro	valor	desvío	t-valor	p-valor
θ_1	-0.73	0.07	-10.58	0.00e+00	θ_1	-0.76	0.06	-12.00	0.00e+00
Θ_2	-0.53	0.12	-4.42	1.92e-05	Φ_1	-0.28	0.09	-3.15	1.99e-03
$AO_{jun/05}$	-8284	3162.5	-2.62	9.77e-03	$AO_{jul/99}$	-6602	3004.8	-2.20	2.96e-02
$TC_{nov/07}$	14955	3025.4	4.94	2.15e-06	$AO_{jun/05}$	-7751	2972.3	-2.61	1.01e-02
$LS_{abr/09}$	11995	2555.4	4.69	6.28e-06	$LS_{nov/07}$	12153	2258.7	5.38	3.02e-07
$AO_{jun/09}$	9976	3353.5	2.97	3.45e-03	$TC_{abr/09}$	16057	2751.0	5.84	3.50e-08
$SLS_{set/09}$	15119	3276	4.62	8.75e-06	$AO_{jun/09}$	8650	3124.2	2.77	6.38e-03
$AO_{ago/10}$	9823	3668.5	2.68	8.30e-03	$AO_{set/09}$	14486	3123.8	4.64	7.98e-06
$AO_{oct/10}$	11837	3717.6	3.18	1.79e-03	$TC_{ago/10}$	15232	3724.6	4.09	7.23e-05
SARIMA(0,1,1)(1,1,1)				SARIMA(0,1,2)(0,1,1)					
parámetro	valor	desvío	t-valor	p-valor	parámetro	valor	desvío	t-valor	p-valor
θ_1	-0.77	0.06	-12.35	0.00e+00	θ_1	-0.57	0.11	-5.03	1.51e-06
Φ_1	0.39	0.15	2.63	9.50e-03	θ_2	-0.30	0.12	-2.55	1.20e-02
Θ_1	-0.86	0.14	-6.03	1.36e-08	Θ_1	-0.37	0.11	-3.42	8.31e-04
$AO_{jun/05}$	-7339	3147.7	-2.33	2.12e-02	$AO_{jul/99}$	-5662	2709.5	-2.09	3.84e-02
$LS_{nov/07}$	11480	2369.1	4.85	3.31e-06	$AO_{jun/05}$	-6447	2736.5	-2.36	1.99e-02
$TC_{abr/09}$	16133	2918.5	5.53	1.54e-07	$TC_{nov/05}$	12872	2656.6	4.85	3.31e-06
$AO_{jun/09}$	9563	3264	2.93	3.96e-03	$TC_{set/08}$	-13385	2746.5	-4.87	2.93e-06
$AO_{set/09}$	14609	3218.8	4.54	1.21e-05	$AO_{ene/09}$	-11826	2709.4	-4.36	2.45e-05
$TC_{ago/10}$	13923	3498.4	3.98	1.10e-04	$SO_{mar/09}$	-11707	2910.5	-4.02	9.39e-05
					$AO_{jun/09}$	9584	2911.9	3.29	1.26e-03
					$AO_{set/09}$	10036	2929.2	3.43	8.03e-04
					$AO_{ago/10}$	8534	3401.1	2.51	1.32e-02

SARIMA(0,1,2)(1,1,0)					SARIMA(3,1,0)(0,1,1)				
parámetro	valor	desvío	t-valor	p-valor	parámetro	valor	desvío	t-valor	p-valor
θ_1	-0.55	0.11	-4.93	2.28e-06	ϕ_1	-0.50	0.08	-5.91	2.49e-08
θ_2	-0.33	0.12	-2.80	5.82e-03	ϕ_2	-0.49	0.09	-5.72	6.23e-08
Φ_1	-0.26	0.09	-2.80	5.79e-03	ϕ_3	-0.29	0.09	-3.34	1.08e-03
$AO_{jul/99}$	-6136	2572.4	-2.39	1.84e-02	Θ_1	-0.50	0.11	-4.63	8.24e-06
$AO_{jun/05}$	-6334	2638.9	-2.40	1.77e-02	$IO_{jul/99}$	-12055	4515.9	-2.67	8.50e-03
$SO_{nov/07}$	-12216	2989.3	-4.09	7.35e-05	$AO_{jun/05}$	-6905	2939.3	-2.35	2.02e-02
$TC_{set/08}$	12591	2579.8	4.88	2.84e-06	$LS_{set/08}$	-12663	2786.6	-4.54	1.19e-05
$AO_{ene/09}$	-12068	2584.9	-4.67	7.03e-06	$LS_{abr/09}$	14084	2922.2	4.82	3.72e-06
$TC_{mar/09}$	-13215	2664.9	-4.96	2.02e-06	$AO_{jun/09}$	8882	3173.7	2.80	5.86e-03
$AO_{jun/09}$	9108	2763.1	3.30	1.24e-03	$AO_{set/09}$	9442	3135.3	3.01	3.09e-03
$AO_{set/09}$	10070	2866.7	3.51	5.98e-04	$AO_{ago/10}$	8309	3658.2	2.27	2.46e-02
$AO_{ago/10}$	8873	3459.2	2.57	1.14e-02					
SARIMA(3,1,0)(1,1,0)									
parámetro	valor	desvío	t-valor	p-valor					
ϕ_1	-0.43	0.09	-4.82	3.73e-06					
ϕ_2	-0.52	0.08	-6.63	6.97e-10					
ϕ_3	-0.31	0.09	-3.36	1.01e-03					
Φ_1	-0.19	0.09	-2.07	4.06e-02					
$IO_{jul/99}$	-11489	3764.2	-3.05	2.74e-03					
$AO_{jun/05}$	-5856	2439.3	-2.40	1.77e-02					
$LS_{set/08}$	9327	2527.4	3.69	3.21e-04					
$IO_{nov/08}$	-11322	3695.7	-3.06	2.63e-03					
$AO_{ene/09}$	-14760	2604.7	-5.67	8.07e-08					
$TC_{abr/09}$	-11783	2792.2	-4.22	4.39e-05					
$AO_{jun/09}$	8049	2580.8	3.12	2.21e-03					
$AO_{set/09}$	8854	2779.7	3.19	1.79e-03					
$AO_{ago/10}$	8332	3476.2	2.40	1.79e-02					

Cuadro 6.1: Significación de parámetros de modelos para fertilizantes.

Todos los modelos tienen todos sus parámetros significativos.

6.1.3. Validación y Diagnóstico

Se investiga el cumplimiento de los supuestos respecto a los residuos.

Con respecto a la prueba de incorrelación, los modelos que pasan las pruebas de Ljung-Box y Box-Pierce son: el $SARIMA(0, 1, 2)(0, 1, 1)$ y el $SARIMA(0, 1, 2)(1, 1, 0)$. El resto de los modelos se descartan en esta etapa.

Ninguno de los dos modelos mencionados anteriormente rechazan la hipótesis de media cero. A su vez ambos verifican los supuestos de homoscedasticidad y normalidad.

El primero de los candidatos tiene menor $AICC$ que el segundo (2769.28 y 2772.42 respectivamente). De todas formas ambos serán probados en la etapa de predicción y será elegido el que tenga mejor performance, esto es, que tenga menor error medido por el error cuadrático medio (ECM) y errores acumulados (absoluto: EAA y relativo: EAR).

6.1.4. Predicción

Se realizan predicciones dentro de la muestra con distintos cortes en la serie y se calculan errores cuadráticos medios y errores acumulados para los modelos candidatos.

El cuadro siguiente muestra los resultados de la predicción con algunos cortes para estos modelos.

horizonte	período	SARIMA(0,1,2)(0,1,1)			SARIMA(0,1,2)(1,1,0)		
		ECM	EAA	EAR(%)	ECM	EAA	EAR(%)
2 pasos	Ene/Feb-2010	5897194	-2862	-9.08	5897194	-1560	-4.95
2 pasos	Mar/Abr-2010	2073764	1224	4.13	1920488	1073	3.62
2 pasos	May/Jun-2010	25581428	-9259	-26.04	36600047	-11318	-31.83
2 pasos	Jul/Ago-2010	47805708	2281	5.76	45835566	2346	5.93
2 pasos	Set/Oct-2010	83492390	18203	25.95	76109791	17211	24.53
2 pasos	Nov/Dic-2010	14342194	-874	-2.42	12006992	-385	-1.06
3 pasos	Ene-Mar/2010	4027197	-2326	-4.92	1523924	-541	-1.15
3 pasos	Abr-Jun/2010	21207865	-11599	-23.44	29567769	-13786	-27.86
3 pasos	Jul-Set/2010	49656700	9586	13.05	43662512	8616	11.73
3 pasos	Oct-Dic/2010	22790772	10345	14.27	25377304	11765	16.23
4 pasos	Ene/Abr-2010	3360303	-3492	-5.71	1387344	-1530	2.5
4 pasos	May/Ago-2010	35872901	-8764	-11.66	40606888	-11028	-14.68
4 pasos	Set/Dic-2010	49593986	24045	22.61	45098269	23406	22.01
6 pasos	Ene-Jun/2010	12627256	-13782	-14.25	15401503	-13857	-14.32
6 pasos	Jul-Dic/2010	41943150	22554	15.46	38961765	22230	15.24
12 pasos	Ene-Dic/2010	22570603	392	0.16	22893564	229	0.09
6 v. 1 paso	Ene-Jun/2010	10590413	-	-	11721539	-	-
6 v. 1 paso	Jul-Dic/2010	41888640	-	-	39923123	-	-
12 v. 1 paso	Ene-Dic/2010	26239527	-	-	25822331	-	-

Cuadro 6.2: Errores cuadráticos medios y acumulados de las predicciones.

Los modelos presentan un comportamiento parecido en predicción, no es claro cual genera predicciones con menores errores. Para decidir entre ellos se recurre al *AIC* corregido. El *AICC* del modelo *SARIMA*(0, 1, 2)(0, 1, 1) es 2769.28 mientras que el del *SARIMA*(0, 1, 2)(1, 1, 0) vale 2772.42, por lo que el primero es el elegido.

El modelo univariado estimado para la serie de ventas en plaza de fertilizantes es:

$$\begin{aligned} \hat{y}_t = & y_{t-1} + y_{t-12} - y_{t-13} - 0,57e_{t-1} - 0,3e_{t-2} - 0,37e_{t-12} + 0,21e_{t-13} + 0,11e_{t-14} \\ & - 5662AO_{jul/99} - 6447AO_{jun/05} + 12872TC_{nov/05} + 12872TC_{nov/07} - 13385TC_{set/08} \\ & - 11826AO_{ene/09} - 11707SO_{mar/09} + 9584AO_{jun/09} + 10036AO_{set/09} + 8534AO_{ago/10} \end{aligned}$$

donde:

\hat{y}_t representa las ventas estimadas de fertilizantes para el mes t .

y_{t-i} representa las ventas reales de fertilizantes en el mes $t - i$.

$e_{t-i} = y_{t-i} - \hat{y}_{t-i}$ es el error de predicción en el mes $t - i$.

$AO_{jul/99}$ a , $AO_{jun/05}$, $AO_{ene/09}$, $AO_{jun/09}$, $AO_{set/09}$, $AO_{ago/10}$ son atípicos aditivos, es decir variables impulso que valen 1 en el mes indicado y cero en el resto.

$TC_{nov/05}$ y $TC_{set/08}$ son cambios transitorios, es decir variables definidas como:

$$TC_i(t) = \frac{1}{1-0.7B} I_i(t) = (1+0.7B+0.7^2B^2+\dots)I_i(t) = (1+0.7I_i(t)+0.7^2I_{i+1}(t)+0.7^3I_{i+2}(t)+\dots),$$

ya que $B^j I_i(t) = I_i(t-j)$ vale 1 si $i = t-j$ y cero en otro caso, lo que equivale a $I_{i+j}(t)$

$SO(t)$ es un atípico estacional, es decir una variable definida como $SO_i(t) = 1$ si $t = i, si, 2si, \dots$ donde s es el período de la estacionalidad (en este caso 12 meses). $SO_i(t) = 0,0909091$ si $t > i$ y $t \neq ksi$ con k entero mayor o igual a cero y $SO_i(t) = 0$ si $t < i$.

A continuación se exhiben resultados con mayor detalle de las predicciones con el modelo elegido.

Predicción a doce pasos

fecha	ventas	predicción	error absoluto	error relativo (%)	lim inf IC	lim sup IC
Ene/2010	20678.05	20147.19	530.86	2.57	13457.14	26837.24
Feb/2010	10851.00	14244.02	-3393.02	-31.27	6951.08	21536.96
Mar/2010	15742.19	15206.28	535.91	3.40	7846.37	22566.19
Abr/2010	13916.61	15082.64	-1166.03	-8.38	7656.36	22508.92
May/2010	16911.62	24222.48	-7310.86	-43.23	16730.41	31714.54
Jun/2010	18645.74	21624.61	-2978.87	-15.98	14067.34	29181.89
Jul/2010	12646.32	19588.85	-6942.53	-54.90	11966.93	27210.78
Ago/2010	26923.38	20321.71	6601.67	24.52	12635.68	28007.74
Set/2010	33867.59	28260.27	5607.32	16.56	20510.66	36009.88
Oct/2010	36288.43	28703.25	7585.18	20.90	20890.58	36515.92
Nov/2010	15456.63	17389.10	-1932.47	-12.50	9513.87	25264.32
Dic/2010	20727.58	17473.12	3254.46	15.70	9535.83	25410.41
Ene-Dic/2010	242263.5	391.62	0.16	-	-	-

$$ECM = 22570603$$

Cuadro 6.3: Predicción año 2010 con información hasta diciembre 2009.

Los errores de predicción son superiores al 10% en valores absolutos en 9 de los 12 meses, el rango varía entre 2,57% y -54,9%. El máximo error se produce en el mes julio de 2010.

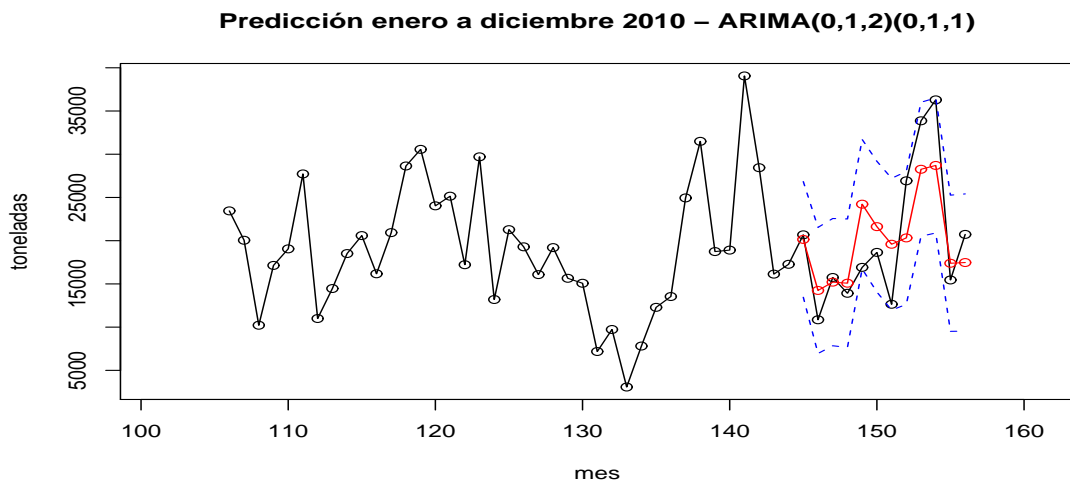


Figura 6.7: Predicción a doce pasos: año 2010 con información hasta diciembre 2010.

Predicción doce veces a un paso

fecha	ventas	predicción	error absoluto	error relativo (%)	lim inf IC	lim sup IC
Ene/2010	20678.05	20147.19	530.86	2.57	13457.14	26837.24
Feb/2010	10851.00	14429.33	-3578.33	-32.98	7764.30	21094.37
Mar/2010	15742.19	13826.60	1915.59	12.17	7154.52	20498.68
Abr/2010	13916.61	15345.82	-1429.21	-10.27	8691.87	21999.78
May/2010	16911.62	23577.98	-6666.36	-39.42	16945.08	30210.87
Jun/2010	18645.74	18094.57	551.17	2.96	11412.39	24776.74
Jul/2010	12646.32	18325.32	-5679.00	-44.91	11667.10	24983.54
Ago/2010	26923.38	16435.77	10487.61	38.95	9736.95	23134.59
Set/2010	33867.59	29590.29	4277.30	12.63	22704.71	36475.86
Oct/2010	36288.43	29151.84	7136.59	19.67	22264.29	36039.40
Nov/2010	15456.63	20098.68	-4642.05	-30.03	13141.47	27055.88
Dic/2010	20727.58	16447.86	4279.72	20.65	9471.86	23423.86

ECM = 26239527

Cuadro 6.4: Predicción doce veces a un paso: año 2010 con información hasta el paso anterior.

Incorporando información del mes anterior se produce una mejora en la mitad de los meses, mientras que en la otra mitad el error es mayor en comparación con la predicción a 12 pasos. Julio de 1020 continúa siendo el mes con el mayor error en valor absoluto.

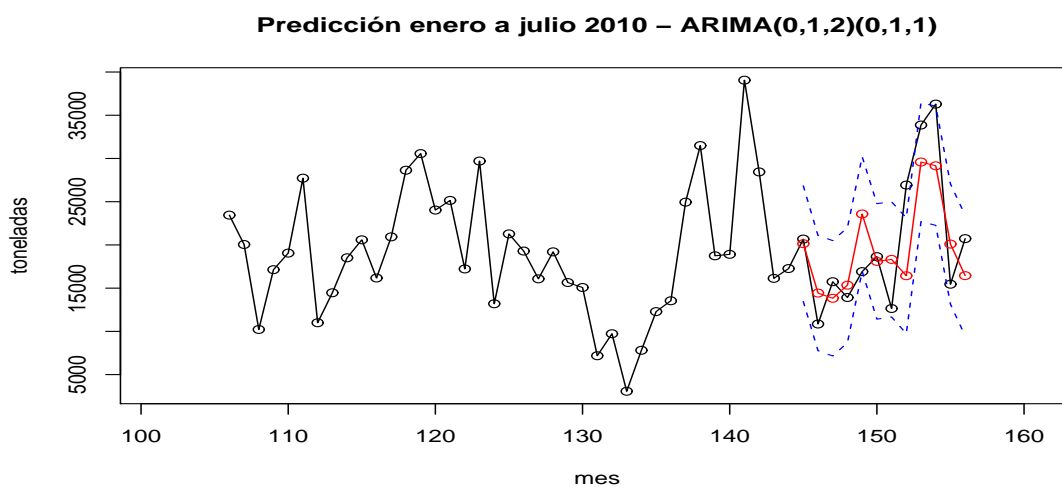


Figura 6.8: Predicción doce veces a un paso: año 2010 con información hasta el paso anterior.

Comentarios sobre el modelo

El modelo elegido capta la dinámica de la serie y verifica todos los supuestos respecto a los residuos que respaldan la teoría de series de tiempo univariadas. A pesar de lo anterior, no tiene buen poder predictivo, que se manifiesta en el alto porcentaje de los errores de predicción, que en algunos casos superan el 50 % del valor observado.

La alternativa de descomponer la serie de ventas de fertilizantes en su conjunto sería de utilidad si los modelos propuestos no hubiesen captado la regularidad de la serie (por tratarse de la suma de muchas series diferentes), pero este inconveniente no tuvo lugar. De igual forma se probó modelizar las tres principales series que conforman las ventas de fertilizantes: ventas de fertilizantes fosfatados, ventas de fertilizantes nitrogenados y ventas de fertilizantes binarios NP. Se llegó a resultados similares a los obtenidos con la serie agregada, por lo que no se consideró oportuno descomponer la serie original.

Una posible explicación para el mal comportamiento del modelo en predicción es su naturaleza univariada, es posible que las ventas de fertilizantes estén más ligadas a otras variables que a su propia historia. En este sentido es de esperar que los modelos multivariados generen predicciones con menores errores, aunque esta afirmación está sujeta a la disponibilidad de información externa y a la capacidad de identificar las variables que explican el fenómeno en estudio.

6.2. Modelo multivariado

Por modelo multivariado se entiende aquel que incluye como variables explicativas además de la serie en estudio otro tipo de variables que se considera pueden explicar dicho fenómeno. Para modelizar la serie de venta de fertilizantes en plaza podrían considerarse como variables explicativas por ejemplo el clima (lluvias), la actividad de la economía (PBI general o de algún sector), los precios de los fertilizantes y de los productos que lo utilizan, el área sembrada en el territorio nacional, variables estacionales y de efecto calendario.

6.2.1. Variables vinculadas a la serie objeto de estudio

A continuación se detalla un conjunto de variables que posiblemente afecten a las ventas de fertilizantes en plaza.

- **Precio de bienes relacionados:** Los fertilizantes tienen la característica de ser un “bien intermedio” en la medida que son un insumo en otros procesos productivos. Esto podría implicar que el nivel de venta del mismo está asociado al nivel de producción y venta de dichos productos. Los consumidores de fertilizantes pertenecen a las siguientes ramas de actividad: agricultura, ganadería, lechería y forestación. Según lo consultado con técnicos de la empresa el agro es el principal comprador dado que los cultivos requieren en mayor o menor medida el uso de fertilizantes para lograr cosechas satisfactorias y obtener mayores rendimientos del suelo. En la ganadería se fertilizan las praderas para proveer al ganadero de mayor cantidad y calidad de alimento. Algo similar ocurre en la lechería, mientras que el uso de fertilizantes en forestación se encuentra aún en expansión.

Para recopilar esta información se consideró la serie mensual del índice de precios de alimentos que elabora la FAO, la cual incluye precios internacionales de carne, leche, cereales, azúcar y aceite ([15]). Esta serie se identifica en el modelo con el nombre **FPI**.

- **Indicador del nivel de actividad económica del país:** Es importante considerar el contexto económico en el que se encuentra el país y en particular el sector agropecuario que es el principal comprador de fertilizantes. En este sentido se utilizó el índice de volumen físico del producto bruto interno agropecuario de Uruguay, dado que se trata exclusivamente de ventas en plaza. Dado que esta serie tiene frecuencia trimestral se utilizó un método estadístico de interpolación para obtener una serie de frecuencia mensual. En el modelo esta variable se denomina *PBI*.
- **Precio de fertilizantes:** Se considera esta variable para investigar la relación existente entre el precio y la cantidad vendida de fertilizantes. Se elaboró una serie de precios promedio ponderando los diferentes tipos de fertilizantes según su participación en el volumen de ventas mensuales de fertilizantes producidos por ISUSA. Se utilizó el CPI³ para transformar los precios nominales en precios reales. Esta variable se identifica con el nombre *pfert* en los cuadros que siguen.
- **Nivel de precipitaciones:** Se cree que pueden influir en las ventas de fertilizantes dado que estas últimas están vinculadas a la aplicación que no puede llevarse a cabo en días de lluvia. Para incorporar esta variables se construyó una serie de precipitaciones mensuales promedio para todo el país, en base a información diaria de las estaciones agrometeorológicas del Instituto Nacional de Investigaciones Agropecuarias (INIA). A esta variable se la denomina *lluvia*.

En la siguiente figura se puede apreciar la representación gráfica de las variables consideradas:

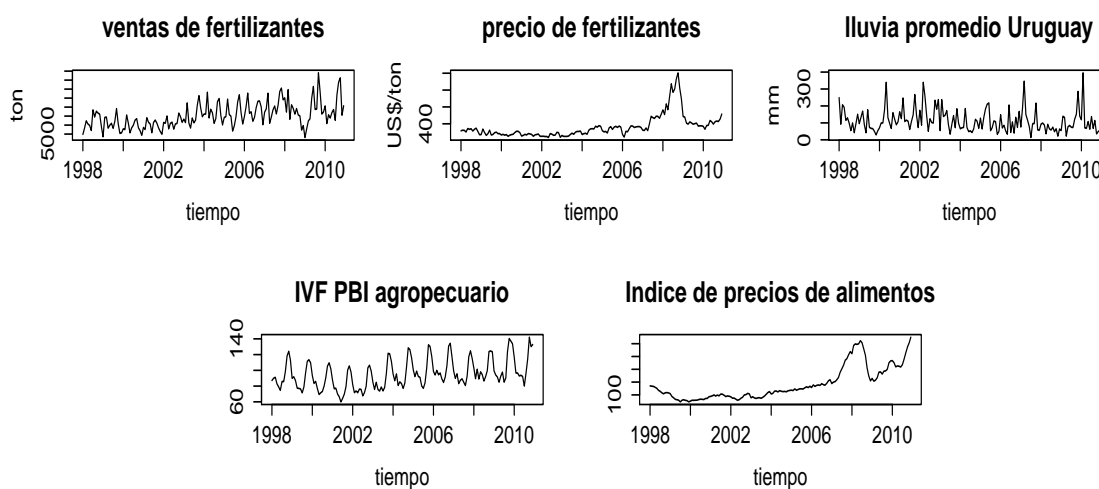


Figura 6.9: series a considerar en un modelo multivariado para ventas de fertilizantes

Mensualización del Índice de Volumen Físico del PBI del agro

La serie de Índice de Volumen Físico del PBI del agro uruguayo tiene frecuencia trimestral, por lo que es necesario mensualizarla para poder incorporarla a un modelo multivariado junto a otras series de frecuencia mensual.

La mensualización de esta serie se llevó a cabo a través del programa Ecotrim ([18]). Se utilizó el método de Chow según se recomienda en Eustat ([8]) utilizando como variable explicativa la serie de ventas en plaza de fertilizantes (serie objeto de estudio). Esta decisión se fundamenta en

³Índice de precios al consumo de Estados Unidos, del inglés Consumer Price Index.

la percepción de que estas dos series se encuentran estrechamente vinculadas y se ven afectadas por los mismos fenómenos.

A continuación se detalla el método de Chow, que fue extraído de las notas de Eustat y ajustado a la mensualización de una serie trimestral.

El problema de la desagregación mensual de $Y_T (T = 1, \dots, N)$, serie trimestral observada, basada en la información del indicador $x_{t,T} (t = 1, 2, 3 \text{ y } T = 1, \dots, N)$, consiste en estimar una serie $y_{t,T} (t = 1, 2, 3 \text{ y } T = 1, \dots, N)$ que minimice la distorsión con $x_{t,T}$ en algún sentido prefijado y satisfaga la restricción temporal:

$$\sum_{i=1}^3 y_{i,r} = Y_r$$

El procedimiento general de Chow y Lin parte de la relación:

$$Y = B'y = B'x\beta + B'\mu$$

donde Y es el vector de datos trimestrales conocidos, y es el vector de datos mensuales desconocidos, x es la matriz de datos mensuales integrada por las variables indicador, β es el vector de parámetros de regresión que relacionan ambas variables, μ es el vector de perturbaciones trimestrales y B es la matriz de agregación temporal (matriz de paso de mensual a trimestral), que para variables flujo tiene la forma:

$$B' = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 \end{pmatrix}$$

Bajo esta especificación del modelo, Chow y Lin obtienen la estimación mínimo cuadrática óptima de los parámetros β y, por consiguiente, la estimación del vector y de datos mensuales:

$$\hat{\beta} = [x'B(B'VB)^{-1}B'x]^{-1}x'B(B'VB)^{-1}Y$$

$$\hat{y} = x\hat{\beta} + VB(B'VB)^{-1}(Y - B'x\hat{\beta})$$

donde V representa la matriz de covarianzas de la perturbación trimestral que es desconocida y cuya estructura dependerá del supuesto realizado sobre el comportamiento de las perturbaciones. La estructura de las perturbaciones propuesta por Chow y Lin consiste en suponer que los errores mensuales se distribuyen según un proceso autorregresivo de orden uno estacionario:

$$\mu_t = \phi\mu_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

$$|\phi| < 1$$

6.2.2. Modelización

Pruebas de raíz unitaria de Dickey Fuller

El primer paso para la modelización multivariada consiste en determinar si las series consideradas individualmente son estacionarias o integradas para elegir el tipo de modelo a utilizar. Si son todas estacionarias se prueba un VAR en niveles, mientras que si son todas integradas del mismo orden se investiga la presencia de un vínculo común estacionario entre ellas; se testea la cointegración. Si las series están cointegradas se implementa un VECM, en caso contrario se prueba un VAR en diferencias.

Se realizaron pruebas de Dickey Fuller aumentado (ADF) a las series individuales para determinar el orden de integración de cada una. Se obtuvo como resultado que la serie de nivel de lluvia es estacionaria ($I(0)$) mientras que las restantes son integradas de orden 1 ($I(1)$)⁴.

Prueba de causalidad de Granger aplicada a las series

Se realizó la prueba de causalidad de Granger a cada par de series, dos a dos, para determinar el sentido de la relación entre ellas. Este test para las series x y y considerando p rezagos realiza la prueba F entre dos modelos, uno donde se estima a x en función de p rezagos tanto de x como de y , a este modelo se lo llama modelo completo. Luego se estima otro modelo donde x es la variable dependiente y los p rezagos de x las variables independientes. A este modelo se lo llama modelo reducido.

La hipótesis nula es que todos los parámetros asociados a los rezagos de y en el modelo completo valen cero. Si esta hipótesis no se rechaza se concluye que no existe causalidad de y a x .

A continuación se muestra el p -valor (en %) de la prueba de causalidad aplicada a cada par de variables. Las variables $I(1)$ están en diferencias y la variable estacionaria nivel de lluvia está en niveles.

orden	Δ fert Δ PBI	Δ fert Δ FPI	Δ fert Δ pfert	Δ fert lluvia	Δ PBI Δ fert	Δ PBI Δ FPI	Δ PBI Δ pfert	Δ PBI lluvia	Δ FPI Δ fert	Δ FPI Δ PBI
1	28.93	38.29	0.90	39.91	0.00	34.45	8.36	32.10	7.17	33.04
2	0.00	83.43	1.10	19.65	0.00	72.99	22.56	47.35	0.03	1.96
3	0.00	63.20	1.62	35.85	0.00	82.37	2.67	55.73	0.05	0.46
4	0.00	55.13	3.86	28.94	0.00	83.14	0.26	54.96	0.07	0.87
5	0.00	70.02	5.54	37.55	0.00	80.61	0.39	6.46	0.02	1.39
6	0.00	72.11	4.01	40.32	0.00	90.50	1.16	9.96	0.01	1.66
7	0.00	77.01	1.47	48.17	0.00	92.13	0.61	6.42	0.00	2.83
8	0.00	55.08	1.76	16.96	0.00	93.38	1.00	5.83	0.00	1.74
9	0.00	53.74	0.98	5.46	0.00	96.46	0.12	8.10	0.00	1.28
10	0.00	31.52	3.30	9.14	0.00	78.07	0.03	4.30	0.00	1.16
11	0.00	35.94	4.36	19.61	0.00	86.83	0.18	9.88	0.00	5.64
12	0.00	26.60	0.39	20.02	0.00	65.41	0.53	10.59	0.00	5.48

⁴Por detalles del procedimiento ver anexo A.

orden	ΔFPI $\Delta pfert$	ΔFPI lluvia	$\Delta pfert$ $\Delta fert$	$\Delta pfert$ ΔPBI	$\Delta pfert$ ΔFPI	$\Delta pfert$ lluvia	lluvia $\Delta fert$	lluvia ΔPBI	lluvia ΔFPI	lluvia $\Delta pfert$
1	0.05	61.24	7.52	70.32	79.35	65.01	29.67	32.23	38.45	98.10
2	0.26	54.58	7.80	22.84	5.70	32.16	38.14	12.34	50.28	43.55
3	0.00	25.68	14.10	47.49	5.51	27.93	52.02	10.72	70.17	21.18
4	0.00	37.20	32.27	2.14	1.93	44.21	70.94	12.36	79.45	30.54
5	0.00	54.64	57.75	1.03	4.76	43.39	74.69	13.88	76.19	41.22
6	0.00	50.05	16.52	2.42	8.56	57.84	78.12	7.00	49.16	54.99
7	0.00	40.74	17.92	4.82	10.87	67.93	30.62	3.98	51.79	58.89
8	0.00	28.86	7.98	10.72	10.23	76.26	38.92	26.40	26.61	61.98
9	0.00	34.91	13.91	57.74	4.65	80.04	43.43	55.92	34.35	65.63
10	0.00	41.92	9.27	26.78	0.96	34.88	64.73	97.03	34.82	74.05
11	0.00	26.96	0.22	4.98	1.25	33.31	60.21	96.91	44.48	81.92
12	0.00	15.97	0.21	1.41	2.35	39.26	66.68	79.21	34.38	18.48

Cuadro 6.5: *P-valor* de la prueba de causalidad de Granger aplicada a cada par de variables consideradas.

Cada valor menor a 5%, nivel de significación considerado, implica que existe causalidad de Granger de la primera variable hacia la segunda, para la cantidad de rezagos que indica la columna orden. Por ejemplo el *p-valor* de la prueba “ $\Delta fert$ lluvia” con un rezago es 39.91% lo cual implica que la venta de fertilizantes no causa al nivel de lluvias.

Las pruebas que más interesan en este caso son las asociadas la serie de fertilizantes, serie objetivo a modelizar. En este sentido se obtiene que existe causalidad de Granger de la serie *PBI* hacia la serie de venta de fertilizantes para cualquier número de rezagos y de la serie *FPI* para más de un rezago, pero la serie de fertilizantes causa únicamente al *PBI* (salvo para el primer rezago) y no al *FPI*, para cualquier cantidad de rezagos.

Estos resultados son consistentes con el conocimiento económico de las variables y sus interrelaciones. Por un lado las ventas de fertilizantes se encuentran estrechamente ligadas al *PBI* del agro del país dado que constituye un insumo fundamental para la producción del sector y la relación entre ambas se da en ambos sentidos, cuanto más produce el agro más fertilizantes se demanda y si se produce más fertilizante aumenta la producción del país en el sector.

Por otro lado la demanda de fertilizantes está asociada al precio internacional de los commodities, los productores nacionales se basan en gran medida en la especulación referente a los precios de distintos productos y en función de ello sumado a otros factores deciden el nivel de inversión que están dispuestos a asumir. En este caso se trata de una relación en un sentido, los precios internacionales de productos básicos afectan a la demanda de fertilizantes, pero la demanda de fertilizantes de ISUSA no afecta a los precios internacionales. También existe causalidad de Granger de la serie de índice de precios de alimentos (*FPI*) hacia el *PBI* del agro, cómo era de esperarse por el análisis realizado anteriormente.

En función de los resultados de la prueba de causalidad y de las conclusiones obtenidas se decide modelizar la demanda de fertilizantes en un sistema de dos variables endógenas: venta de fertilizantes y *PBI* del agro, y una variable exógena que afecta a las dos endógenas: índice de precios de alimentos (*FPI*).

Las otras variables consideradas, lluvia y precio de fertilizantes, no causan a las ventas de fertilizantes por lo que se descartan del análisis.

6.2.2.1. Prueba de cointegración y estimación

En la sección anterior se determinó a partir de un conjunto de variables vinculadas a la serie objeto de estudio cuales realmente estaban relacionadas y en que sentido se daba esa relación. Una vez especificado este punto es necesario determinar el modelo adecuado a estimar para el sistema.

Dado que las tres series involucradas en el sistema son $I(1)$, se aplica en esta sección la prueba de cointegración para determinar si existe una combinación lineal entre las series endógenas que sea estacionaria.

Se utiliza la prueba de cointegración de Johansen⁵ y el estadístico de la traza. Se consideran las series de fertilizantes y PBI agropecuario en logaritmos. Se incluyen variables determinísticas e indicatrices de los meses. Se aplica la prueba a todas las configuraciones resultantes de combinar los siguientes elementos:

- Entre entre 2 y 13 rezagos.
- Incluyendo constante no restringida, constante restringida o sin constante⁶.
- Distintas formas de incluir a la variable exógena *FPI* en diferencias: variando cantidad de rezagos entre 0 y 6.

Después de varias pruebas se decidió incluir a la variable exógena *FPI* en diferencias con uno o dos rezagos, según el modelo.

Conjuntamente con las pruebas de cointegración se estimaron los modelos *VAR* o *VECM* según corresponda y se analizó la significación de parámetros. Para cada modelo se quitaron las variables determinísticas que representan intervenciones no significativas.

En el cuadro 6.6 se presentan las 36 variantes de modelos considerados. Para cada variante se realizó la prueba de cointegración de Johansen, utilizando el estadístico de la traza. Para cada modelo se muestra su especificación: orden, que variables determinísticas incluye, cuantos rezagos de la variable *FPI* en diferencias (*DFPI*) se incluyen y cuantas variables intervención (de tipo *AO*) se consideraron. Junto a la especificación de cada modelo se muestra el resultado de la prueba de cointegración y si el modelo es válido o no según la significación de parámetros.

Los resultados de la prueba de cointegración pueden ser “NN”, “NR” y “RR”. La primera letra hace referencia a la primera hipótesis de la prueba, $r \leq 1$, mientras que la segunda letra refiere a la segunda hipótesis, $r = 0$, donde r es el rango de cointegración. “N” significa no rechazo y “R” rechazo, entonces, “NN” equivale a $r = 0$, no hay cointegración, “NR” equivale a $r = 1$, hay cointegración y el rango es uno. “RR” significa que r no es cero ni uno, pero tampoco puede ser mayor a uno, por lo que se descartan las variantes con este resultado.

Se incluye una columna “válido” que refiere a la validez en lo que respecta a la significación de parámetros. La idea es que si para un modelo no son significativos los parámetros que lo definen, por ejemplo la constante en el caso de un modelo con constante no restringida, ese modelo no es válido. También puede perder validez si los parámetros asociados al orden no son significativos, por ejemplo si los rezagos de orden 10 de las variables endógenas no son significativos y el orden del modelo es 10, esa variante pierde validez desde el punto de vista de la significación de parámetros.

⁵Por detalle de la prueba ver anexo C.

⁶Dado que las series en diferencias no presentan tendencia determinística se excluyen las variantes de tendencia en la especificación.

modelo	orden	caso	rezagos DFPI	cantidad interv.	resultados coint.	válido
1	2	const. no restr.	1	16	NR	no
2	2	const. restr.	2	10	NR	no
3	2	sin const. ni tend.	2	10	NR	no
4	3	const no restr.	2	11	NR	no
5	3	const. restr.	2	11	NR	no
6	3	sin const. ni tend.	2	9	NR	no
7	4	const. no restr.	2	11	NR	no
8	4	const. restr.	1	9	NR	no
9	4	sin const. ni tend.	1	12	NR	si
10	5	const. no restr.	2	9	NR	no
11	5	const. restr.	2	9	NR	no
12	5	sin const. ni tend.	2	9	NR	no
13	6	const. no restr.	2	8	NR	no
14	6	const. restr.	2	8	NR	no
15	6	sin const. ni tend.	2	8	NR	no
16	7	const. no restr.	2	7	NR	no
17	7	const. restr.	2	7	NR	no
18	7	sin const. ni tend.	2	7	NR	no
19	8	const. no restr.	2	7	NR	si
20	8	const. restr.	2	6	NR	si
21	8	sin const. ni tend.	2	6	NR	si
22	9	const. no restr.	2	4	NR	si
23	9	const. restr.	1	6	NR	si
24	9	sin const. ni tend.	1	6	NR	no
25	10	const. no restr.	2	6	NR	no
26	10	const. restr.	2	7	NR	no
27	10	sin const. ni tend.	2	7	NR	no
28	11	const. no restr.	2	6	NR	no
29	11	const. restr.	2	6	NR	no
30	11	sin const. ni tend.	2	6	NR	no
31	12	const. no restr.	1	5	NR	si
32	12	const. restr.	1	6	NR	si
33	12	sin const. ni tend.	1	6	NR	si
34	13	const. no restr.	1	4	NR	no
35	13	const. restr.	1	4	NR	no
36	13	sin const. ni tend.	1	4	NR	no

Cuadro 6.6: Especificación de los modelos estimados.

En todas las especificaciones se obtiene como resultado de la prueba de cointegración un rango igual a uno (resultado de la prueba de cointegración igual a “NR”).

Se preseleccionan 9 modelos válidos, es decir, donde los parámetros que lo definen son significativos.

6.2.2.2. Validación y Diagnóstico

En esta sección se aplican pruebas diagnósticas a los 9 modelos preseleccionados en la sección anterior. Se estudia el comportamiento de los residuos individuales (univariado) y en conjunto (multivariado). Un modelo será válido si los residuos cumplen con los supuestos de normalidad, incorrelación y homoscedasticidad tanto individualmente como en conjunto.

Para testear la normalidad se utiliza la prueba de Jarque-Bera, para la incorrelación se utilizan los test de Portmanteau (*PT*) y Breusch-Godfrey (*BG*) de correlación serial y para la homoscedasticidad se estima un modelo *ARCH*⁷.

El cuadro que se presenta a continuación muestra el *p-valor* de cada una de las pruebas aplicadas a los modelos candidatos.

mod.	NORMALIDAD					INCORRELACIÓN				HOMOSCED.		
	fert. univ.	PBI. univ.	mult.	sim	kur	PT. asint.	PT. ajust.	BG	ES	fert. univ.	PBI. univ.	mult.
9	58.56	87.89	57.29	46.61	50.07	0.04	0.01	0.11	2.28	79.34	19.79	4.74
19	50.46	77.56	63.10	31.02	88.89	0.14	0.09	0.50	13.60	56.72	0.80	59.28
20	56.58	45.32	74.97	52.91	72.22	0.15	0.09	0.51	12.38	55.18	35.60	90.56
21	54.31	48.06	73.13	52.43	69.32	0.21	0.13	0.86	14.98	67.62	33.32	90.39
22	0.00	7.47	0.00	0.15	0.00	0.75	0.44	36.03	87.21	92.35	44.04	39.46
23	12.83	14.60	27.82	18.42	42.61	0.08	0.05	0.30	10.72	47.86	40.71	93.95
31	85.36	65.14	25.62	75.49	9.28	0.00	0.00	0.68	21.79	0.19	0.15	0.00
32	17.33	3.99	44.23	18.12	85.04	0.00	0.00	0.64	27.41	33.20	70.38	43.26
33	16.79	4.59	43.20	17.47	85.07	0.00	0.00	0.76	27.56	33.66	70.81	44.90

La prueba de normalidad utilizada tiene como hipótesis nula que los residuos siguen una distribución normal, por lo que sólo se descarta esta hipótesis para el modelo 22.

En la prueba de incorrelación la hipótesis nula es que los residuos están incorrelacionados, rechazo esta hipótesis para todos los modelos según los estadísticos *PT* asintótico y ajustado. No rechazo únicamente para el modelo 22 utilizando el estadístico *BG* mientras que solamente rechazo la hipótesis nula para el modelo 9 en el caso del estadístico *ES*.

En lo que respecta a la prueba de homoscedasticidad, la hipótesis nula es que los residuos son homoscedásticos, se obtiene que para el modelo 9 se rechaza la hipótesis nula en la versión multivariada, en el modelo 19 se rechaza la hipótesis nula para los residuos de *PBI* y para el modelo 31 se rechaza la hipótesis nula en las tres variantes. Para el resto de los modelos no se rechaza la hipótesis nula.

En función de los resultados anteriores se descartan los modelos 9, 19, 22 y 31.

Los 5 modelos preseleccionados serán evaluados en predicción en la siguiente sección.

6.2.2.3. Predicción

En esta sección se realizan predicciones a distintos horizontes de los modelos candidatos. En principio se utilizan los valores reales de la variable exógena *FPI* de forma de evaluar la performance del modelo en sí. Luego se evalúa el modelo con valores predichos de la serie exógena, para la cual se estimó un *ARMA*(1,1)⁸.

En ambos casos se realizaron predicciones a 3, 6, 9 y 12 pasos.

⁷Modelo autorregresivo de heteroscedasticidad condicional.

⁸Se estima un modelo para la serie estacionaria *FPI* en diferencias.

En los cuadros 6.7 y 6.8 se muestra el *ECM* de la predicción para cada modelo y distintos horizontes.

mod.	pasos			
	3	6	9	12
20	38,832,331.86	36,878,716.3	23,402,111.92	65,011,543.62
21	37,518,614.92	36,828,733.16	23,847,102.39	49,041,624
23	36,393,415.17	40,562,828.55	28,196,989.4	97,883,808.33
32	55,290,813.17	43,678,523.52	33,864,528.45	51,947,894.3
33	54,282,849.14	42,806,350.73	32,426,937.72	43,779,304.73

Cuadro 6.7: *ECM* de predicción en serie logfert para distintos horizontes usando datos reales del *FPI*, según modelo.

mod.	pasos			
	3	6	9	12
20	49.17	39.15	24.58	40.97
21	50.99	39.27	24.69	38.48
23	87.32	51.35	36.3	90.52
32	88.77	52.19	38.85	32.36
33	89.54	52.59	39.34	31.72

Cuadro 6.8: *ECM* de predicción en serie logpbi para distintos horizontes usando datos reales del *FPI*, según modelo.

De la comparación de los 5 candidatos restantes se obtiene que los mejores tanto para predecir la serie logfert como la serie logpbi son el 20 y el 21.

A continuación se muestran los resultados de las predicciones a 3, 6, 9 y 12 pasos para los candidatos finales y los errores asociados.

pasos	modelo fert		modelo PBI	
	20	21	20	21
3	12.55	13.80	2.26	2.16
	-62.88	-59.43	-8.88	-9.13
	5.52	8.49	1.42	1.17
6	-21.55	-19.17	-5.87	-5.87
	15.65	18.11	3.60	3.52
	9.20	12.42	3.03	2.87
	11.11	14.42	3.85	3.64
	-83.77	-76.68	-9.32	-9.51
	7.54	11.56	2.58	2.52
9	-16.17	-13.61	-0.60	-0.56
	-16.99	-12.69	-3.03	-3.08
	6.99	10.71	-0.32	-0.50
	-14.83	-9.47	-6.60	-6.75
	16.93	21.17	2.54	2.50
	5.41	10.19	1.37	1.38
	10.67	14.75	2.76	2.77
	-79.84	-72.48	-9.57	-9.54
3.87	7.52	1.82	1.85	
12	-37.42	-32.77	2.70	2.73
	-84.10	-73.84	-0.36	-0.42
	-82.22	-72.17	-4.23	-4.43
	-65.76	-53.19	-4.17	-4.14
	-32.45	-20.76	-6.26	-6.07
	-6.94	1.91	-5.30	-5.12
	-38.51	-26.23	-13.75	-13.36
	18.01	25.21	-1.09	-0.68
	2.26	9.99	-1.52	-1.15
	-4.29	3.47	-0.57	-0.25
	-110.95	-93.96	-12.50	-12.07
	-14.97	-4.96	-0.61	-0.21

Cuadro 6.9: Errores relativos de predicción para los modelos candidatos usando datos reales de variable exógena.

Al comparar los resultados obtenidos en predicción en los tres modelos candidatos se concluye que el que presenta menor *ECM* y en general menor error relativo en las predicciones a distintos horizontes es el modelo 21.

Para este modelo se prueba incorporar las intervenciones del modelo univariado que no están presentes en el multivariado, de forma de evaluar el impacto sobre las predicciones. Los resultados se muestran en el cuadro 6.10.

pasos	modelo fert		modelo PBI	
	21 antes	21 después	21 antes	21 después
3	13.8	22.70	2.16	4.19
	-59.43	-32.57	-9.13	-4.89
	8.49	22.58	1.17	5.14
6	-19.17	-19.93	-5.87	-6.15
	18.11	18.99	3.52	3.44
	12.42	20.78	2.87	5.17
	14.42	26.07	3.64	7.31
	-76.68	-45.23	-9.51	-3.81
	11.56	26.86	2.52	7.76
9	-13.61	-25.96	-0.56	-2.98
	-12.69	-18.08	-3.08	-5.28
	10.71	17.61	-0.50	-0.56
	9.47	-6.38	-6.75	-7.46
	21.17	17.99	2.5	1.11
	10.19	14.71	1.38	2.37
	14.75	27.22	2.77	6.29
	-72.48	-47.77	-9.54	-4.72
7.52	20.96	1.85	6.25	
12	-32.77	-20.98	2.73	3.53
	-73.84	-54.51	-0.42	1.07
	-72.17	-73.88	-4.43	-6.40
	-53.19	-60.65	-4.17	-7.19
	-20.76	-14.58	-6.07	-7.31
	1.91	11.22	-5.12	-5.17
	-26.23	-26.22	-13.36	-14.51
	25.21	23.64	-0.68	-1.60
	9.90	14.41	-1.15	-0.36
	3.47	7.33	-0.25	1.23
	-93.96	-72.08	-12.07	-8.66
	-4.96	12.60	-0.21	3.77

Cuadro 6.10: Comparación de errores relativos de predicción para el modelo 21 actual (antes) y con las intervenciones del modelo univariado (después).

De la comparación de los resultados del cuadro anterior se concluye que incorporar las intervenciones del modelo univariado en el multivariado no mejora las predicciones, por lo que se mantiene el modelo candidato estimado anteriormente, sin incorporar otras intervenciones.

El modelo $ARMA(1, 1)$ estimado para la serie $DFPI$ no tiene buena performance en predicción, por lo que no se analiza su utilización en el modelo de fertilizantes.

El modelo multivariado estimado para las ventas de fertilizantes es:

$$\begin{aligned}
\hat{y}_t &= y_{t-1} - 0,49\Delta y_{t-1} - 0,33\Delta y_{t-2} - 0,31\Delta y_{t-3} - 0,47\Delta y_{t-4} - 0,23\Delta y_{t-5} - 0,25\Delta y_{t-6} \\
&\quad - 0,53\Delta y_{t-7} - 0,57\Delta x_{t-1} - 1,70\Delta x_{t-2} - 1,51\Delta x_{t-3} + 1,36\Delta x_{t-4} - 0,32\Delta x_{t-5} \\
&\quad + 0,38\Delta x_{t-6} + 3,32\Delta x_{t-7} - 0,41y_{t-8} + 0,03DFPI_{t-1} - 0,007DFPI_{t-2} - 1,28AO_{ene/99} \\
&\quad - 0,75AO_{may/00} - 0,99AO_{jun/01} + 0,76AO_{ene/02} - 0,83AO_{jun/05} - 1,09AO_{ene/09} \\
&\quad + 0,11I_{ene} - 0,37I_{feb} - 0,39I_{mar} - 1,31I_{abr} - 0,79I_{may} - 0,45I_{jun} - 0,55I_{jul} \\
&\quad + 0,30I_{ago} + 0,20I_{set} + 0,07I_{oct} + 0,30I_{nov} \\
\hat{x}_t &= x_{t-1} - 0,05\Delta y_{t-1} - 0,01\Delta y_{t-2} + 0,03\Delta y_{t-3} - 0,001\Delta y_{t-4} + 0,04\Delta y_{t-5} + 0,04\Delta y_{t-6} \\
&\quad - 0,01\Delta y_{t-7} - 0,12\Delta x_{t-1} - 0,36\Delta x_{t-2} - 0,49\Delta x_{t-3} + 0,06\Delta x_{t-4} - 0,19\Delta x_{t-5} \\
&\quad - 0,13\Delta x_{t-6} + 0,33\Delta x_{t-7} + 0,0003x_{t-8} + 0,005DFPI_{t-1} - 0,003DFPI_{t-2} - 0,07AO_{ene/99} \\
&\quad - 0,10AO_{may/00} - 0,17AO_{jun/01} + 0,05AO_{ene/02} - 0,11AO_{jun/05} - 0,01AO_{ene/09} \\
&\quad - 0,12I_{ene} - 0,13I_{feb} - 0,07I_{mar} - 0,23I_{abr} - 0,14I_{may} - 0,11I_{jun} - 0,17I_{jul} \\
&\quad + 0,05I_{ago} + 0,13I_{set} - 0,19I_{oct} + 0,11I_{nov}
\end{aligned}$$

donde:

\hat{y}_t y \hat{x}_t son los valores estimados para las series de fertilizantes y PBI en el mes t , respectivamente.

y_i y x_i son los valores reales de las series en el mes i .

$DFPI_i$ es el valor real de la serie FPI en diferencias en el mes i .

$AO_{ene/99}$ a $AO_{ene/09}$ son intervenciones de tipo AO en el mes indicado.

I_{ene}, \dots, I_{nov} son variables indicatrices de los meses.

Comentarios sobre el modelo

El modelo elegido para estimar de forma multivariada las ventas de fertilizantes es un *VECM*. El sistema está integrado por dos variables endógenas (fertilizantes y PBI) y una variable exógena (FPI) que se incluye tomando los primeros dos rezagos de la serie en diferencias. El modelo incluye ocho rezagos de las variables endógenas. A su vez se incluyen seis intervenciones de tipo AO e indicatrices de los meses.

Este modelo si bien es el mejor de todos los modelos multivariados estimados, no presenta buen comportamiento en predicción como era de esperarse por la inclusión de variables relacionadas al fenómeno de estudio.

Entre los motivos que pueden explicar este hecho se encuentran la carencia de otras variables que ayuden a explicar el fenómeno en estudio y la complejidad al identificar el PGD⁹ por tratarse de una serie que es la suma de muchas series seguramente con distintos PGD.

⁹Proceso generador de los datos.

Capítulo 7

Conclusiones

El objetivo central de esta investigación se basa en encontrar modelos de series temporales para las ventas en plaza de ácido sulfúrico, sulfato de aluminio y fertilizantes, producidos por ISUSA. Para esto se decidió en un principio ajustar modelos univariados a estas series siguiendo la metodología propuesta por Box y Jenkins.

Para la serie de ácido sulfúrico se ajustó un modelo de medias móviles de segundo orden con estacionalidad autorregresiva, $SARIMA(0, 0, 2)(1, 0, 0)$. Para la serie de sulfato de aluminio se ajustó un modelo integrado de medias móviles de primer orden, $ARIMA(0, 1, 1)$. Para la serie de fertilizantes se ajustó un modelo integrado de medias móviles de segundo orden con estacionalidad de media móvil, $SARIMA(0, 1, 2)(0, 1, 1)$.

Los modelos ajustados para las series de ventas de ácido sulfúrico y sulfato de aluminio tienen en general buena performance según las pruebas efectuadas de evaluación dentro de la muestra. El modelo ajustado para fertilizantes si bien es válido desde el punto estadístico, es deficiente al generar predicciones.

Con el objetivo de obtener mejores predicciones de esta serie se implementó un modelo multivariado para los fertilizantes, generalizando la idea de los modelos univariados a un conjunto de variables.

Para su construcción se consideraron las siguientes variables: índice de precios de alimentos, PBI agropecuario, precios de fertilizantes y nivel de precipitaciones. Luego de analizar las interrelaciones entre estas variables se decide incorporar como endógenas en el sistema a las variables venta de fertilizantes y PBI agropecuario y como exógena al índice de precios de alimentos.

El modelo resultante es un $VECM$ de orden ocho sin constante ni tendencia, incluyendo las indicatrices de los meses y variables de intervención determinísticas. Este modelo tampoco tiene buena performance en predicción y no se logró mejorarlo a pesar de considerar distintas alternativas.

Para encontrar un modelo con mejor performance para la serie de fertilizantes con respecto a los presentados en este trabajo se puede seguir alguna de las siguientes recomendaciones:

- Incluir otras variables al sistema, por ejemplo el área sembrada en el territorio nacional. Esta variable es determinante según la opinión de los técnicos, pero no se encontraron series de datos suficientes para su implementación.
- Descomponer la serie agregada en subseries, por ejemplo fertilizantes fosfatados, nitrogenados y binarios. Otra alternativa puede ser clasificarlos según cómo se comercializan, por ejemplo sólidos y líquidos o a granel y envasados.
- Utilizar otras técnicas como por ejemplo simulación de procesos estocásticos.

Bibliografía

- [1] Box, G. and Jenkins, G. (1970) Time series analysis: Forecasting and control, San Francisco: Holden-Day.
- [2] Brockwell, P., Davis, R. (2002) Introduction to Time Series and Forecasting, Springer.
- [3] Chen, C.,Liu, L. (1993) Joint Estimation of Model Parameters and Outlier Effect in Time Series, Journal of the American Statistical Association, Vol 88, No 421.
- [4] Cryer, J., Chan, K. (2008) Time Series Analysis With Applications in R, Springer.
- [5] Dickey, D. and Fuller, W. (1979) Distribution of the Estimators for Autorregresive Time Series With a Unit Root, Journal of the American Statistical Association, Vol 74, Issue 366.
- [6] Dickey, D. and Fuller, W. (1981) Likelihood Ratio Statistics for Autorregresive Time Series with a Unit Root, Econometrica, Vol 49, No 4.
- [7] Enders, W. (1995), Applied Econometric Time Series, John Wiley & Sons, inc.
- [8] Eustat. Nota Metodológica de las Cuentas Económicas Trimestrales de la C.A. de Euskadi.
- [9] Hamilton, J. (1994) Time Series Analysis, Princeton University Press.
- [10] Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?, Journal of Econometrics, No 54.
- [11] Peña, D. (2005), Análisis de Series Temporales, Alianza Editorial.
- [12] Pfaff, B. (2008) Analysis of Integrated and Cointegrated Time series with R, Second Edition. Springer, New York. ISBN 0-387-27960-1, URL <http://www.pfaffikus.de>.
- [13] Juselius, K. (2006) The Cointegrated VAR model methodology and applications, Oxford University Press.
- [14] Índice de volumen físico del producto bruto interno: Banco Central del Uruguay (BCU) - www.bcu.gub.uy.
- [15] Índice de precios de alimentos: Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) - www.fao.org.
- [16] Precipitaciones: Instituto Nacional de Investigación Agropecuaria (INIA) - www.inia.org.uy.

- [17] R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [18] Ecotrim (2002). European Commission Statistical Office Directorate B. Economic Statistics and Economic and Monetary Convergence. Eurostat. URL <http://circa.europa.eu/Public/irc/dsis/ecotrim/library>.
- [19] Adrian Trapletti and Kurt Hornik (2009) tseries: Time Series Analysis and Computational Finance. R package version 0.10-22, URL <http://CRAN.R-project.org/package=tseries>.
- [20] Kung-Sik Chan (2010) TSA: Time Series Analysis. R package version 0.98, URL <http://CRAN.R-project.org/package=TSA>.
- [21] Pfaff, B. (2010) urca: Unit root and cointegration tests for time series data. R package version 1.2-4, URL CRAN.R-project.org/package=urca.
- [22] Rob J Hyndman (2010) forecast: Forecasting functions for time series. R package version 2.08, URL <http://CRAN.R-project.org/package=forecast>.

Apéndice A

Anexo de figuras

A.1. Ácido sulfúrico

Gráficos de los residuos

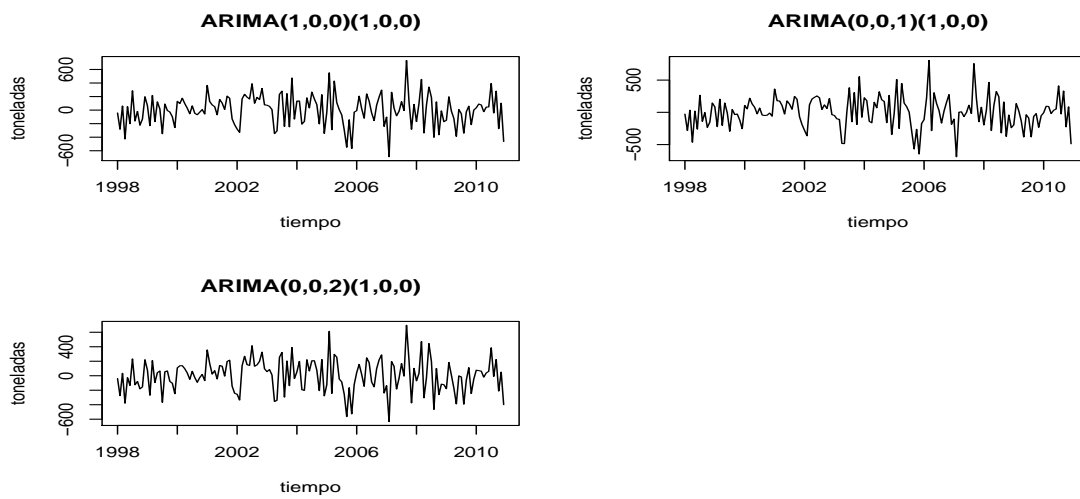


Figura A.1: Residuos de los modelos con detección automática de atípicos, serie de ventas de ácido sulfúrico.

Incorrelación

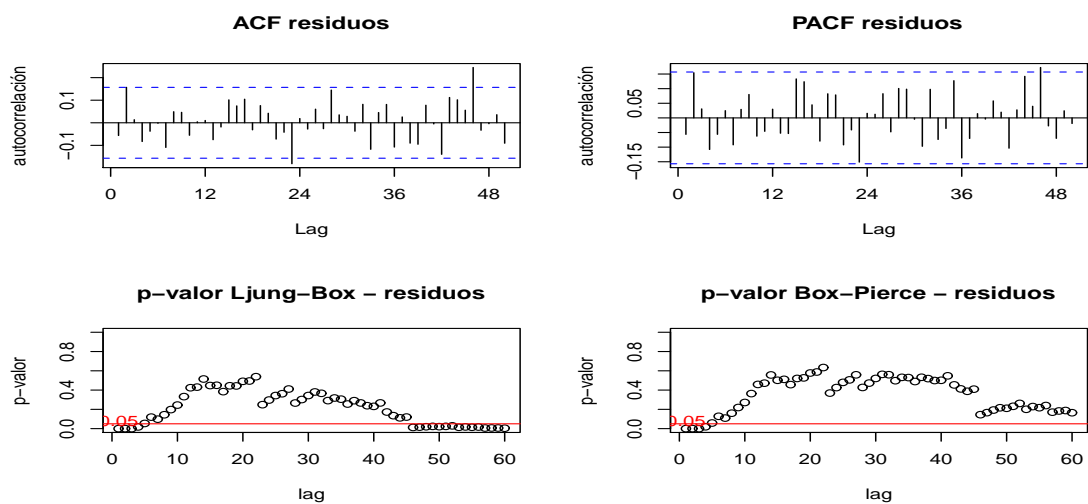


Figura A.2: Prueba de incorrelación de residuos para el modelo $SARIMA(1,0,0)(1,0,0)$.

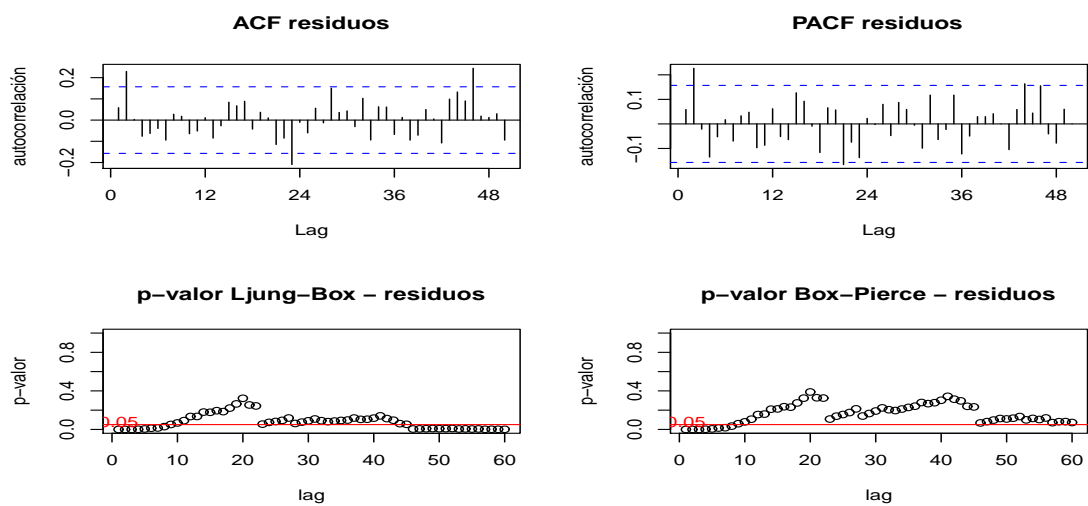


Figura A.3: Prueba de incorrelación de residuos para el modelo $SARIMA(0,0,1)(1,0,0)$.

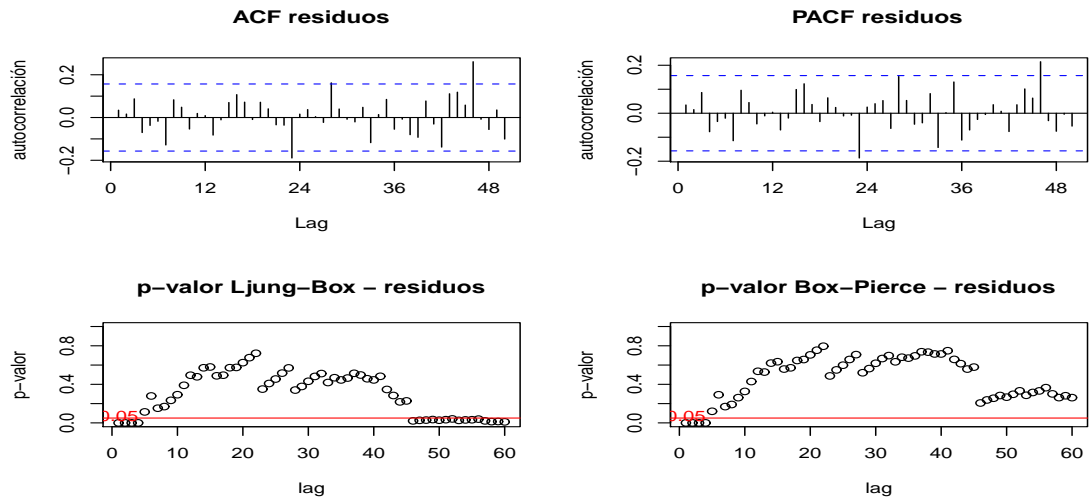


Figura A.4: Prueba de incorrelación de residuos para el modelo $SARIMA(0,0,2)(1,0,0)$.

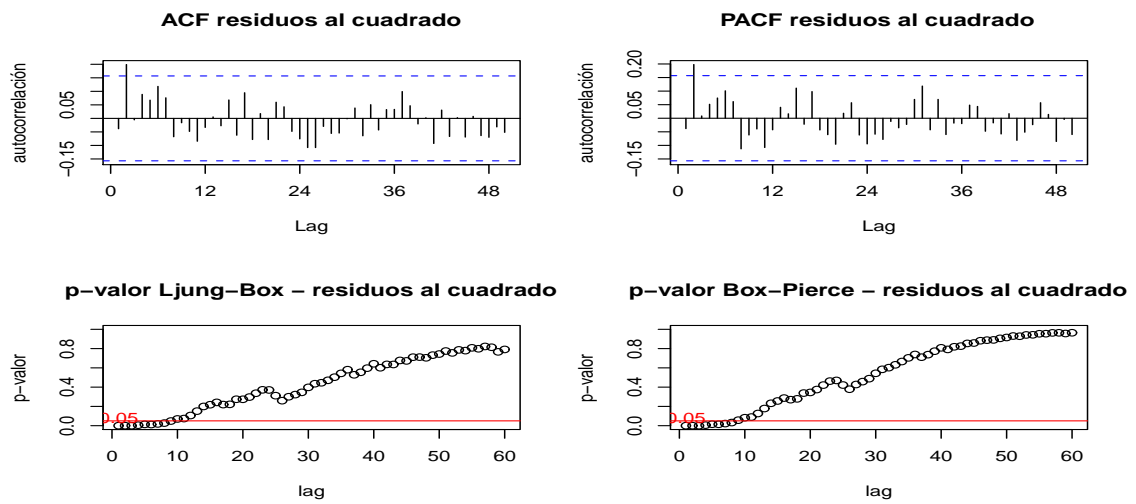


Figura A.5: Prueba de incorrelación de residuos al cuadrado para el modelo $SARIMA(1,0,0)(1,0,0)$.

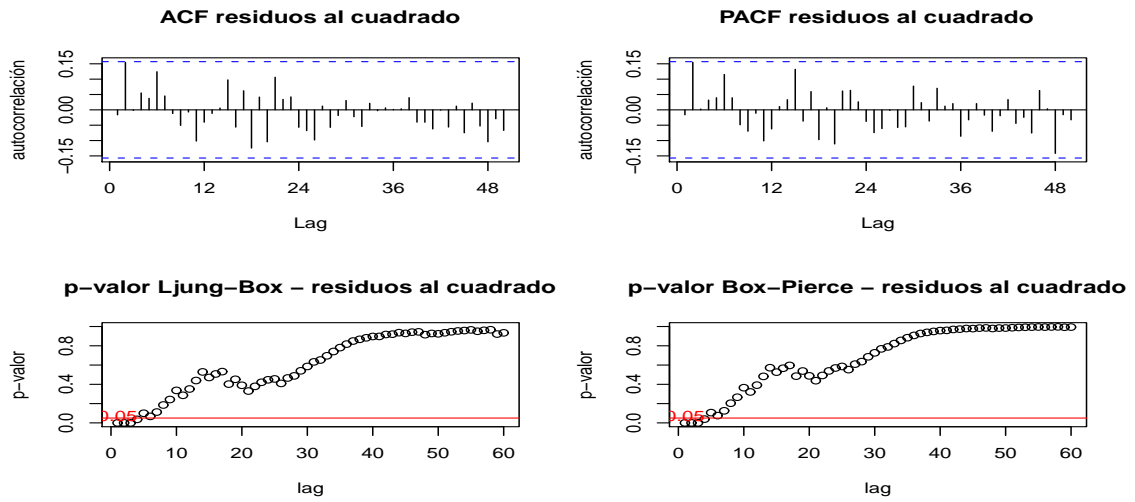


Figura A.6: Prueba de incorrelación de residuos al cuadrado para el modelo $SARIMA(0, 0, 1)(1, 0, 0)$.

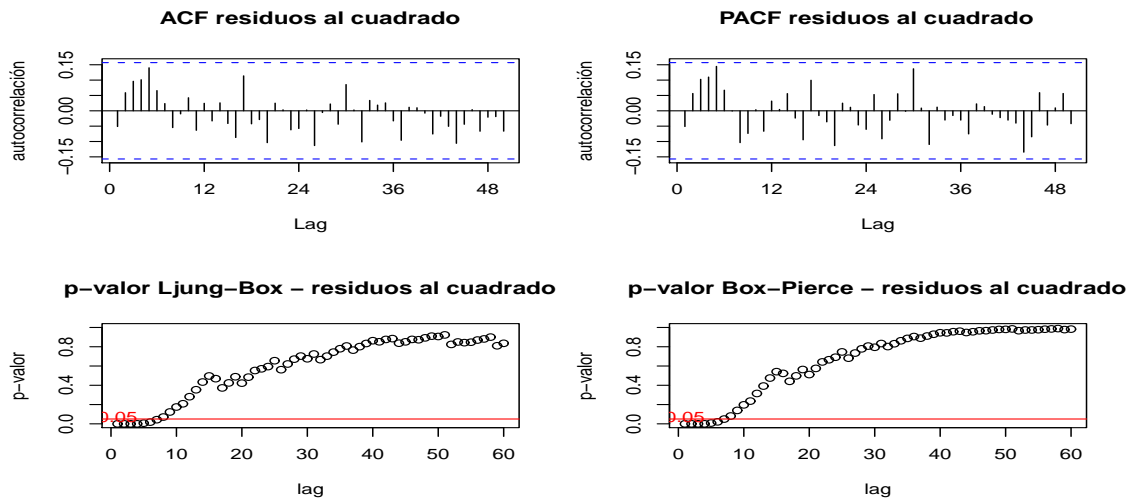


Figura A.7: Prueba de incorrelación de residuos al cuadrado para el modelo $SARIMA(0, 0, 2)(1, 0, 0)$.

Media cero

modelo	estimación media	estimación desvío	t-valor	p-valor
$SARIMA(1, 0, 0)(1, 0, 0)$	2.98	206.61	0.18	0.86
$SARIMA(0, 0, 2)(1, 0, 0)$	2.98	201.63	0.18	0.85

Cuadro A.1: P -valor prueba de media cero.

Homoscedasticidad

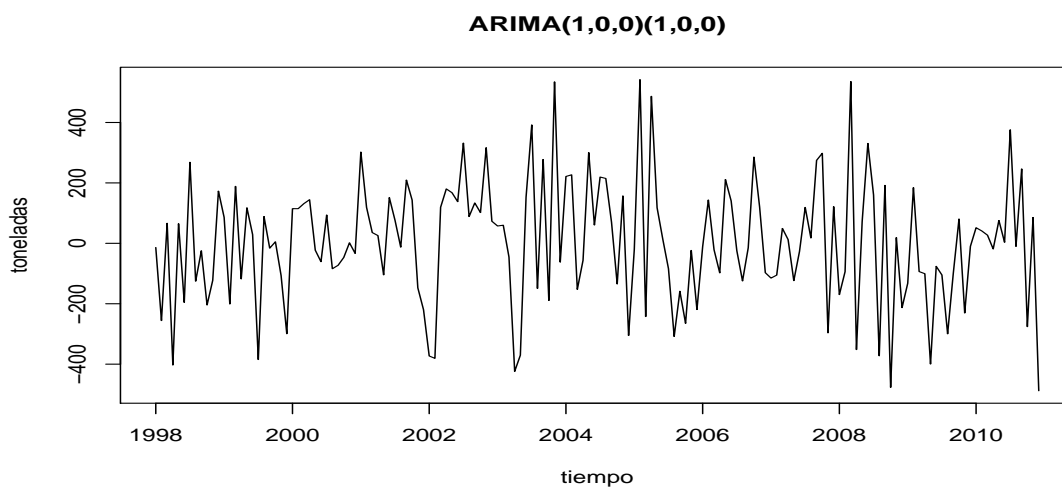


Figura A.8: Residuos del modelo $SARIMA(1,0,0)(1,0,0)$.

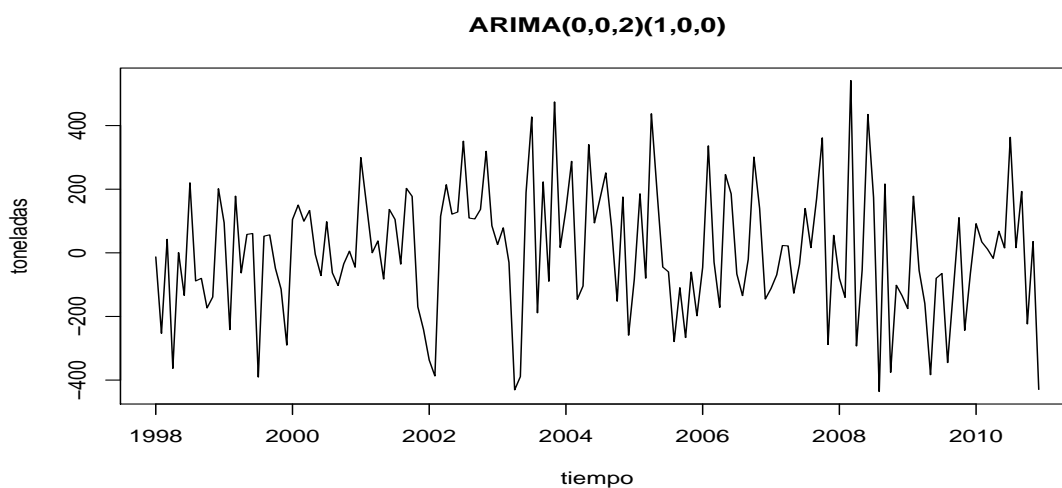


Figura A.9: Residuos del modelo $SARIMA(0,0,2)(1,0,0)$.

Normalidad

modelo	Jarque-Bera	Shapiro-Wilks
$SARIMA(1,0,0)(1,0,0)$	0.9807	0.3324
$SARIMA(0,0,2)(1,0,0)$	0.9378	0.4528

Cuadro A.2: P -valor pruebas de normalidad.

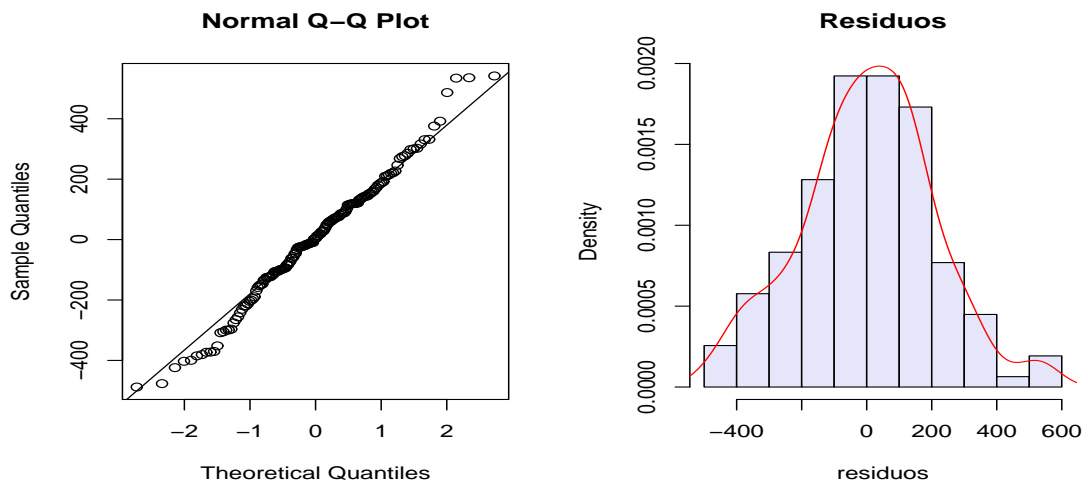


Figura A.10: Qq-plot e histograma de los residuos para el modelo $SARIMA(1, 0, 0)(1, 0, 0)$.

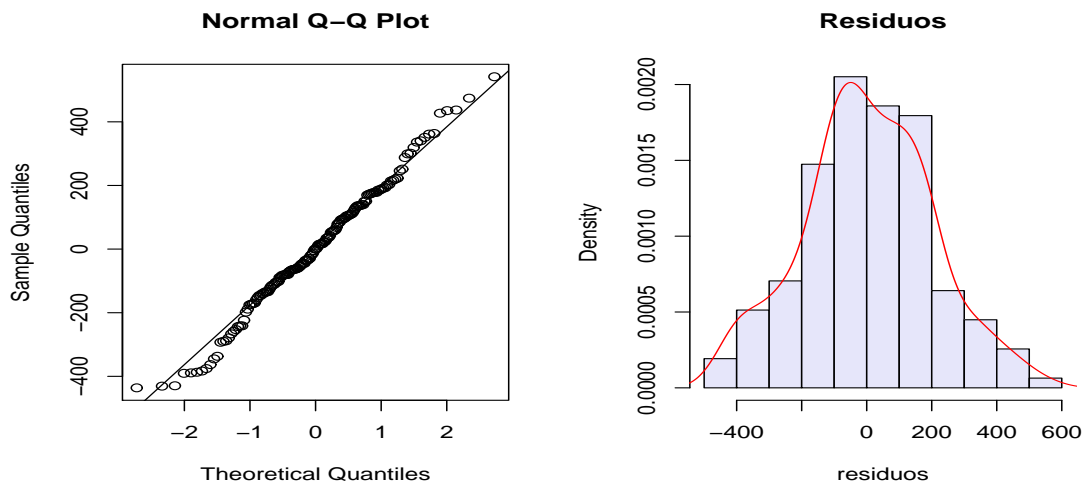


Figura A.11: Qq-plot e histograma de los residuos para el modelo $SARIMA(0, 0, 2)(1, 0, 0)$.

A.2. Sulfato de aluminio

ARIMA(0,1,1)					ARIMA(1,1,1)				
parámetro	valor	desvío	t-valor	p-valor	parámetro	valor	desvío	t-valor	p-valor
θ_1	-0.69	0.09	-7.55	3.48e-12	ϕ_1	0.21	0.11	1.98	4.96e-02
$AO_{may/00}$	389.48	76.38	5.10	9.88e-07	θ_1	-0.82	0.06	-13.61	0.00e+00
$AO_{ago/03}$	898.96	74.18	12.12	0.00e+00	$AO_{may/00}$	376.02	74.30	5.06	1.18e-06
$AO_{oct/08}$	347.82	73.84	4.71	5.47e-06	$AO_{ago/003}$	878.15	73.80	11.90	0.00e+00
					$AO_{oct/08}$	350.41	72.83	4.81	3.56e-06

ARIMA(0,1,2)				
parámetro	valor	desvío	t-valor	p-valor
θ_1	-0.64	0.08	-8.19	9.48e-14
θ_2	-0.12	0.08	-1.54	1.26e-01
$AO_{may/00}$	385.51	74.37	5.18	6.79e-07
$AO_{ago/03}$	883.47	74.48	11.86	0.00e+00
$AO_{oct/08}$	347.97	73.45	4.74	4.90e-06

Cuadro A.3: Significación de parámetros de modelos con tres intervenciones, sulfato de aluminio.

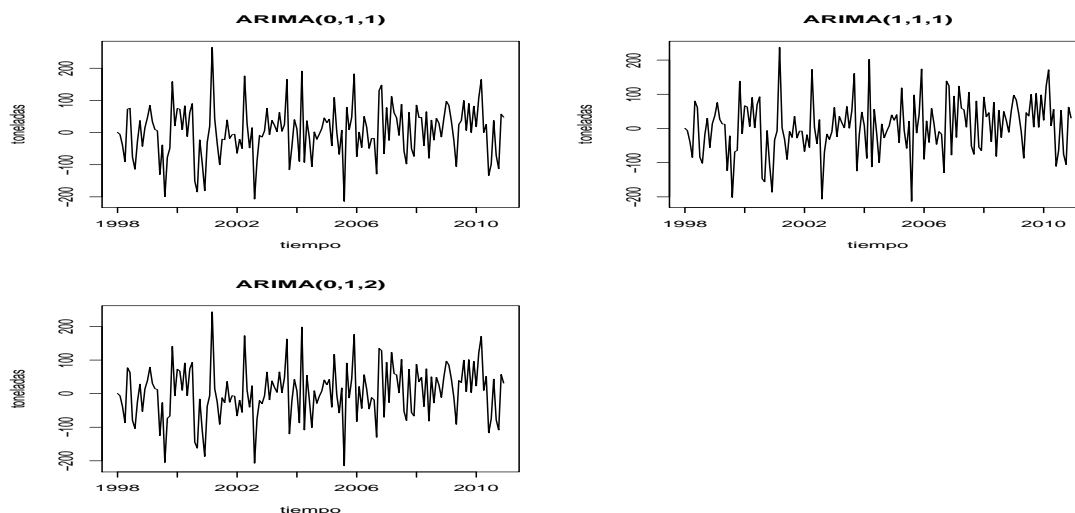


Figura A.12: Residuos de los modelos intervenidos con 3 AO.

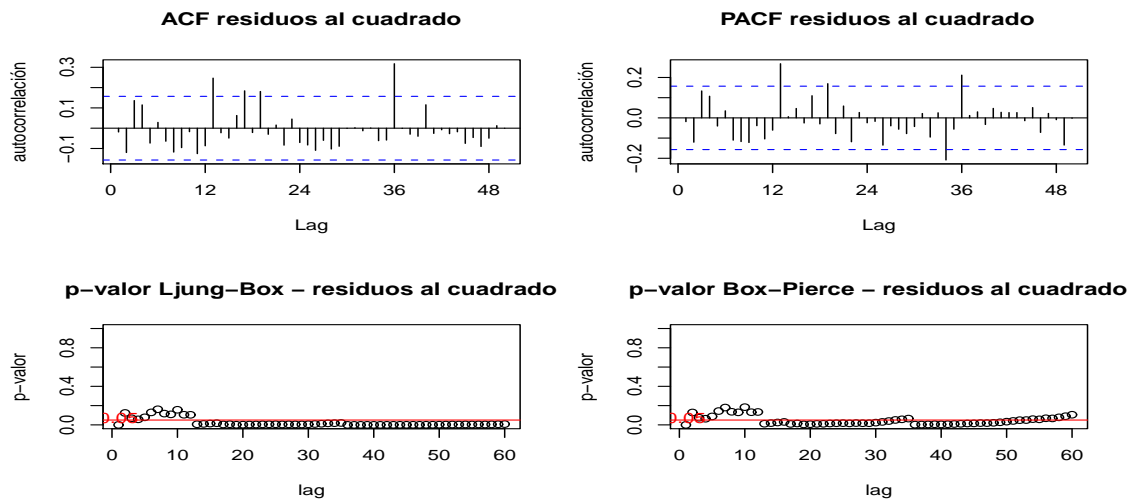


Figura A.13: Prueba de incorrelación para los residuos al cuadrado del modelo $ARIMA(0,1,1)$ intervenido con 3 AO.

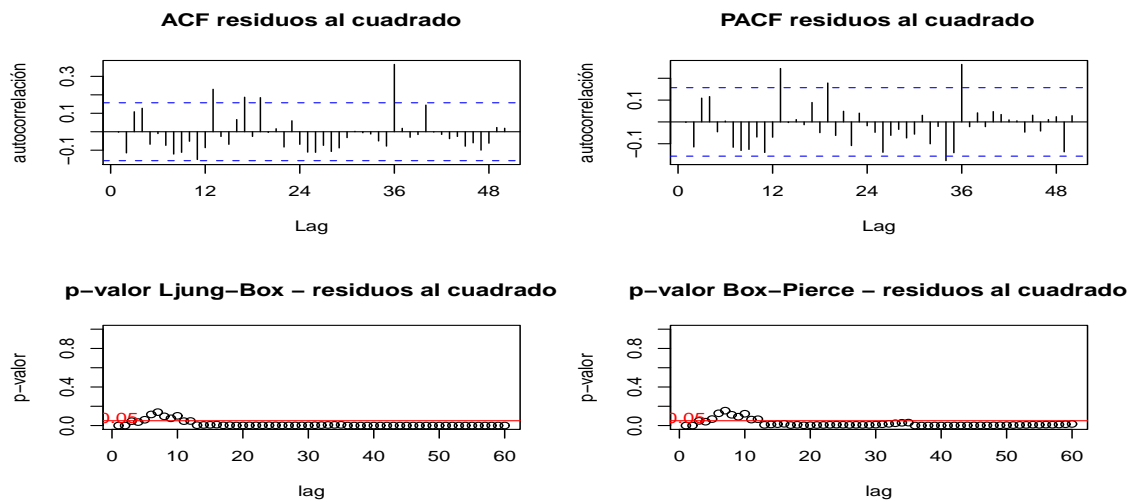


Figura A.14: Prueba de incorrelación para los residuos al cuadrado del modelo $ARIMA(1,1,1)$ intervenido con 3 AO.

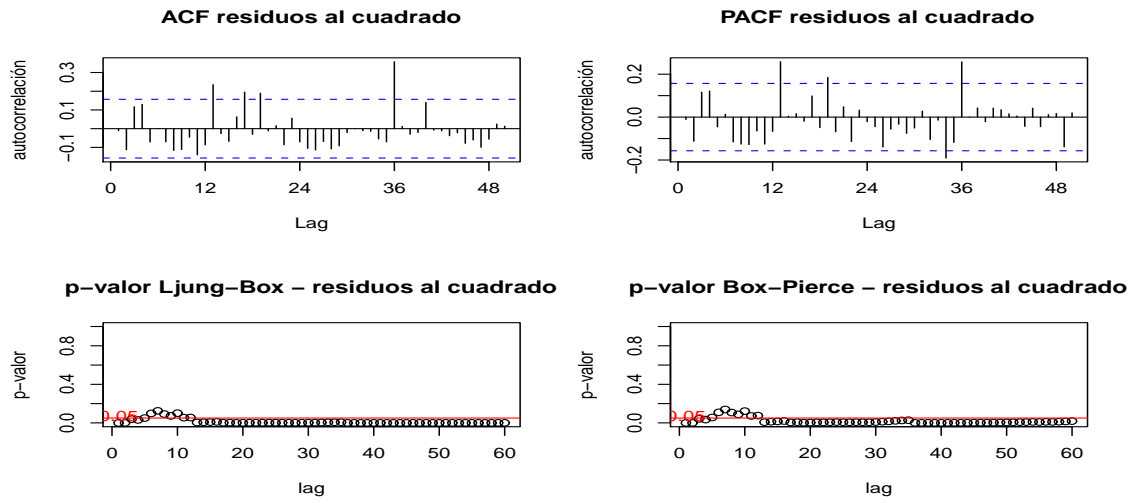


Figura A.15: Prueba de incorrelación para los residuos al cuadrado del modelo $ARIMA(0,1,2)$ intervenido con 3 AO.

Gráficos de los residuos

Incorrelación

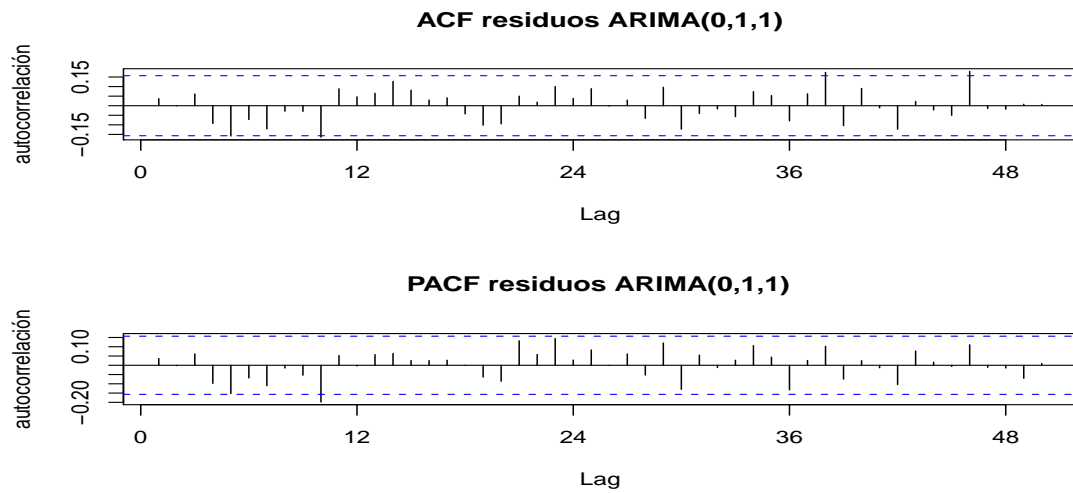


Figura A.16: Autocorrelogramas de los residuos, modelo $ARIMA(0,1,1)$ intervenido con seis AO.

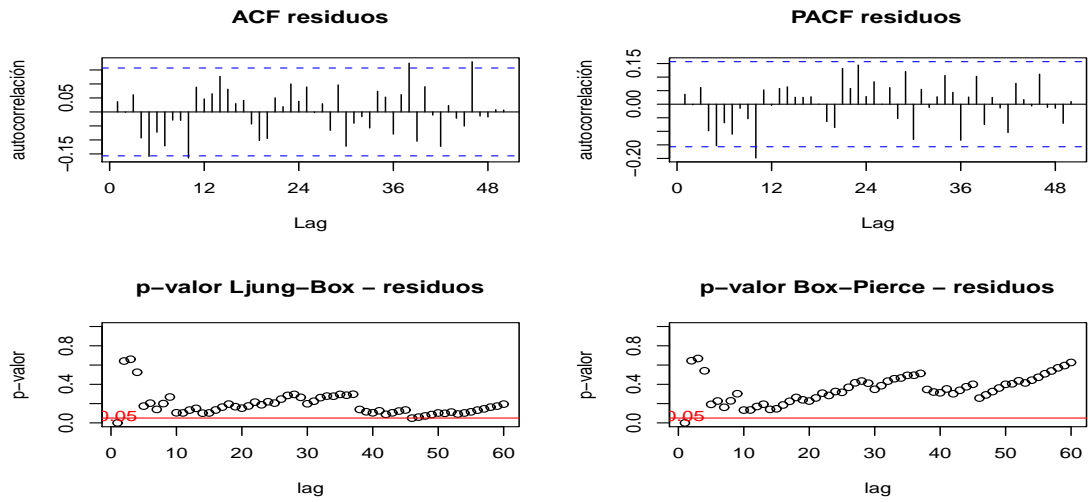


Figura A.17: Prueba de incorrelación de residuos para el modelo $ARIMA(0, 1, 1)$.

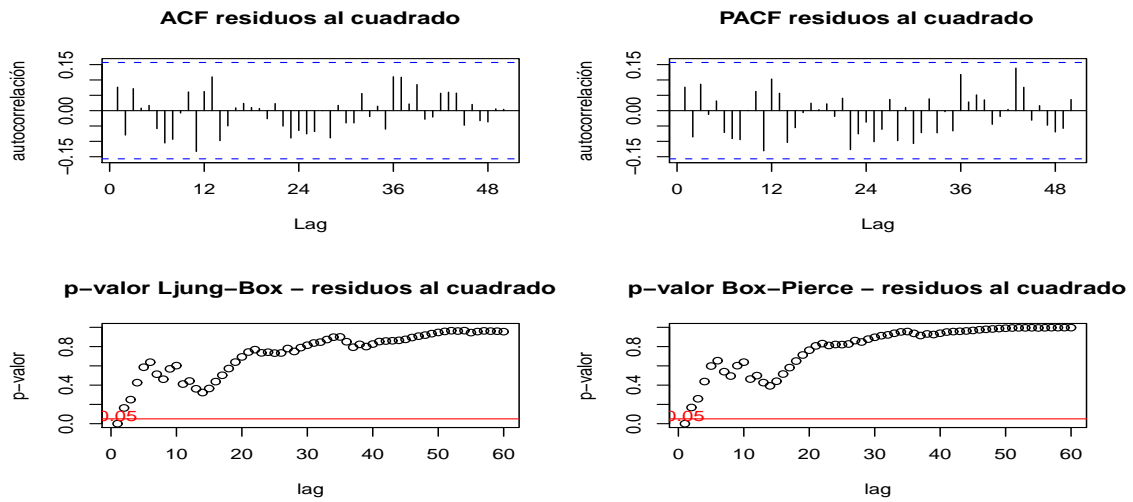


Figura A.18: Prueba de incorrelación de los residuos al cuadrado para el modelo $ARIMA(0, 1, 1)$.

Media cero

modelo	estimación media	estimación desvío	t-valor	p-valor
$ARIMA(0, 1, 1)$	2.69	73.92	0.45	0.32

Cuadro A.4: P -valor prueba de media cero.

Homoscedasticidad

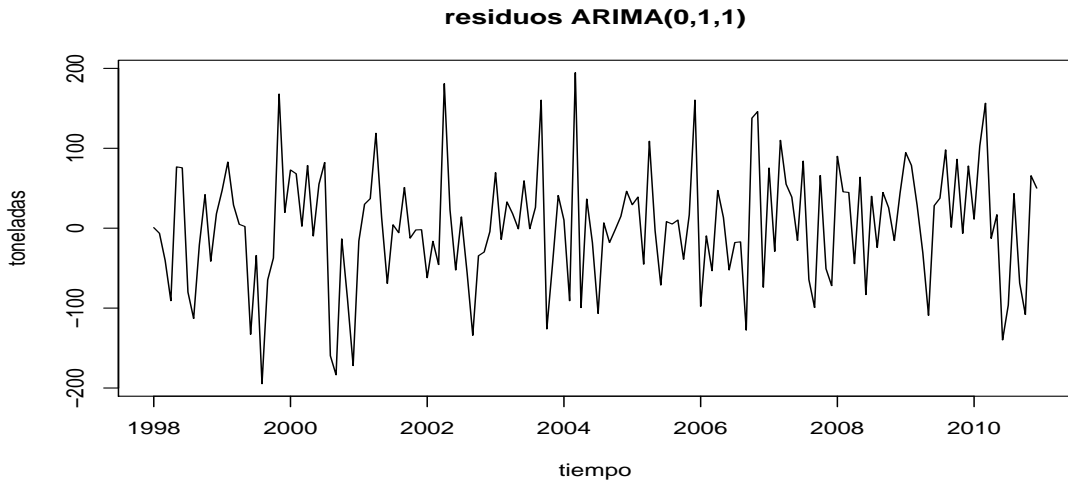


Figura A.19: Residuos modelo $ARIMA(0, 1, 1)$ con seis intervenciones.

Normalidad

modelo	Jarque-Bera	Shapiro-Wilks
$ARIMA(0, 1, 1)$	0.8375	0.5367

Cuadro A.5: *P-valor* pruebas de normalidad.

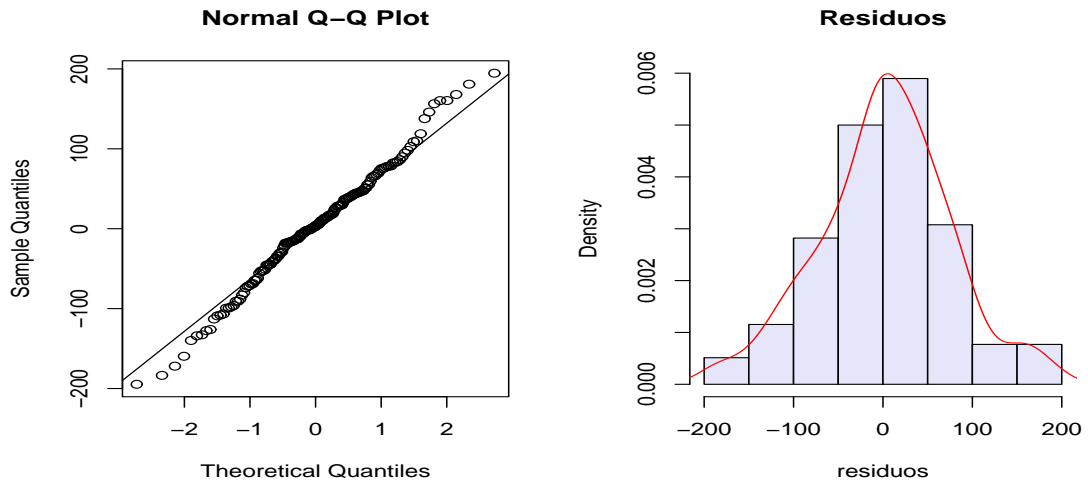
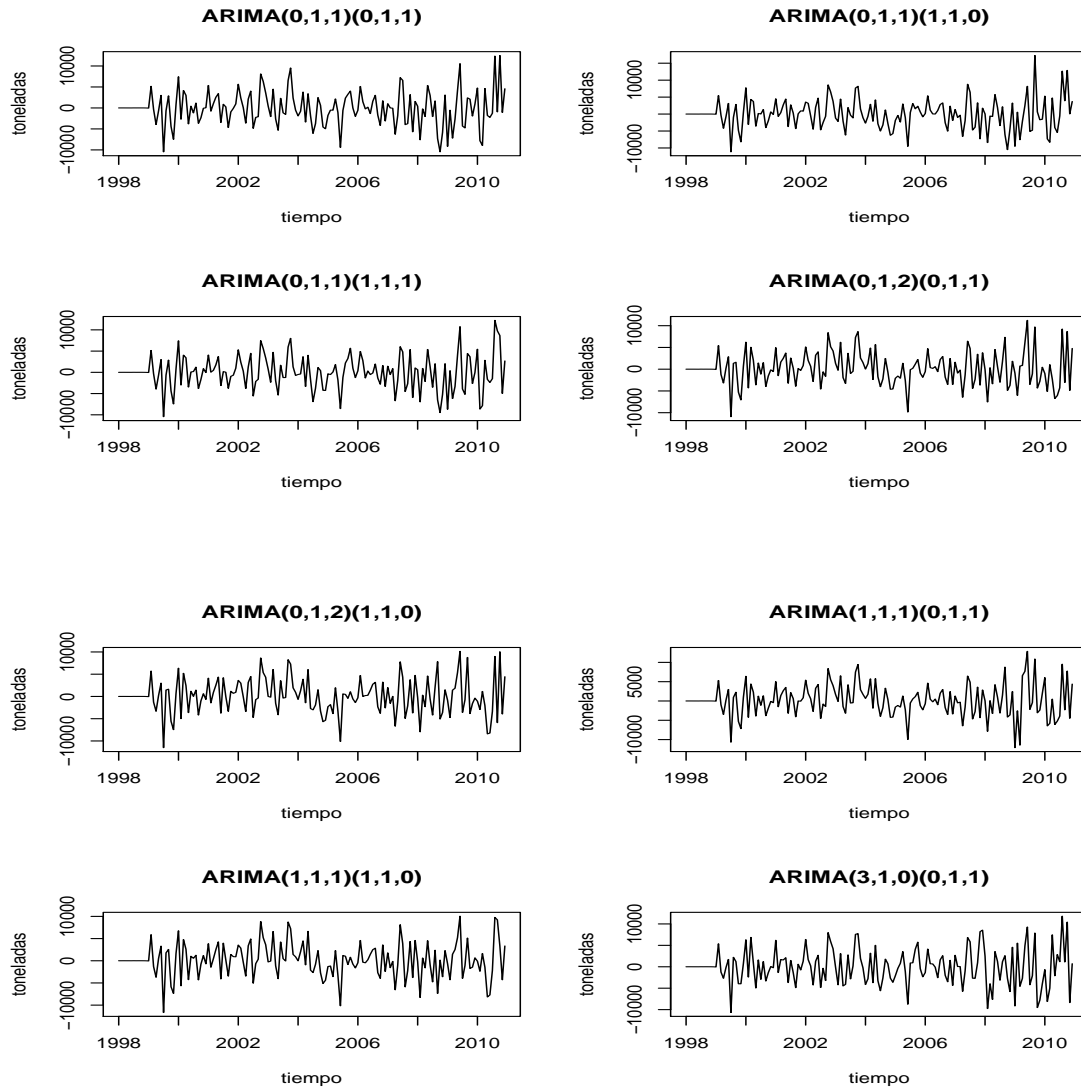


Figura A.20: Qq-plot e histograma de los residuos para el modelo $ARIMA(0, 1, 1)$.

A.3. Fertilizantes

A.3.1. Modelo univariado

Estimación



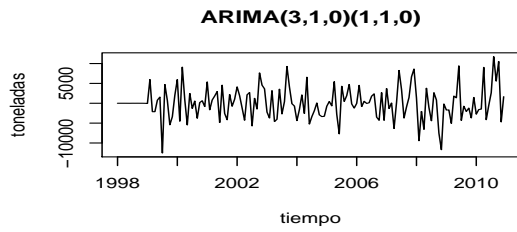
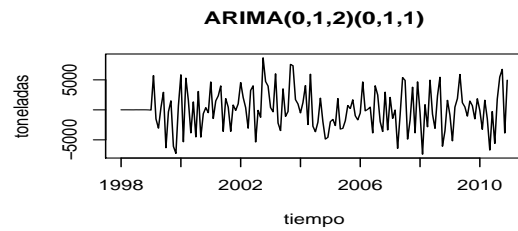
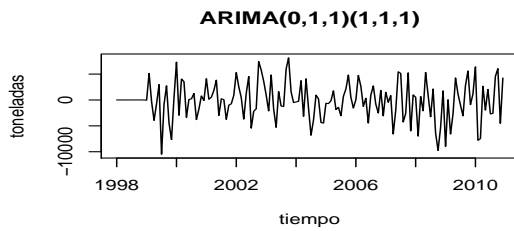
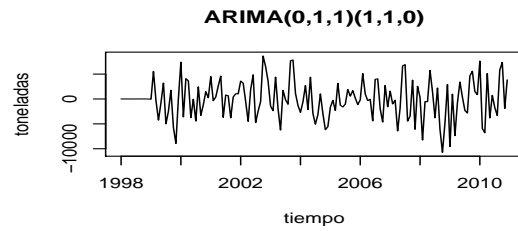
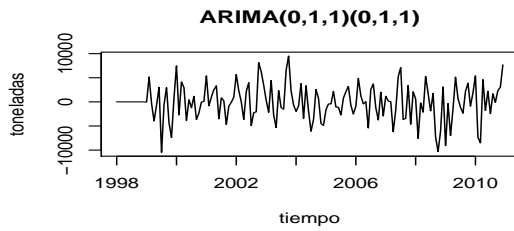


Figura A.21: Residuos de los modelos con detección automática de atípicos, serie de ventas de fertilizantes plaza.



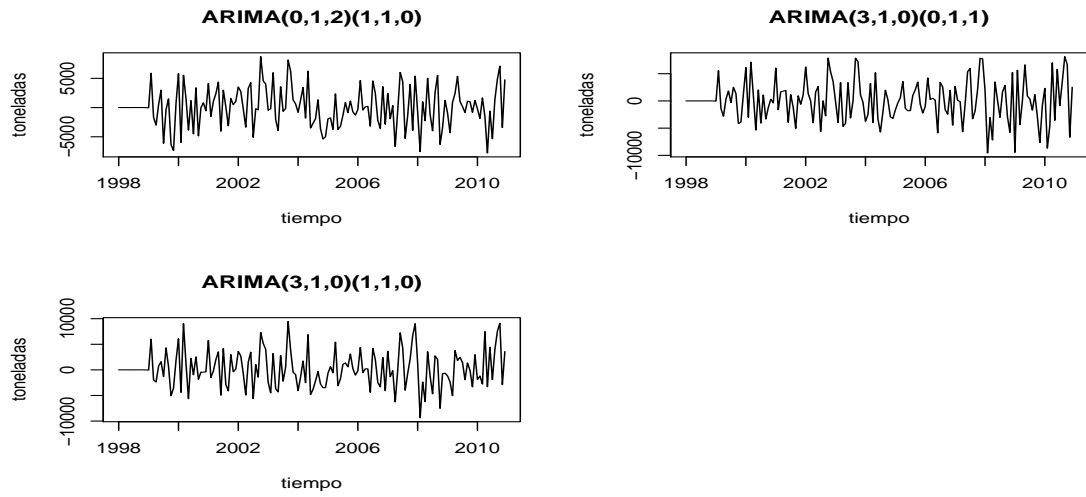


Figura A.22: Residuos de los modelos candidatos luego de incorporar nuevas intervenciones.

Incorrelación

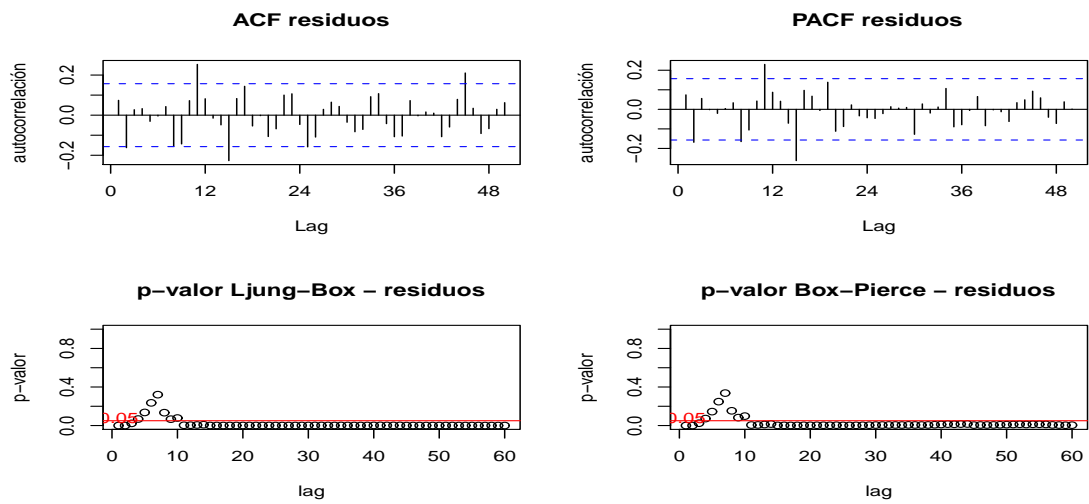


Figura A.23: Prueba de incorrelación de residuos para el modelo $SARIMA(0, 1, 1)(0, 1, 1)$.

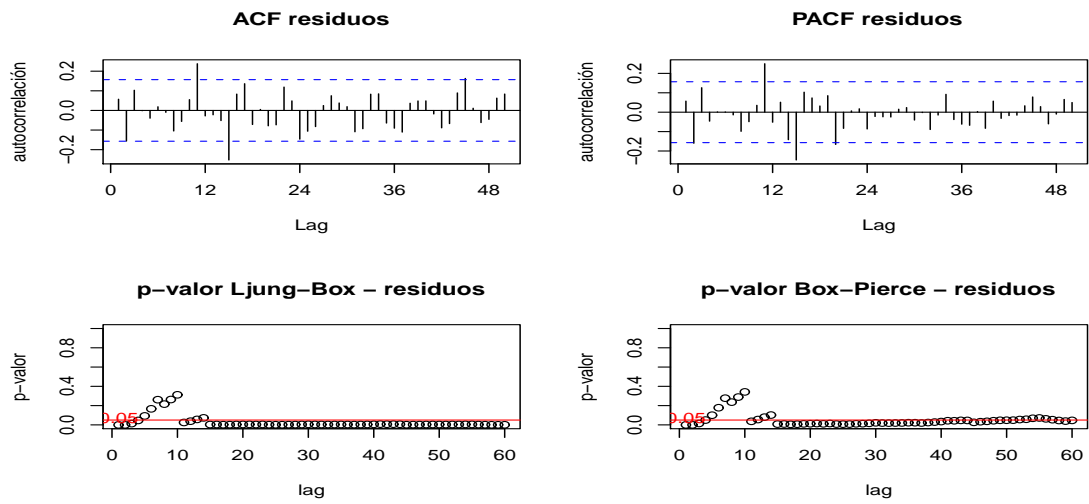


Figura A.24: Prueba de incorrelación de residuos para el modelo $SARIMA(0, 1, 1)(1, 1, 0)$.

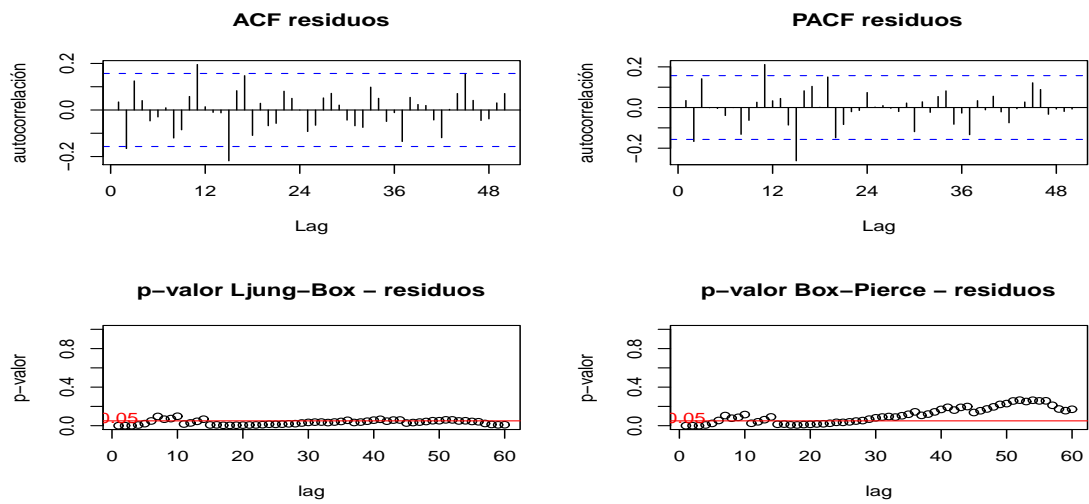


Figura A.25: Prueba de incorrelación de residuos para el modelo $SARIMA(0, 1, 1)(1, 1, 1)$.

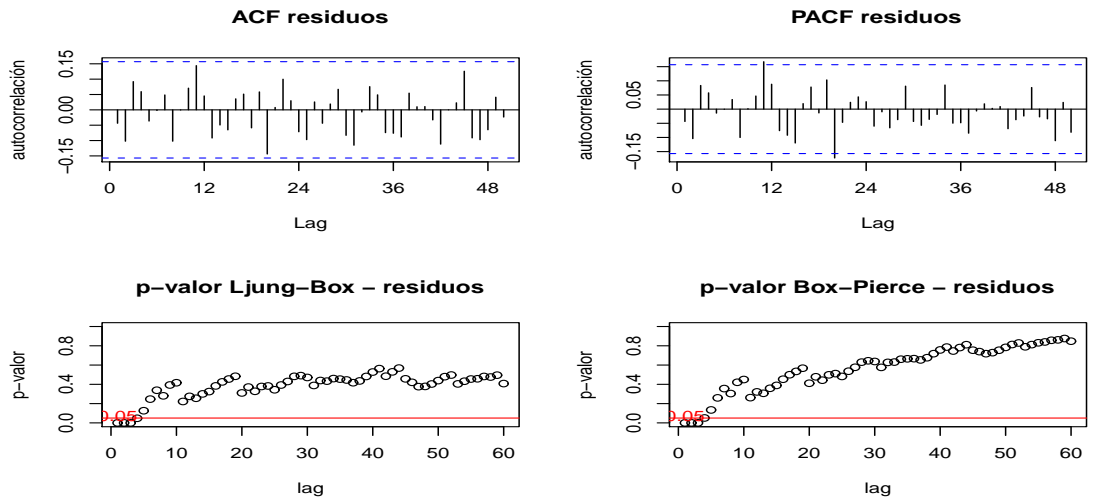


Figura A.26: Prueba de incorrelación de residuos para el modelo $SARIMA(0, 1, 2)(0, 1, 1)$.

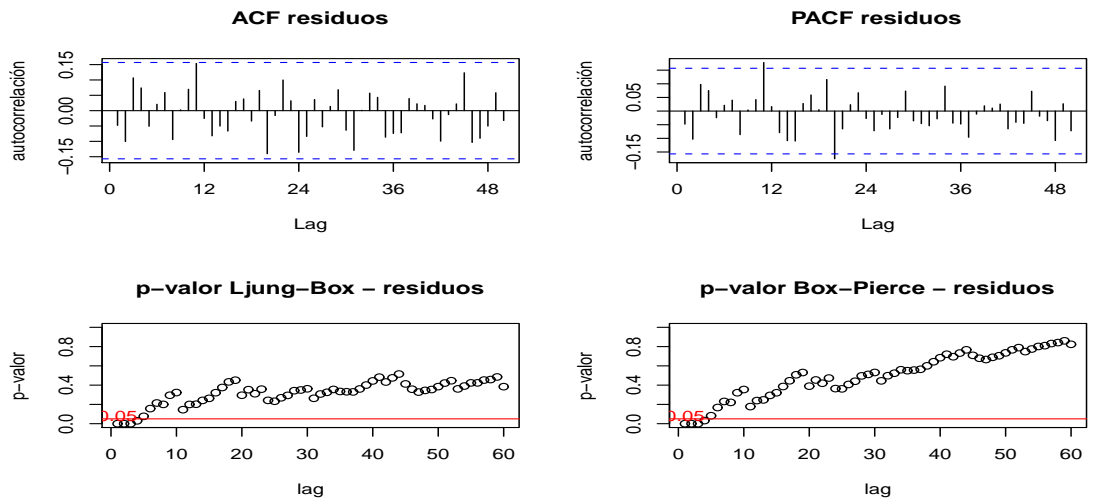


Figura A.27: Prueba de incorrelación de residuos para el modelo $SARIMA(0, 1, 2)(1, 1, 0)$.

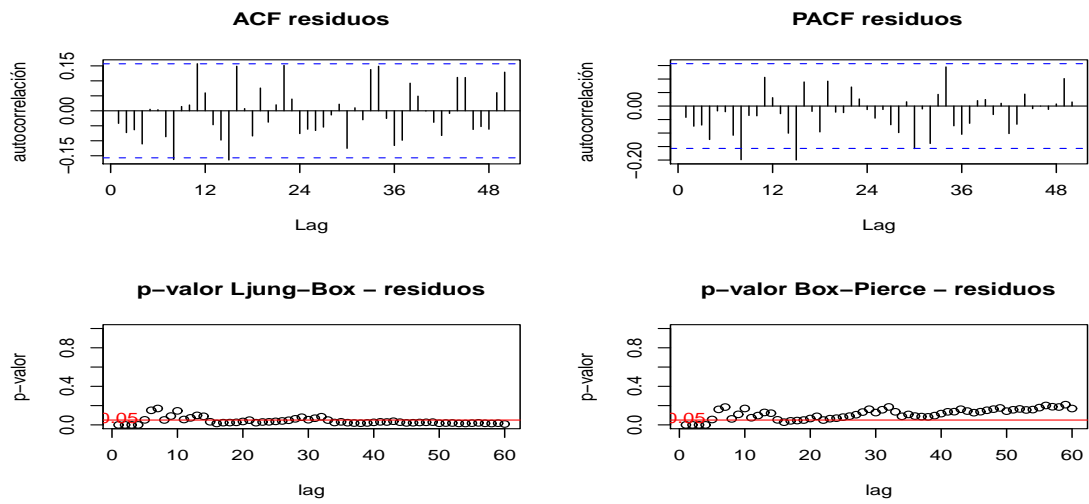


Figura A.28: Prueba de incorrelación de residuos para el modelo $SARIMA(3, 1, 0)(0, 1, 1)$.

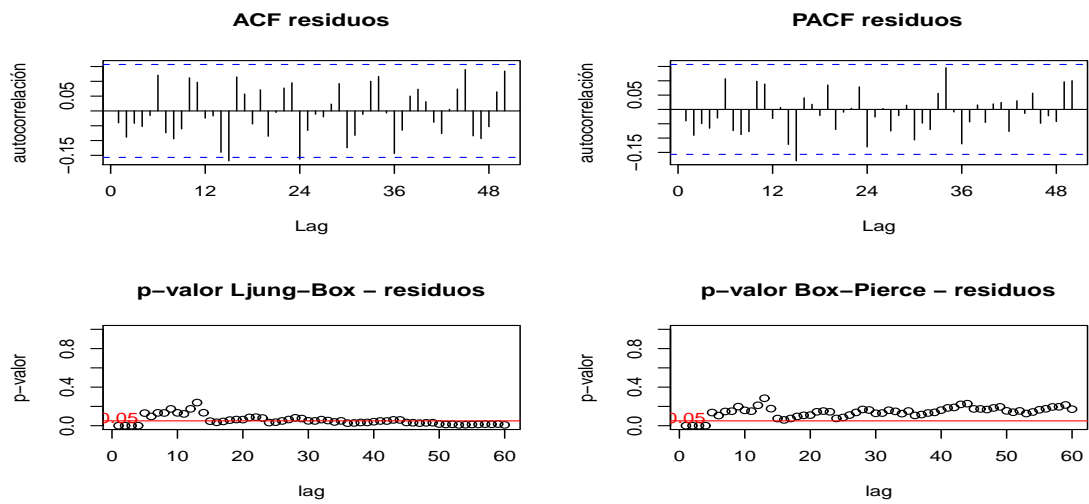


Figura A.29: Prueba de incorrelación de residuos para el modelo $SARIMA(3, 1, 0)(1, 1, 0)$.

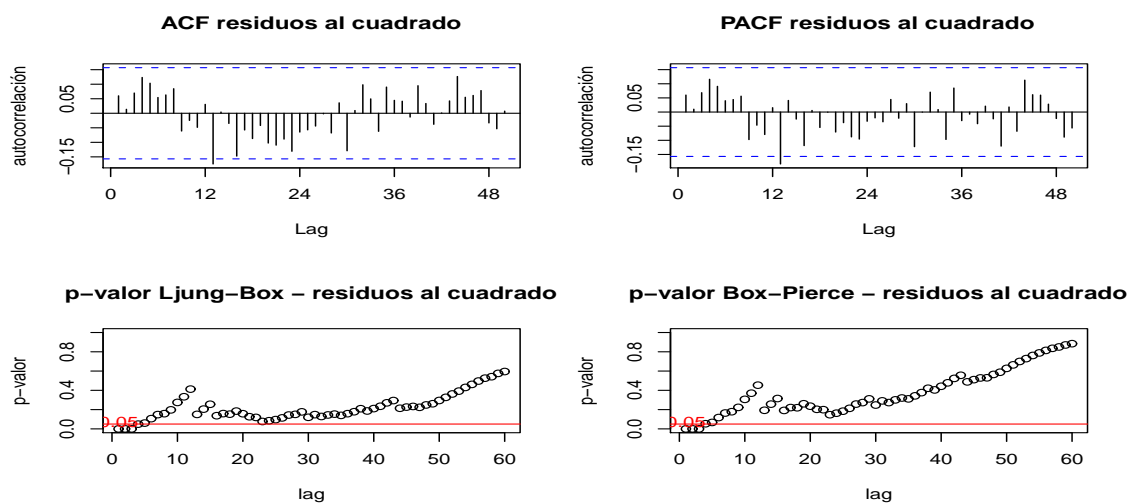


Figura A.30: Prueba de incorrelación de los residuos al cuadrado para el modelo $SARIMA(0, 1, 2)(0, 1, 1)$.

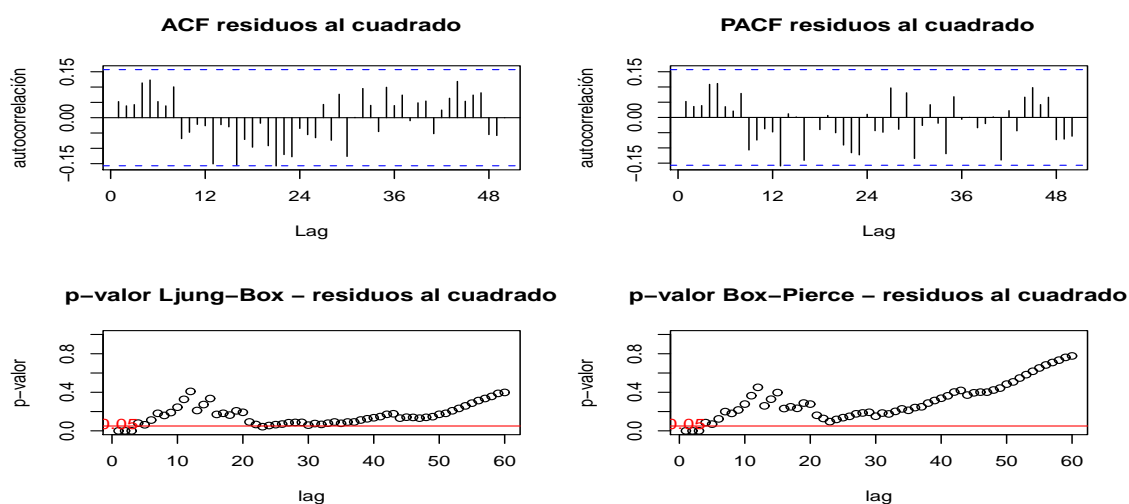


Figura A.31: Prueba de incorrelación de los residuos al cuadrado para el modelo $SARIMA(0, 1, 2)(1, 1, 0)$.

Media cero

modelo	estimación media	estimación desvío	t-valor	p-valor
$SARIMA(0, 1, 2)(0, 1, 1)$	70.23	3529.45	0.24	0.81
$SARIMA(0, 1, 2)(1, 1, 0)$	60.97	3580.43	0.20	0.84

Cuadro A.6: P -valor prueba de media cero.

Homoscedasticidad

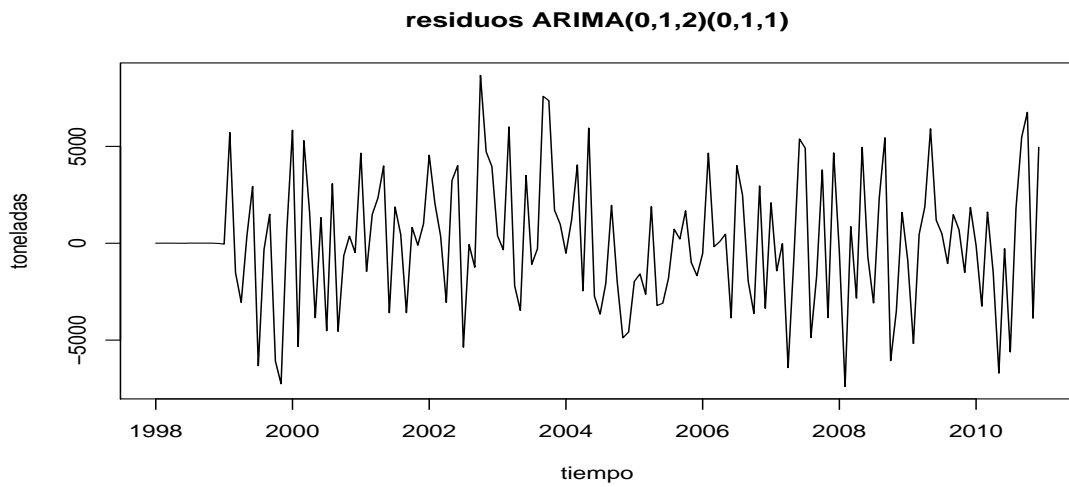


Figura A.32: Residuos del modelo $SARIMA(0, 1, 2)(0, 1, 1)$.

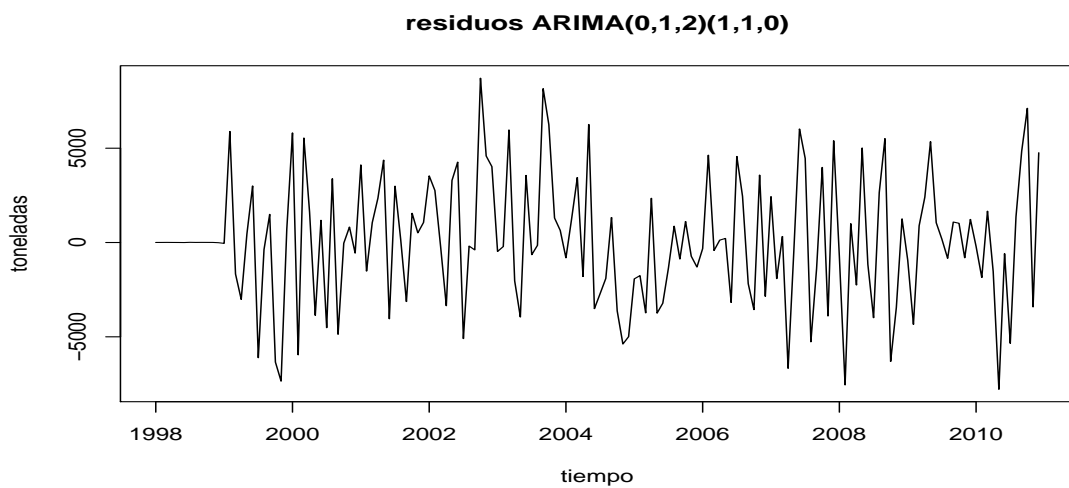


Figura A.33: Residuos del modelo $SARIMA(0, 1, 2)(1, 1, 0)$.

Normalidad

modelo	Jarque-Bera	Shapiro-Wilks
$SARIMA(0, 1, 2)(0, 1, 1)$	0.6304	0.2184
$SARIMA(0, 1, 2)(1, 1, 0)$	0.7672	0.2279

Cuadro A.7: P -valor pruebas de normalidad.

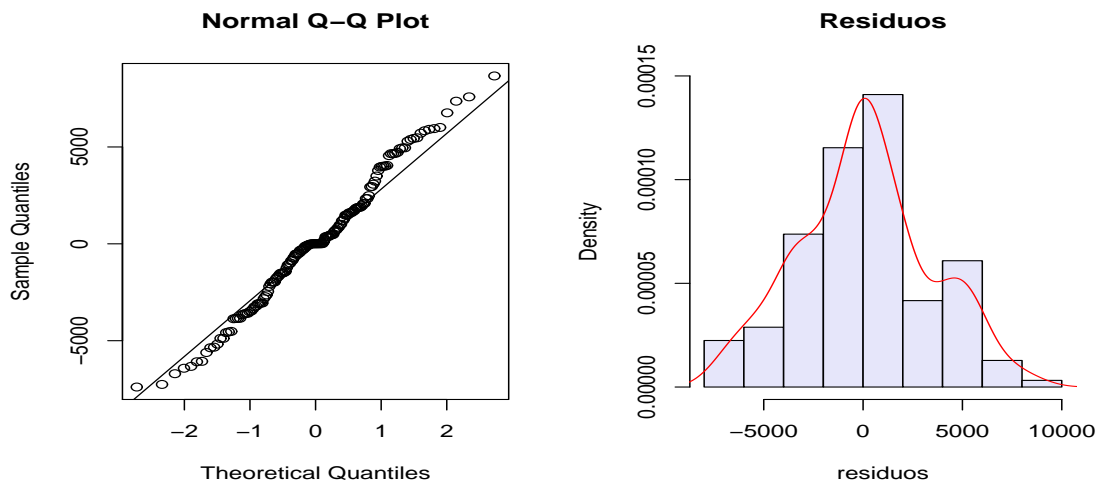


Figura A.34: Qq-plot e histograma de los residuos para el modelo $SARIMA(0, 1, 2)(0, 1, 1)$.

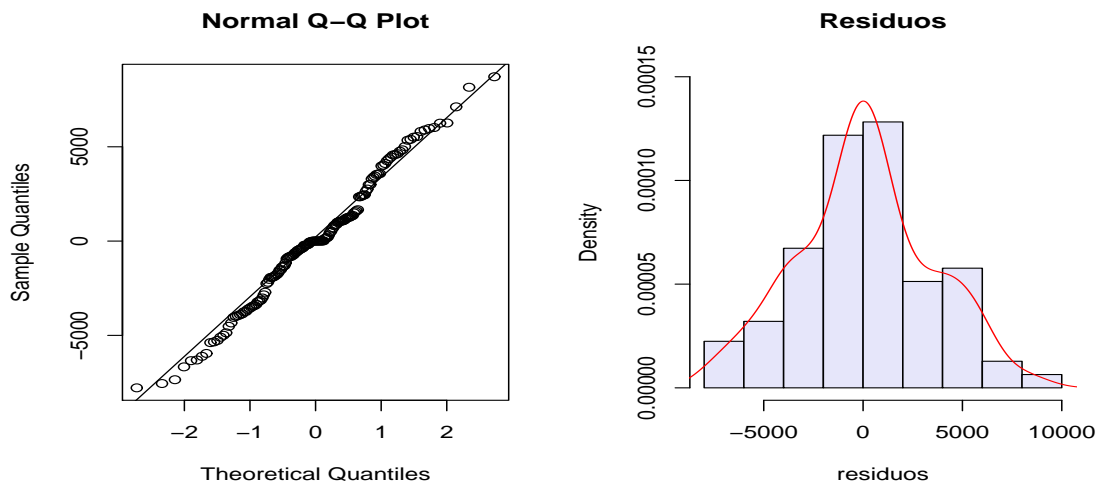


Figura A.35: Qq-plot e histograma de los residuos para el modelo $SARIMA(0, 1, 2)(1, 1, 0)$.

A.3.2. Modelo multivariado

Pruebas ADF aplicadas a la serie de venta de fertilizantes

En esta sección se sigue el procedimiento que se describe en el Anexo B.

En primer lugar se estima el modelo más general, incluyendo constante y tendencia lineal, con 12 rezagos.

Como el valor del estadístico τ_3 es mayor al valor crítico al 5% no se rechaza la hipótesis nula de que $\gamma = 0$. A su vez como el valor de ϕ_3 es menor al valor crítico correspondiente tampoco se rechaza la hipótesis conjunta de que $\gamma = a_2 = 0$ ¹. Los resultados se muestran en la tabla A.8.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_3	-2.90	-3.99	-3.43	-3.13
ϕ_2	3.03	6.22	4.75	4.07
ϕ_3	4.22	8.43	6.49	5.47

Cuadro A.8: Resultado de la prueba ADF a la serie de fertilizantes, modelo con constante y tendencia lineal.

Luego se estima el modelo sin tendencia y con constante, cuyos resultados se muestran en la tabla A.9.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_2	-1.29	-3.46	-2.88	-2.57
ϕ_1	1.13	6.52	4.63	3.81

Cuadro A.9: Resultado de la prueba ADF a la serie de fertilizantes, modelo con constante y sin tendencia lineal.

Como el valor del estadístico τ_2 es mayor al valor crítico al 5% no se rechaza que $\gamma = 0$. Por otro lado como el valor del estadístico ϕ_1 es menor al valor crítico correspondiente al 5% no se rechaza que $\gamma = a_0 = 0$ y se estima el modelo sin constante ni tendencia lineal. Los resultados figuran en la tabla A.10.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	0.42	-2.58	-1.95	-1.62

Cuadro A.10: Resultado de la prueba ADF a la serie de fertilizantes, modelo sin constante ni tendencia lineal.

En esta prueba el valor del estadístico τ_1 es mayor al valor crítico para el nivel de significación considerado por lo que no se rechaza la hipótesis nula $\gamma = 0$. Lo anterior permite concluir que no se rechaza la hipótesis de que la serie de fertilizantes posee una raíz unitaria. Se aplica la prueba

¹ver anexo

ADF a la serie en diferencias para determinar el orden de integración. En este caso se estima el modelo sin constante ni tendencia lineal, los resultados figuran en la tabla A.11.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	-4.89	-2.58	-1.95	-1.62

Cuadro A.11: Resultado de la prueba ADF a la serie de fertilizantes en diferencias, modelo sin constante ni tendencia lineal.

Se rechaza las hipótesis $\gamma = 0$ dado que el valor del estadístico es menor al valor crítico al 5% por lo que se concluye que la serie de fertilizantes en diferencias no posee raíz unitaria, entonces la serie de fertilizantes en niveles es integrada de orden uno (I(1)).

Pruebas ADF aplicadas a la serie de PBI agropecuario

Nuevamente se aplica el procedimiento de estimar secuencialmente distintos modelos y se emplea la prueba ADF para determinar en este caso si la serie de PBI agropecuario es integrada y cual es el orden de integración.

Al igual que con la serie de fertilizantes se estimaron modelos con 12 rezagos de la propia serie como variables explicativas para obtener residuos incorrelacionados. Los resultados se muestran en las tablas A.12, A.13 y A.14.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_3	-2.31	-3.99	-3.43	-3.13
ϕ_2	2.61	6.22	4.75	4.07
ϕ_3	3.62	8.43	6.49	5.47

Cuadro A.12: Resultado de la prueba ADF a la serie de PBI agropecuario, modelo con constante y tendencia lineal.

El estadístico τ_3 es mayor al valor crítico correspondiente y el estadístico ϕ_3 es menor a su respectivo valor crítico por lo que no se rechazan las hipótesis $\gamma = 0$ ni $\gamma = a_2 = 0$.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_2	-0.33	-3.46	-2.88	-2.57
ϕ_1	0.34	6.52	4.63	3.81

Cuadro A.13: Resultado de la prueba ADF a la serie de PBI agropecuario, modelo con constante y sin tendencia lineal.

El valor de τ_2 es mayor al valor crítico y el de ϕ_1 es menor a su valor crítico por lo que no se rechaza que $\gamma = 0$ ni que $\gamma = a_0 = 0$.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	0.73	-2.58	-1.95	-1.62

Cuadro A.14: Resultado de la prueba ADF a la serie de PBI agropecuario, modelo sin constante ni tendencia lineal.

En este caso τ_1 es mayor al valor crítico por lo que no se rechaza que $\gamma = 0$.

En el primer modelo, con constante y tendencia, no se rechazan ninguna de las dos hipótesis, tanto la individual como la conjunta por lo que se estima el segundo modelo, con constante pero sin tendencia, donde tampoco se rechazan las dos hipótesis. Luego se estima el tercer modelo, sin constante ni tendencia, en este caso tampoco se rechaza la hipótesis de que $\gamma = 0$ por lo que se concluye que la serie de fertilizantes posee raíz unitaria.

En una segunda instancia se aplica el procedimiento ADF a las serie de PBI agropecuario en diferencias, en este caso se estima el modelo sin constante ni tendencia y se concluye que la serie en diferencias no es integrada, por lo que la serie de PBI agropecuario es $I(1)$. Los resultados de la prueba se muestran en la tabla A.15.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	-3.37	-2.58	-1.95	-1.62

Cuadro A.15: Resultado de la prueba ADF a la serie de PBI agropecuario en diferencias, modelo sin constante ni tendencia lineal.

Como el valor de τ_1 es menor al valor crítico, se rechaza que $\gamma = 0$ y se obtiene que la serie de PBI agropecuario en diferencias es estacionaria.

Pruebas ADF aplicadas a la serie de índice de precios de alimentos

En este caso se estimaron modelos con 2 rezagos, lo cual es suficiente para garantizar la incorrelación de los residuos.

Los resultados de aplicar el procedimiento de las pruebas ADF a esta serie se muestran en las tablas A.16, A.17 y A.18.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_3	-2.32	-3.99	-3.43	-3.13
ϕ_2	3.09	6.22	4.75	4.07
ϕ_3	3.87	8.43	6.49	5.47

Cuadro A.16: Resultado de la prueba ADF a la serie de índice de precios de alimentos, modelo con constante y tendencia lineal.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_2	0.19	-3.46	-2.88	-2.57
ϕ_1	0.75	6.52	4.63	3.81

Cuadro A.17: Resultado de la prueba ADF a la serie de índice de precios de alimentos, modelo con constante y sin tendencia lineal.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	1.23	-2.58	-1.95	-1.62

Cuadro A.18: Resultado de la prueba ADF a la serie de índice de precios de alimentos, modelo sin constante ni tendencia lineal.

Los resultados indican que la serie es integrada. Para determinar el orden de integración se aplicó el procedimiento ADF a la serie en diferencias estimando un modelo sin constante ni tendencia, donde se obtuvieron los resultados que se muestran en la tabla A.19.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	-4.01	-2.58	-1.95	-1.62

Cuadro A.19: Resultado de la prueba ADF a la serie de índice de precios de alimentos en diferencias, modelo sin constante ni tendencia lineal.

Se rechaza la hipótesis nula de $\gamma = 0$ para la serie de índice de precios de alimentos en diferencias, la serie de índice de precios de alimentos en niveles es $I(1)$.

Pruebas ADF aplicadas a la serie de precio de fertilizantes

Para esta serie se estimaron modelos con 1 rezago, con lo cual se obtuvo incorrelación en los residuos.

Los resultados de aplicar el procedimiento de las pruebas ADF a esta serie se muestran en las tablas A.20, A.21 y A.22.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_3	-2.90	-3.99	-3.43	-3.13
ϕ_2	2.88	6.22	4.75	4.07
ϕ_3	4.26	8.43	6.49	5.47

Cuadro A.20: Resultado de la prueba ADF a la serie de precio de fertilizantes, modelo con constante y tendencia lineal.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_2	-2.12	-3.46	-2.88	-2.57
ϕ_1	2.31	6.52	4.63	3.81

Cuadro A.21: Resultado de la prueba ADF a la serie de precio de fertilizantes, modelo con constante y sin tendencia lineal.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	-0.42	-2.58	-1.95	-1.62

Cuadro A.22: Resultado de la prueba ADF a la serie de precio de fertilizantes, modelo sin constante ni tendencia lineal.

Una vez más los resultados indican que la serie es integrada. Para determinar el orden de integración se aplicó el procedimiento ADF a la serie en diferencias estimando un modelo sin constante ni tendencia, donde se obtuvieron los resultados que se muestran en la tabla A.23.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_1	-8.55	-2.58	-1.95	-1.62

Cuadro A.23: Resultado de la prueba ADF a la serie de precio de fertilizantes en diferencias, modelo sin constante ni tendencia lineal.

Se rechaza la hipótesis nula de $\gamma = 0$, la serie de precio de fertilizantes es $I(1)$.

Pruebas ADF aplicadas a la serie de nivel de lluvias en el país

Se estimaron modelos sin incluir rezagos de la serie en diferencias ya que así se obtuvo incorrelación en los residuos.

El resultado de la prueba para esta serie se muestra en el cuadro A.24. Se estimó un modelo con constante pero sin tendencia.

	estadístico	p-valor		
		1 %	5 %	10 %
τ_2	-10.09	-3.46	-2.88	-2.57
ϕ_1	50.97	6.52	4.63	3.81

Cuadro A.24: Resultado de la prueba ADF a la serie de nivel de lluvia en el país, modelo con constante y sin tendencia lineal.

En el caso de esta serie se rechaza la hipótesis de que $\gamma = 0$, por lo que se concluye que la misma no tiene raíz unitaria.

Apéndice B

Pruebas de existencia de raíz unitaria de Dickey Fuller

Se tiene el modelo

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

Se dice que el modelo posee una *raíz unitaria* cuando $a_1 = 1$. Restando de ambos lados y_{t-1} , se puede reexpresar al modelo como:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

donde $\Delta y_t = y_t - y_{t-1}$ y $\gamma = a_1 - 1$. En esta situación, testear la hipótesis $a_1 = 1$ es equivalente a testear $\gamma = 0$.

Dickey y Fuller consideran tres modelos diferentes que pueden ser usados para testear la presencia de raíz unitaria:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t \tag{B.1}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t \tag{B.2}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t \tag{B.3}$$

La diferencia entre los tres modelos se debe a la presencia de los elementos determinísticos a_0 y $a_2 t$. El modelo (B.1) es un paseo aleatorio¹ puro, el (B.2) es un modelo con constante y el (B.3) incluye una constante y una tendencia lineal.

En todos los modelos el parámetro de interés es γ ; si $\gamma = 0$ el proceso posee una raíz unitaria. La prueba implica estimar una o más de las ecuaciones anteriores usando MCO² para obtener valores estimados de γ y su desvío estándar asociado.

¹Caso particular de un proceso AR(1): $y_t = c + \phi y_{t-1} + \varepsilon_t$.

²Mínimos Cuadrados Ordinarios.

El estadístico no sigue una distribución t de Student, por lo que Dickey y Fuller crearon tablas para cada modelo que dependen del tamaño de la muestra. A su vez proveen tres estadísticos F adicionales (llamados ϕ_1 , ϕ_2 y ϕ_3) para testear hipótesis conjuntas sobre los coeficientes. El cuadro B.1 resume las pruebas de hipótesis que se pueden realizar a partir de los modelos y los estadísticos a utilizar.

modelo	hipótesis	estadístico de prueba
$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$	$\gamma = 0$	τ_τ
	$a_0 = 0 \gamma = 0$	$\tau_{\alpha\tau}$
	$a_2 = 0 \gamma = 0$	$\tau_{\beta\tau}$
	$\gamma = a_2 = 0$	ϕ_3
$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$	$a_0 = \gamma = 0$	ϕ_2
	$\gamma = 0$	τ_μ
	$a_0 = 0 \gamma = 0$	$\tau_{\alpha\mu}$
$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	$a_0 = \gamma = 0$	ϕ_1
	$\gamma = 0$	τ

Cuadro B.1: resumen de las pruebas de Dickey-Fuller

Los estadísticos ϕ_1 , ϕ_2 y ϕ_3 son construidos como estadísticos F :

$$\phi_i = \frac{[SCR(reducido) - SCR(completo)]/r}{SCR(completo)/(T - k)} \quad (B.4)$$

donde $SCR(reducido)$ ³ y $SCR(completo)$ son las suma de cuadrados de los residuos del modelo reducido (bajo hipótesis nula) y la suma de cuadrados de los residuos del modelo completo (bajo hipótesis alternativa) respectivamente. r es el número de restricciones, T en número de observaciones a usar y k el número de parámetros estimados en el modelo completo, entonces $T - k$ son los grados de libertad del modelo completo.

La hipótesis nula es que los datos fueron generados por el modelo reducido y la alternativa que fueron generados por el modelo completo. Si se cumple la hipótesis nula, $SCR(reducido)$ será cercano a $SCR(completo)$ y ϕ_i será pequeño; entonces, grandes valores de ϕ_i sigieren que el modelo adecuado es el completo (los errores del modelo reducido superan a los del modelo completo), por lo que se rechaza la hipótesis nula. En síntesis, si los valores del estadístico ϕ_i superan los valores críticos reportados por Dickey-Fuller, se rechaza la hipótesis nula y se descarta modelo reducido.

Extensión de las pruebas de Dickey-Fuller No todas las series pueden ser bien representadas por el proceso autorregresivo de primer orden: $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$. Se pueden usar las pruebas de Dickey-Fuller para ecuaciones de mayor orden que (B.1), (B.2) y (B.3). Se considera el proceso autorregresivo de orden p :

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p} + \varepsilon_t \quad (B.5)$$

Si a esta ecuación se le suma y resta $a_p y_{t-p+1}$ se obtiene:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - a_p \Delta y_{t-p+1} + \varepsilon_t \quad (B.6)$$

³ $SCR = \sum_{i=1}^T e_t^2$, donde e_t son los residuos del modelo.

Luego, si se le suma y resta $(a_{p-1} + a_p)y_{t-p+2}$ se obtiene:

$$y_t = a_0 + a_1y_{t-1} + a_2y_{t-2} + a_3y_{t-3} + \cdots - (a_{p-1} + a_p)\Delta y_{t-p+2} - a_p\Delta y_{t-p+1} + \varepsilon_t \quad (\text{B.7})$$

Si se continúa sumando y restando de esa forma se obtiene:

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (\text{B.8})$$

donde $\gamma = 1 - (\sum_{i=1}^p a_i)$, $\beta_i = \sum_{j=i}^p a_j$

En (B.8) el coeficiente de interés es γ ; si $\gamma = 0$ la ecuación es en primeras diferencias y entonces tiene raíz unitaria. Se puede testear la presencia de raíz unitaria usando los mismos estadísticos de Dickey-Fuller mencionados en la sección anterior.

Si $\gamma = 0$, entonces los coeficientes de la ecuación en diferencias suman uno y esto implica que al menos una de las raíces características es la unidad.

Para testear la presencia de múltiples raíces unitarias en el proceso, se pueden realizar las pruebas de Dickey-Fuller a la serie diferenciada d veces, donde d indica la cantidad de raíces unitarias encontradas hasta el momento.

Comentarios sobre las pruebas: Las pruebas de Dickey Fuller asumen que los errores son independientes y tienen varianza constante. Estos supuestos generan algunos problemas importantes relativos al hecho de que el verdadero proceso generador de los datos es desconocido. En primer lugar, el verdadero proceso generador de los datos puede contener tanto componentes autorregresivos como de medias móviles. Es necesario saber como realizar el test cuando el orden de los términos de medias móviles (si existe) es desconocido. En segundo lugar, no se puede estimar apropiadamente γ ni su desvío estándar si no se incluyen todos los términos autorregresivos en la ecuación, si bien el verdadero orden del proceso autorregresivo por lo general es desconocido. El problema es entonces determinar el valor apropiado de p .

El tercer problema consiste en que las pruebas de Dickey-Fuller se basan en una única raíz unitaria (si bien se resolvió el problema realizando sucesivas pruebas a la serie diferenciada según la cantidad de raíces unitarias encontradas).

Por último, dado que el proceso generador de los datos es desconocido, no es sencillo decidir cuando incluir constante y/o tendencia determinística, es decir cual de los modelos (B.1), (B.2) o (B.3) emplear.

En la sección siguiente se detalla un procedimiento recomendado para efectuar las pruebas de Dickey-Fuller cuando se desconoce el verdadero proceso generador de los datos.

Procedimiento para testear presencia de raíz unitaria Se parte del modelo más general,
 $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum \beta_i \Delta y_{t-i} + \varepsilon_t$

1. se testea $H_0) \gamma = 0$ con el modelo $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum \beta_i \Delta y_{t-i} + \varepsilon_t$
 - rechazo $H_0) \Rightarrow$ no hay evidencia para suponer raíz unitaria, pero
 - no rechazo $H_0) \Rightarrow$ testeo presencia de tendencia: $H_0) a_2 = \gamma = 0$
 - rechazo $H_0) \Rightarrow$ testeo $H_0) \gamma = 0$ usando distribución normal
 - rechazo $H_0) \Rightarrow$ concluyo que no hay raíz unitaria, pero
 - no rechazo $H_0) \Rightarrow$ concluyo que hay raíz unitaria, pero
 - no rechazo $H_0) \Rightarrow$ estimo modelo sin tendencia y con constante

2. Se testea $H_0) \gamma = 0$ con el modelo $\Delta y_t = a_0 + \gamma y_{t-1} + \sum \beta_i \Delta y_{t-i} + \varepsilon_t$
 - rechazo $H_0) \Rightarrow$ concluyo que no hay raíz unitaria, pero
 - no rechazo $H_0) \Rightarrow$ testeo presencia de constante: $H_0) a_0 = \gamma = 0$
 - rechazo $H_0) \Rightarrow$ testeo $H_0) \gamma = 0$ usando distribución normal
 - rechazo $H_0) \Rightarrow$ concluyo que no hay raíz unitaria, pero
 - no rechazo $H_0) \Rightarrow$ concluyo que hay raíz unitaria, pero
 - no rechazo $H_0) \Rightarrow$ estimo modelo sin tendencia ni constante

3. Se testea $H_0) \gamma = 0$ con el modelo $\Delta y_t = \gamma y_{t-1} + \sum \beta_i \Delta y_{t-i} + \varepsilon_t$
 - rechazo $H_0) \Rightarrow$ concluyo que no hay raíz unitaria, pero
 - no rechazo $H_0) \Rightarrow$ concluyo que hay raíz unitaria, pero

Apéndice C

Prueba de cointegración de Johansen

Se considera la siguiente generalización de n variables

$$x_t = A_1 x_{t-1} + \varepsilon_t$$

de manera que

$$\begin{aligned}\Delta x_t &= A_1 x_{t-1} - x_{t-1} + \varepsilon_t \\ &= (A_1 - I)x_{t-1} + \varepsilon_t \\ &= \pi x_{t-1} + \varepsilon_t\end{aligned}\tag{C.1}$$

donde

x_t y ε_t son vectores de dimensión $n \times 1$

A_1 es una matriz de parámetros de dimensión $n \times n$

I es una matriz identidad de dimensión $n \times n$

$$\pi = A_1 - I$$

El rango de $(A_1 - I)$ es igual al número de vectores de cointegración. Por analogía con el caso univariado, si $(A_1 - I)$ es una matriz de ceros, entonces el rango de π es cero, todos los $\{\Delta x_{it}\}$ son procesos de raíz unitaria. Como no hay combinación lineal de los procesos $\{x_{it}\}$ que sea estacionaria, las variables no son cointegradas. Si se excluye el caso en que las raíces características sean superiores a la unidad, si el rango de π es n , (C.1) representa un sistema convergente de ecuaciones en diferencias, por lo que las variables son estacionarias.

Si se considera una constante

$$\Delta x_t = A_0 \pi x_{t-1} + \varepsilon_t$$

donde $A_0 = (a_{01}, a_{02}, \dots, a_{0n})'$.

El efecto de incluir los a_{0i} es permitir una posible tendencia lineal en el proceso generador de los datos.

Al igual que con el test de Dickey-Fuller aumentado, el modelo multivariado también puede generalizarse para permitir un proceso autorregresivo de mayor orden. Se considera

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + \varepsilon_t\tag{C.2}$$

donde

$$x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$$

ε_t es un vector iid de dimensión n , con media cero y matriz de varianzas Σ_ε . Restando a ambos lados de (C.2) se obtiene

$$\Delta x_t = (A_1 - I)x_{t-1} + A_2x_{t-2} + A_3x_{t-3} + \dots + A_px_{t-p} + \varepsilon_t$$

luego, sumando y restando $(A_1 - I)x_{t-2}$

$$\Delta x_t = (A_1 - I)\Delta x_{t-1} + (A_2 + A_1 - I)x_{t-2} + A_3x_{t-3} + \dots + A_px_{t-p} + \varepsilon_t$$

después, sumando y restando $(A_2 + A_1 - I)x_{t-2}$

$$\Delta x_t = (A_1 - I)\Delta x_{t-1} + (A_2 + A_1 - I)\Delta x_{t-2} + (A_3 + A_2A_1 - I)x_{t-3} + \dots + A_px_{t-p} + \varepsilon_t$$

continuando se esta forma, se obtiene

$$\Delta x_t = \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \pi x_{t-p} + \varepsilon_t \quad (\text{C.3})$$

donde

$$\pi = -\left(I - \sum_{i=1}^p A_i\right)$$

$$\pi_i = -\left(I - \sum_{j=1}^i A_j\right)$$

La característica clave es que (C.3) es el rango de la matriz π ; el rango de la matriz π es igual al número de vectores de cointegración independientes. Si rango de π es cero, la matriz es nula y (C.3) es un VAR en primeras diferencias. Por otro lado, si π es de rango n , el proceso vectorial es estacionario. En casos intermedios, si rango de π es uno, existe un único vector de cointegración y la expresión πx_{t-p} es el factor de error de corrección. Para otros casos en donde el rango de π se encuentra entre 1 y n , existen múltiples vectores de cointegración.

El número de vectores de cointegración distintos puede obtenerse chequeando la significación de las raíces características de π . El rango de una matriz es igual al número de raíces características diferentes de cero. Se supone que se obtiene la matriz π y se ordenan las n raíces características tal que $\lambda_1 > \lambda_2 > \dots > \lambda_n$. Si las variables en x_t no están cointegradas, el rango de π es cero y todas las raíces características también valen cero. Como $\log(1) = 0$, cada una de las expresiones $\log(1 - \lambda_i)$ será cero si las variables no están cointegradas. De forma similar, si el rango de π es uno, $0 < \lambda_1 < 1$ tal que la primera expresión $\log(1 - \lambda_i)$ será negativa y todas las otras $\lambda_i = 0$ tal que $\log(1 - \lambda_2) = \log(1 - \lambda_3) = \dots = \log(1 - \lambda_n) = 0$.

En la práctica se pueden obtener únicamente estimadores de π y las raíces características. La prueba para el número de raíces características que son no significativamente diferentes de uno puede realizarse utilizando los siguientes estadísticos:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^n \log(1 - \hat{\lambda}_i) \quad (\text{C.4})$$

$$\lambda_{max}(r, r+1) = -T \log(1 - \hat{\lambda}_{r+1}) \quad (\text{C.5})$$

donde

$\hat{\lambda}_i$ son los valores estimados de las raíces características (valores propios) obtenidos a partir de la matriz π estimada

T el número de observaciones usables

Cuando los valores apropiados de r son claros, los estadísticos se denominan simplemente λ_{trace} y λ_{max} .

El primer estadístico prueba la hipótesis nula de que el número de vectores de cointegración distintos es menor o igual que r contra una alternativa general. λ_{trace} es igual a cero cuando todos los λ_i son iguales a cero. Cuando mayor que cero sean las raíces características estimadas, el más negativo es $\log(1 - \lambda_i)$ y el mayor es λ_{trace} . El segundo estadístico prueba la hipótesis nula de que el número de vectores de cointegración es r , contra la alternativa de $r + 1$ vectores de cointegración. Nuevamente, si el valor estimado de la raíz característica es cercano a cero, λ_{max} será pequeño.

Johansen y Juselius (1990) proveen los valores críticos para λ_{trace} y λ_{max} . la distribución de estos estadísticos depende de:

1. el número de componentes estacionarios bajo la hipótesis nula ($n-r$)
2. la forma del vector A_0

Apéndice D

Anexo de Código

```
#####  
##### ACIDO SULFURICO #####  
#####  
  
#####  
# 1: SERIE #  
#####  
  
## directorio de trabajo  
setwd("C:/PASANTIA/VENTAS/SERIES VENTAS")  
  
## bibliotecas  
library(tseries); library(TSA); library(urca); library(forecast)  
  
## funciones  
source("C:\\PASANTIA\\VENTAS\\log.test") # para testear conv de usar log  
source("C:\\PASANTIA\\VENTAS\\ analisis") # estima un modelo y devuelve el diagnos  
source("C:\\PASANTIA\\VENTAS\\DC") # estima regresoras comerciales  
nolaborables=read.table("nolaborables.txt") # regresora que modela feriados no lab  
turismo=read.table("turismo.txt") # regresora que modela turismo  
source("C:\\PASANTIA\\VENTAS\\efcal") # prueba la sign. de regres. del ef. calend  
source("C:\\PASANTIA\\VENTAS\\atipico.txt") # estima tipo y ubicacion del atipico  
source("C:\\PASANTIA\\VENTAS\\preds") # hace todo tipo de predicciones  
  
## serie  
acido = read.ts("acido.txt", start = c(1998, 1), frequency = 12)  
n = length(acido)  
  
## grafico  
axis = paste("01", c("98","99","00","01","02","03","04","05","06","07",  
"08","09","10"), sep = "/" )  
plot(acido, main = "Ventas mensuales de acido sulfurico - ISUSA", xlab =  
"mes/a~no", ylab = "toneladas", col = "lightskyblue4", axes = FALSE,  
frame.plot = TRUE); axis(1, 1998:2010, axis); axis(2)  
abline(v = 1998:2010, lty = 2, col = "grey")  
abline(h = seq(1000, 5000, 500),lty = 2, col = "grey")
```

```

plot(acido, xlab = "tiempo", ylab = "toneladas",
main = "ventas mensuales de acido sulfurico")
points(acido, pch = c("E","F","M","A","M","J","J","A","S","O","N","D"))

#####
# 2: ESTACIONARIEDAD y ESTACIONALIDAD #
#####

# series de los meses
acido_mes = matrix(acido, ncol = 12, byrow = T)
plot(acido_mes[,1], col = 1, cex = 1.5, xlab = "tiempo",
ylab = "toneladas", ylim = c(800, 3500), type = 'l')
for(i in 2:12){ lines(acido_mes[,i], col = i) }
legend(1997, 3500, c('ene','feb','mar','abr','may','jun','jul','ago',
'set','oct','nov','dic'), text.col = 1:12, bty = "n")

## ACF y PACF serie
par(mfrow = c(2,1)); acf(acido, lag.max = 50, main =
"ACF serie de ventas de acido sulfurico - ISUSA",
axes = F, frame.plot = T, ylab = "autocorrelacion")
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)
pacf(acido, lag.max = 50, main =
"PACF serie de ventas de acido sulfurico - ISUSA",
axes = F, frame.plot = T, ylab = "autocorrelacion")
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)

## transformacion logaritmica ##

# prueba para saber si es necesaria la transformacion logaritmica
logtest = log.test(acido); logtest # no aplico transf log

## TESTS DE RAIZ UNITARIA DE DICKEY FULLER ##

# 1) modelo general:  $diff(y) = \phi_2 + \tau_3 * y.lag1 + \phi_3 * t + e_t$ 
test1_acido = ur.df(acido, type = "trend", lags = 1)
summary(test1_acido) # todos sign

par(mfrow = c(2, 1))
acf(test1_acido@res, lag.max = 50)
pacf(test1_acido@res, lag.max = 50)

par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test1_acido@res, type = "Ljung-Box",
fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1), main =
"p-valor test Ljung-Box - residuos test1_acido", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test1_acido@res, type = "Box-Pierce",
fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1), main =
"p-valor test Box-Pierce - residuos test1_acido", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")

```

```

# se cumple incorrelacion, no se capta totalmente la dinamica pero mejora
# si aumento el num de lags no capto mas la dinamica

resumen1_acido = cbind(t(test1_acido@teststat), test1_acido@cval)
resumen1_acido
# testeo al 1%: Ho) tau3=0, H1) tau3<0
# el estadistico es menor al valor critico al 5% => rechazo que tau3 = 0 => no hay RU

#####
# 3: IDENTIFICACION #
#####

## ACF y PACF acido
par(mfrow = c(2,1))
acf(acido, lag.max = 50, main = "ACF de la serie acido - ISUSA")
pacf(acido, lag.max = 50, main = "PACF de la serie acido - ISUSA")

# modelos posibles: AR(2), ARMA(1,1), ARMA(2,1), AR(1)
# con tend lineal y estacionalidad determ.

#####
# 4: ESTIMACION, INTERVENCION, VALIDACION Y DIAGNOSTICO #
#####

# estimacion con tendencia lineal y estacionalidad deterministica

m = matrix(0, ncol = 11, nrow = n); tend = seq(1, n)
for(i in 1:ncol(m)){ m[seq(i, n, 12), i] = 1}

ar2_tm = analisis(acido, 2,0,0,0,0,0, xreg = cbind(tend, m), plot = F)
arma11_tm = analisis(acido, 1,0,1,0,0,0, xreg = cbind(tend, m), plot = F)
arma21_tm = analisis(acido, 2,0,1,0,0,0, xreg = cbind(tend, m), plot = F)
ar1_tm = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m), plot = F)

ar2_tm$sign # ar2 no sign
arma11_tm$sign # ma1 no sign
arma21_tm$sign # ar1, ar2 y ma1 no sign, descarto
ar1_tm$sign

# meses sign: marzo, junio, julio, set
# unico modelo posible: ar1

par(mfrow = c(2, 1)); acf(ar1_tm$res, lag.max = 50); pacf(ar1_tm$res, lag.max = 50)
# no se capta totalmente la dinamica

##### estimacion con efecto calendario #####

estreg.ar1_tm = efcac(acido, 1,0,0,0,0,0, xreg = cbind(tend, m))
estreg.ar1_tm
# errores (por nolab)

```

```

op = DC(n); op1 = op$op1; op2 = op$op2; op3 = op$op3; op4 = op$op4
op5 = op$op5; op6 = op$op6; op7 = op$op7
tur = turismo[1:n,]; nolab = nolaborables[1:n,]

ar1_t = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, tur), plot = F)
ar1_nl = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, nolab), plot = F)
ar1_op1 = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op1), plot = F)
ar1_op2 = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op2), plot = F)
ar1_op3 = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op3), plot = F)
ar1_op4 = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op4), plot = F)
ar1_op5 = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op5), plot = F)
ar1_op6 = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op6), plot = F)
ar1_op7 = analisis(acido , 1,0,0,0,0,0, xreg = cbind(tend, m, op7), plot = F)

ar1_t$sign # tur no sign
ar1_nl$sign # error
ar1_op1$sign # op1 sign
ar1_op2$sign # op2.2 no sign
ar1_op3$sign # op3 sign
ar1_op4$sign # op4.2 no sign
ar1_op5$sign # op5 sign, cte no sign
ar1_op6$sign # op6 no sign
ar1_op7$sign # op7 no sign

# posibles
ar1_op1$mod; ar1_op3$mod; ar1_op5$mod
# me quedo con el que ajuste mejor: op1

plot(ar1_op1$res); par(mfrow = c(2,1))
acf(ar1_op1$res, lag.max = 50); pacf(ar1_op1$res, lag.max = 50)

int.ar1 = detecta(acido, ar1_op1$mod)
int.ar1$significacion # detecta LS en jul 2002
ar1_int = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op1,
int.ar1$regresores), plot = "inc")

ar1_int$sign # meses sign: marzo, junio, julio, setiembre

ts.plot(ar1_op1$res, ar1_int$res, col = c(2, 4))
# ar1_int no pasa inc
# pruebo intervenir a mano
plot(ar1_op1$res)

atipicosA0 = c(71, 86, 87, 93, 99, 110, 117)
A0 = matrix(0, nrow = n, ncol = length(atipicosA0))
colnames(A0) = paste('A0', atipicosA0, sep = '')
for(i in 1:ncol(A0)){ A0[atipicosA0[i], i] = 1 }
A0 = as.data.frame(A0); attach(A0)

ar1_int0 = analisis(acido, 1,0,0,0,0,0, xreg = cbind(tend, m, op1, A087,
A093, A0110, A071, A0117, A086, A099),plot = "inc"); ar1_int0$sign

```

```

plot(ar1_int0$res, xlab = "tiempo", ylab = "toneladas"); ar1_int0$norm

ar1_int0$sign # meses sign: marzo, junio, julio, setiembre

plot(acido)
points(time(acido)[c(87, 93, 110, 71, 117, 86, 99)],
acido[c(87, 93, 110, 71, 117, 86, 99)], pch = 18, col = 2)

par(mfrow = c(2, 1))
acf(ar1_int0$res, lag.max = 50, main = "ACF residuos AR(1)")
pacf(ar1_int0$res, lag.max = 50, main = "PACF residuos AR(1)")

# hay muchos meses no sign y le falta modelizar parte estacional
# estimo modelos con tendencia lineal

par(mfrow = c(2, 1)); acf(acido, lag.max = 50); pacf(acido, lag.max = 50)

arima100100_t = analisis(acido, 1,0,0,1,0,0, xreg = tend, plot = "inc")
arima200100_t = analisis(acido, 2,0,0,1,0,0, xreg = tend, plot = "inc")
arima300100_t = analisis(acido, 3,0,0,1,0,0, xreg = tend, plot = "inc")
arima001100_t = analisis(acido, 0,0,1,1,0,0, xreg = tend, plot = "inc")
arima002100_t = analisis(acido, 0,0,2,1,0,0, xreg = tend, plot = "inc")
arima003100_t = analisis(acido, 0,0,3,1,0,0, xreg = tend, plot = "inc")
arima004100_t = analisis(acido, 0,0,4,1,0,0, xreg = tend, plot = "inc")
arima101100_t = analisis(acido, 1,0,1,1,0,0, xreg = tend, plot = "inc")
arima101101_t = analisis(acido, 1,0,1,1,0,1, xreg = tend, plot = "inc")
arima101001_t = analisis(acido, 1,0,1,0,0,1, xreg = tend, plot = "inc")

arima100100_t$sign
arima200100_t$sign # ar2 no sign
arima300100_t$sign # ar3 no sign
arima001100_t$sign
arima002100_t$sign
arima003100_t$sign # ma3 no sign
arima004100_t$sign # ma3 y ma4 no sign
arima101100_t$sign # ma1 no sign
arima101101_t$sign # ma1, sar1 y sma1 no sign
arima101001_t$sign # ma1 no sign

# posibles
plot(arima100100_t$res)
plot(arima001100_t$res)
plot(arima002100_t$res)

### reg comerciales

estreg.arima100100_t = efcac(acido, 1,0,0,1,0,0, xreg = tend)
estreg.arima001100_t = efcac(acido, 0,0,1,1,0,0, xreg = tend)
estreg.arima002100_t = efcac(acido, 0,0,2,1,0,0, xreg = tend)

estreg.arima100100_t # op1, op3

```

```

estreg.arima001100_t # op1, op3
estreg.arima002100_t # op1

arima100100_op1 = analisis(acido, 1,0,0,1,0,0, xreg = cbind(tend, op1),
plot = "inc"); arima100100_op1$sign
arima001100_op1 = analisis(acido, 0,0,1,1,0,0, xreg = cbind(tend, op1),
plot = "inc"); arima001100_op1$sign
arima002100_op1 = analisis(acido, 0,0,2,1,0,0, xreg = cbind(tend, op1),
plot = "inc"); arima002100_op1$sign

arima100100_op3 = analisis(acido, 1,0,0,1,0,0, xreg = cbind(tend, op3),
plot = "inc"); arima100100_op3$sign
arima001100_op3 = analisis(acido, 0,0,1,1,0,0, xreg = cbind(tend, op3),
plot = "inc"); arima001100_op3$sign

arima100100_op1$mod; arima100100_op3$mod # op1
arima001100_op1$mod; arima001100_op3$mod # op1

# deteccion automatica

int.arima100100 = detecta(acido, arima100100_op1$mod)
int.arima001100 = detecta(acido, arima001100_op1$mod)
int.arima002100 = detecta(acido, arima002100_op1$mod)

int.arima100100$significacion
int.arima001100$significacion
int.arima002100$significacion

arima100100_int = analisis(acido, 1,0,0,1,0,0, xreg = cbind(tend, op1,
int.arima100100$regresores), plot = "inc"); arima100100_int$sign # LS no sign
arima001100_int = analisis(acido, 0,0,1,1,0,0, xreg = cbind(tend, op1,
int.arima001100$regresores), plot = "inc"); arima001100_int$sign
arima002100_int = analisis(acido, 0,0,2,1,0,0, xreg = cbind(tend, op1,
int.arima002100$regresores), plot = "inc"); arima002100_int$sign # LS no sign

arima100100_int = analisis(acido, 1,0,0,1,0,0, xreg = cbind(tend, op1,
int.arima100100$regresores[,1]), plot = "inc"); arima100100_int$sign # LS no sign
arima002100_int = analisis(acido, 0,0,2,1,0,0, xreg = cbind(tend, op1,
int.arima002100$regresores[,1]), plot = "inc"); arima002100_int$sign # LS no sign

par(mfrow = c(2, 2))
plot(arima100100_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(1,0,0)(1,0,0)")
plot(arima001100_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,0,1)(1,0,0)")
plot(arima002100_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,0,2)(1,0,0)")

for(i in c(86, 93, 95, 99, 110, 117)){
  AO[, paste('AO', i, sep = '')] = 0; AO[i, paste('AO', i, sep = '')] = 1 }
AO = AO[, order(as.numeric(gsub('AO', '', names(AO))))]; attach(AO)

```

```

arima100100_int1 = analisis(acido, 1,0,0,1,0,0, xreg = cbind(op1, tend,
int.arima100100$regresores[,1], A0110, A0117, A093, A095), plot = "inc")
arima001100_int1 = analisis(acido, 0,0,1,1,0,0, xreg = cbind(op1, tend,
int.arima001100$regresores, A099, A0117, A095, A0110, A093), plot = "inc")
arima002100_int1 = analisis(acido, 0,0,2,1,0,0, xreg = cbind(op1, tend,
int.arima002100$regresores[,1], A086, A0117, A093, A095, A0110), plot = "inc")

arima100100_int1$sign
arima001100_int1$sign
arima002100_int1$sign

plot(arima100100_int1$res)
plot(arima001100_int1$res)
plot(arima002100_int1$res)

### posibles: arima100100_int1, arima001100_int1 y arima002100_int1

# residuos #

plot(arima100100_int1$res, xlab = "tiempo", ylab = "ton",
main = "ARIMA(1,0,0)(1,0,0)")
plot(arima001100_int1$res, xlab = "tiempo", ylab = "ton",
main = "ARIMA(0,0,1)(1,0,0)")
plot(arima002100_int1$res, xlab = "tiempo", ylab = "ton",
main = "ARIMA(0,0,2)(1,0,0)")

# incorrelacion #

arima100100_int1 = analisis(acido, 1,0,0,1,0,0, xreg = cbind(op1, tend,
int.arima100100$regresores[,1], A0110, A0117, A093, A095), plot = "inc")
arima001100_int1 = analisis(acido, 0,0,1,1,0,0, xreg = cbind(op1, tend,
int.arima001100$regresores, A099, A0117, A095, A0110, A093), plot = "inc")
arima002100_int1 = analisis(acido, 0,0,2,1,0,0, xreg = cbind(op1, tend,
int.arima002100$regresores[,1], A086, A0117, A093, A095, A0110), plot = "inc")
# descarto arima001100

# homoscedasticidad #

arima100100_int1 = analisis(acido, 1,0,0,1,0,0, xreg = cbind(op1, tend,
int.arima100100$regresores[,1], A0110, A0117, A093, A095), plot = "hom")
arima002100_int1 = analisis(acido, 0,0,2,1,0,0, xreg = cbind(op1, tend,
int.arima002100$regresores[,1], A086, A0117, A093, A095, A0110), plot = "hom")

# media cero #

med = mean(arima100100_int1$res)
sd = sqrt(sum((arima100100_int1$res - med)^2)/(156-3))
t = sqrt(156)*med/sd; pv.arima100100_int1 = 1 - pnorm(abs(t))
pv.arima100100_int1

```

```

med = mean(arima002100_int1$res)
sd = sqrt(sum((arima002100_int1$res - med)^2)/(156-3))
t = sqrt(156)*med/sd; pv.arima002100_int1 = 1 - pnorm(abs(t))
pv.arima002100_int1
# pasan

# normalidad #

arima100100_int1$norm
arima002100_int1$norm
# pasan

# AICc #

arima100100_int1$mod # 2
arima002100_int1$mod # 1

# QQ plots #

qqnorm(arima100100_int1$res); qqline(arima100100_int1$res)
qqnorm(arima002100_int1$res); qqline(arima002100_int1$res)

# histograma de los residuos #

par(mfrow = c(1, 2))

hist(arima100100_int1$res, freq = FALSE, col = "lavender",
main = "Residuos", xlab = "")
lines(density(arima100100_int1$res), col = "red")

hist(arima002100_int1$res, freq = FALSE, col = "lavender",
main = "Residuos", xlab = "")
lines(density(arima002100_int1$res), col = "red")

#####
# 5: PREDICCIÓN DENTRO DE LA MUESTRA #
#####

# p es la cantidad de pasos
# m es el mes de 2010 en que se corta la serie

## predicciones a varios pasos
pred100100 = pred002100 = list()
for(p in c(2:4, 6, 12)){
  pred100100[[as.character(p)]] = list()
  pred002100[[as.character(p)]] = list()
  for(m in seq(p, 12, p)){
    aux = window(acido, start = c(1998, 1), end = c(2010, m))
    pred100100[[as.character(p)]] [[as.character(m)]] = preds(aux,
1,0,0,1,0,0, xreg = cbind(op1, tend, int.arima100100$regresores[,1],
A0110, A0117, A093, A095)[1:length(aux),], pasos = p)

```



```

    pred002100[[as.character(p)]] [[as.character(m)]] = preds(aux,
    0,0,2,1,0,0, xreg = cbind(op1, tend, int.arima002100$regresores[,1],
    A086, A0117, A093, A095, A0110)[1:length(aux),], pasos = p)
  }
}

## predicciones varias veces a un paso
pred100100_2 = pred002100_2 = list()
for(p in c(6, 12)){
  pred100100_2[[as.character(p)]] = list()
  pred002100_2[[as.character(p)]] = list()
  for(m in seq(p, 12, p)){
    aux = window(acido, start = c(1998, 1), end = c(2010, m))
    pred100100_2[[as.character(p)]] [[as.character(m)]] = preds(aux,
    1,0,0,1,0,0, xreg = cbind(op1, tend, int.arima100100$regresores[,1],
    A0110, A0117, A093, A095)[1:length(aux),], pasos = 1, cuantos = p)
    pred002100_2[[as.character(p)]] [[as.character(m)]] = preds(aux,
    2,0,0,1,0,0, xreg = cbind(op1, tend, int.arima002100$regresores[,1],
    A086, A0117, A093, A095, A0110)[1:length(aux),], pasos = 1, cuantos = p)
  }
}

#####
##### SULFATO DE ALUMINIO #####
#####

#####
# 1: SERIE #
#####

## directorio de trabajo
setwd("C:/PASANTIA/VENTAS/INFORME FINAL")

## bibliotecas
library(tseries); library(TSA); library(urca); library(forecast)

## codigo externo
source("C:\\PASANTIA\\VENTAS\\log.test") # para testear conv de usar log
source("C:\\PASANTIA\\VENTAS\\ analisis") # estima un modelo y devuelve el diagnos
source("C:\\PASANTIA\\VENTAS\\DC") # estima regresoras comerciales
nolaborables=read.table("nolaborables.txt") # regresora que modela feriados no lab
turismo=read.table("turismo.txt") # regresora que modela turismo
source("C:\\PASANTIA\\VENTAS\\efcal") # prueba la sign. de regres. del ef. calend
source("C:\\PASANTIA\\VENTAS\\atipico.txt") # estima tipo y ubicacion del atipico
source("C:\\PASANTIA\\VENTAS\\preds") # hace todo tipo de predicciones

## serie
sulfa = read.ts("scripts/series/sulfa.txt", start = c(1998, 1), frequency = 12)
n = length(sulfa)

## grafico

```

```

axis = paste("01",c(98:99, paste(0, 0:9, sep = ""), 10), sep = "/" )
plot(sulfa, main = "Ventas de sulfato de aluminio - ISUSA", xlab = "mes/a\~no",
ylab = "toneladas", col = "bisque4", axes = FALSE, frame.plot = TRUE)
axis(1,1998:2010, axis); axis(2)
abline(v = 1998:2010, lty = 2, col = "grey")
abline(h = seq(200, 1400, 200), lty = 2, col = "grey")

plot(sulfa, xlab = "tiempo", ylab = "toneladas", main =
"ventas mensuales de sulfato de aluminio")
points(sulfa, pch = c("E","F","M","A","M","J","J","A","S","O","N","D"))

#####
# 2: ESTACIONARIEDAD y ESTACIONALIDAD #
#####

# series de los meses
mes = c("ene", "feb", "mar", "abr", "may", "jun", "jul", "ago","set", "oct",
"nov", "dic"); sulfa_mes = matrix(sulfa, ncol = 12, byrow = TRUE)

sulfa_meses = list()
for(i in 1:12){
  sulfa_meses[[i]] = ts(sulfa[seq(i, n, 12)], start = 1998,frequency = 1)}
  ts.plot(do.call(cbind, sulfa_meses), col = 1:12, lwd = 2, xlab = "tiempo",
  ylab = "toneladas", ylim = c(400,1600))
  legend("top", mes, fill = 1:12, bty = "n", horiz = TRUE)
}

## ACF y PACF serie
par(mfrow = c(2, 1))
acf(sulfa, lag.max = 50, main = "ACF", axes = F, frame.plot = T, xlab =
"rezago", ylab = "autocorrelacion")
axis(1, 0:4, as.character(seq(0, 48, 12))); axis(2)
pacf(sulfa, lag.max = 50, main = "PACF", axes = F, frame.plot = T, xlab =
"rezago", ylab = "autocorrelacion")
axis(1,0:4, as.character(seq(0, 48, 12))); axis(2)

## transformacion logaritmica ##

# prueba para saber si es necesaria la tranformacion logaritmica
logtest = log.test(sulfa, plot = T); logtest # no transformo

## TESTS DE RAIZ UNITARIA DICKEY FULLER ##

# la serie presenta tres atipicos importantes
seq(1,n)[sulfa > 1000] # los tres primeros
# imputo la media para quitar el efecto de esos atipicos en la prueba
sulfa1 = sulfa; sulfa1[c(29, 68, 130)] = mean(sulfa[-c(29, 68, 130)])
plot(sulfa1)

# 1) modelo general:  $\text{diff}(y) = \phi_2 + \tau_3 \cdot y_{\text{lag}1} + \phi_3 \cdot t + e_t$ 
test1_sulfa1 = ur.df(sulfa1, type = "trend", lags = 1)

```

```

par(mfrow = c(2, 1))
acf(test1_sulfa1@res, lag.max = 50)
pacf(test1_sulfa1@res, lag.max = 50)

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test1_sulfa1@res, type =
"Ljung-Box", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - residuos test1_sulfa1", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test1_sulfa1@res, type =
"Box-Pierce", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - residuos test1_sulfa1", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorrelacion (aunque no se capta totalmente la dinamica)

resumen1_sulfa1 = cbind(t(test1_sulfa1@teststat), test1_sulfa1@cval)
resumen1_sulfa1
# testeo al 1%: Ho) tau3=0, H1) tau3<0
# tau3 es mayor al valor critico al 1% => no rechazo Ho) => testeo tendencia
# phi3 es menor al vc => no rechazo ho) => tend no sign

# 2) modelo con constante: diff(y) = phi1 + tau2*y.lag1 + e_t
test2_sulfa1 = ur.df(sulfa1, type = "drift", lags = 1)

par(mfrow = c(2, 1))
acf(test2_sulfa1@res, lag.max = 50)
pacf(test2_sulfa1@res, lag.max = 50)

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test2_sulfa1@res, type =
"Ljung-Box", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - residuos test2_sulfa1", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test2_sulfa1@res, type =
"Box-Pierce", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - residuos test2_sulfa1", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple

resumen2_sulfa1 = cbind(t(test2_sulfa1@teststat), test2_sulfa1@cval)
resumen2_sulfa1
# testeo al 1%: Ho) tau2=0, H1) tau2<0
# estad tau2>vc => no rechazo Ho) => me fijo sign de cte
# estad phi1<vc => no rechazo Ho) => cte no sign

# 3) modelo sin nada: diff(y) = tau1*y.lag1 + e_t
test3_sulfa1 = ur.df(sulfa1, type = "none", lags = 6) # con estos lags se cumple incorr

```

```

par(mfrow = c(2, 1))
acf(test3_sulfa1@res, lag.max = 50)
pacf(test3_sulfa1@res, lag.max = 50)

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test3_sulfa1@res, type =
"Ljung-Box", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - residuos test3_sulfa1", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test3_sulfa1@res, type =
"Box-Pierce", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - residuos test3_sulfa1", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorr

resumen3_sulfa1 = cbind(t(test3_sulfa1@teststat), test3_sulfa1@cval)
resumen3_sulfa1
# testeo al 1%:, Ho) tau1=0, H1) tau1<0
# estad tau1>vc => no rechazo Ho) => hay RU

### testeo si posee otra RU

dsulfa1 = diff(sulfa1)
plot(dsulfa1)

# 1) modelo general:  $\text{diff}(y) = \phi_2 + \tau_3 * y.\text{lag}1 + \phi_3 * t + e_t$ 
test1_dsulfa1 = ur.df(dsulfa1, type = "trend", lags = 7) # con estos lags se cumple incorr

par(mfrow = c(2, 1))
acf(test1_dsulfa1@res, lag.max = 50)
pacf(test1_dsulfa1@res, lag.max = 50)

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test1_dsulfa1@res, type =
"Ljung-Box", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - residuos test1_sulfa", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 1:60){ i[j] = Box.test(test1_dsulfa1@res, type =
"Box-Pierce", fitdf = 1, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - residuos test1_sulfa", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")

resumen1_dsulfa1 = cbind(t(test1_dsulfa1@teststat), test1_dsulfa1@cval)
resumen1_dsulfa1
# tau3 < vc => rechazo Ho)
# phi3 > vc => rechazo H0)
# no hay RU en dsulfa => sulfa es I(1)

## grafico serie en diferencias

```

```

plot(dsulfa, main = "Ventas de sulfato de aluminio en diferencias - ISUSA",
xlab = "mes/a\~no", ylab = "toneladas", col = "bisque4", axes = FALSE,
frame.plot = TRUE); axis(1, 1998:2010, axis); axis(2)
abline(v = 1998:2010, lty = 2, col = "grey")
abline(h = seq(-500, 1000, 500), lty = 2, col = "grey")

## ACF y PACF serie en diferencias
par(mfrow = c(2, 1))
acf(dsulfa, lag.max = 50, main = "ACF", axes = F, frame.plot = T, xlab = "rezago",
ylab = "autocorrelacion"); axis(1, 0:4, as.character(seq(0, 48, 12))); axis(2)
pacf(dsulfa, lag.max = 50, main = "PACF", axes = F, frame.plot = T, xlab = "rezago",
ylab = "autocorrelacion"); axis(1,0:4, as.character(seq(0, 48, 12))); axis(2)

#####
# 3: IDENTIFICACION #
#####

# modelos posibles: ARIMA(0,1,1), ARIMA(2,1,1), ARIMA(4,1,1)

#####
# 4: ESTIMACION, INTERVENCION, VALIDACION Y DIAGNOSTICO #
#####

# estimacion inicial (sin ninguna regresora)

arima011 = analisis(sulfa, 0, 1, 1, plot = F)
arima211 = analisis(sulfa, 2, 1, 1, plot = F)
arma411 = analisis(sulfa, 4, 1, 1, plot = F)

arima011$sign
arima211$sign # ar2 no sign
arma411$sign # ar2, ar3 y ar4 no sign

arima311 = analisis(sulfa, 3, 1, 1, plot = F); arima311$sign # ar2 y ar3 no sign
arima211 = analisis(sulfa, 2, 1, 1, plot = F); arima211$sign # ar2 no sign

# otros
arima111 = analisis(sulfa, 1, 1, 1, plot = "inc"); arima111$sign
arima012 = analisis(sulfa, 0, 1, 2, plot = "inc"); arima012$sign
arima112 = analisis(sulfa, 1, 1, 2, plot = F); arima112$sign # ar1 y ma2 no sign

# estimacion con efecto calendario

estreg.arima011 = efcalsulfa(sulfa, 0, 1, 1); estreg.arima011
estreg.arima111 = efcalsulfa(sulfa, 1, 1, 1); estreg.arima111
estreg.arima012 = efcalsulfa(sulfa, 0, 1, 2); estreg.arima012

# no hay reg sign

# intervencion

```

```

# deteccion automatica
int.arima011 = detecta(sulfa, arima011$mod, maxiter = 5)
int.arima011$significacion # 2 AO y 1 IO
int.arima111 = detecta(sulfa, arima111$mod, maxiter = 5) # da error
int.arima012 = detecta(sulfa, arima012$mod, maxiter = 5)
int.arima012$significacion # 2 AO y 1 IO

arima011_int = analisis(sulfa, 0, 1, 1, xreg = int.arima011$regresores,
plot = "hom"); arima011_int$sign
arima111_int = analisis(sulfa, 1, 1, 1, xreg = int.arima011$regresores,
plot = "hom"); arima111_int$sign # le asigno las mismas reg que a arima011
arima012_int = analisis(sulfa, 0, 1, 2, xreg = int.arima012$regresores,
plot = "hom"); arima012_int$sign # p-valor ma2 grande

# intervencion 'manual'

atipicosAO = c(29, 39, 56, 68, 92, 130)
AO = matrix(0, nrow = n, ncol = length(atipicosAO))
colnames(AO) = paste('AO', atipicosAO, sep = '')
for(i in 1:ncol(AO)){ AO[atipicosAO[i], i] = 1 }
AO = as.data.frame(AO); attach(AO)

arima011_int = analisis(sulfa, 0,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056), plot = "hom"); arima011_int$sign
arima111_int = analisis(sulfa, 1,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056), plot = "hom"); arima111_int$sign
arima012_int = analisis(sulfa, 0,1,2, xreg = cbind(int.arima012$regresores,
A039, A092, A056), plot = "hom"); arima012_int$sign # ma2 no sign
arima011_int0 = analisis(sulfa, 0,1,1, xreg = cbind(A029, A068, A0130, A039,
A092, A056), plot = "hom"); arima011_int0$sign
arima111_int0 = analisis(sulfa, 1,1,1, xreg = cbind(A029, A068, A0130, A039,
A092, A056), plot = "hom"); arima111_int0$sign # ar1 no sign
arima012_int0 = analisis(sulfa, 0,1,2, xreg = cbind(A029, A068, A0130, A039,
A092, A056), plot = "hom"); arima012_int0$sign # ma2 no sign

# se cumple homosc y se capta la dinamica

# posibles
arima011_int$mod
arima111_int$mod
arima011_int0$mod

plot(arima011_int$res)
plot(arima111_int$res)
plot(arima011_int0$res)

plot(sulfa, xlab = "tiempo", ylab = "toneladas", main =
"ventas mensuales en plaza de sulfato de aluminio")
points(time(sulfa)[c(29, 39, 56, 68, 92, 130)],
sulfa[c(29, 39, 56, 68, 92, 130)], pch = 15, col = 2, cex = 1)

```

```

# pruebo significacion de las regresoras del efecto calendario

estreg.arima011_int = efcalsulfa(0,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056)); estreg.arima011_int # nada
estreg.arima111_int = efcalsulfa(1,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056)); estreg.arima111_int # nada
estreg.arima011_int0 = efcalsulfa(0,1,1, xreg = cbind(A029, A068, A0130, A039,
A092, A056)); estreg.arima011_int0 # nada
estreg.arima111_int0 = efcalsulfa(1,1,1, xreg = cbind(A029, A068, A0130, A039,
A092, A056)); estreg.arima111_int0 # nada
estreg.arima012_int0 = efcalsulfa(0,1,2, xreg = cbind(A029, A068, A0130, A039,
A092, A056)); estreg.arima012_int0 #

# me fijo si se capta la dinamica
par(mfrow = c(2, 1))
acf(arima011_int$res, lag.max = 50); pacf(arima011_int$res, lag.max = 50) # 10

par(mfrow= c(2, 1))
acf(arima111_int$res, lag.max = 50)
pacf(arima111_int$res, lag.max = 50) # 10, 38 y 46

par(mfrow = c(2, 1))
acf(arima011_int0$res, lag.max = 50, main = "ACF residuos ARIMA(0,1,1)", axes = F,
frame.plot = T, ylab = "autocorrelacion") # 10
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)
pacf(arima011_int0$res, lag.max = 50, main = "PACF residuos ARIMA(0,1,1)", axes = F,
frame.plot = T, ylab = "autocorrelacion") # 10
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)

arima011_int$sign
arima111_int$sign
arima011_int0$sign

##### verificacion de supuestos #####

# incorrelacion
arima011_int = analisis(sulfa, 0,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056), plot = "inc")
arima111_int = analisis(sulfa, 1,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056), plot = "inc")
arima011_int0 = analisis(sulfa, 0,1,1, xreg = cbind(A029, A068, A0130, A039,
A092, A056), plot = "inc")
# pasan

# homoscedasticidad
arima011_int = analisis(sulfa, 0,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056), plot = "hom")
arima111_int = analisis(sulfa, 1,1,1, xreg = cbind(int.arima011$regresores,
A039, A092, A056), plot = "hom")
arima011_int0 = analisis(sulfa, 0,1,1, xreg = cbind(A029, A068, A0130, A039,

```

```

A092, A056), plot = "hom")
# pasan

# prueba de media cero
m = mean(arima011_int$res); sd = sqrt(sum((arima011_int$res - m)^2)/(156-1))
t = sqrt(156)*m/sd; pv.arima011_int = 1 - pnorm(abs(t))

m = mean(arima111_int$res); sd = sqrt(sum((arima111_int$res - m)^2)/(156-2))
t = sqrt(156)*m/sd; pv.arima111_int = 1 - pnorm(abs(t))

m = mean(arima011_int0$res); sd = sqrt(sum((arima011_int0$res - m)^2)/(156-1))
t = sqrt(156)*m/sd; pv.arima011_int0 = 1 - pnorm(abs(t))

pv.arima011_int; pv.arima111_int; pv.arima011_int0
# pasan media cero

# normalidad:
arima011_int$norm # pasa
arima111_int$norm # pasa
arima011_int0$norm # pasa

par(mfrow = c(2, 1))
plot(arima011_int$res, xlab = "tiempo", ylab = "toneladas",
main = "residuos ARIMA(0,1,1)")
plot(arima111_int$res, xlab = "tiempo", ylab = "toneladas",
main = "residuos ARIMA(1,1,1)")
plot(arima011_int0$res, xlab = "tiempo", ylab = "toneladas",
main = "residuos ARIMA(0,1,1)")

arima011_int$mod
arima111_int$mod
arima011_int0$mod

# son muy parecidos

qqnorm(arima011_int$res); qqline(arima011_int$res)
qqnorm(arima111_int$res); qqline(arima111_int$res)
qqnorm(arima011_int0$res); qqline(arima011_int0$res)

# histograma de los residuos

par(mfrow = c(1, 2))

hist(arima011_int$res, freq = FALSE, col = "lavender",
main = "Residuos", xlab = "residuos", ylim = c(0,0.007))
lines(density(arima011_int$res), col = "red")

hist(arima111_int$res, freq = FALSE, col = "lavender",
main = "Residuos", xlab = "residuos", ylim = c(0,0.007))
lines(density(arima111_int$res), col = "red")

```



```

hist(arima011_int0$res,freq = FALSE, col = "lavender",
main = "Residuos", xlab = "residuos", ylim = c(0,0.006))
lines(density(arima011_int0$res), col = "red")

# pruebo los tres en prediccion

#####
# 6: PREDICCION DENTRO DE LA MUESTRA #
#####

# p es la cantidad de pasos
# m es el mes de 2010 en que se corta la serie

## predicciones a varios pasos
pred011int = pred111int = pred011int0 = list()
for(p in c(2:4, 6, 12)){
  pred011int[[as.character(p)]] = list()
  pred111int[[as.character(p)]] = list()
  pred011int0[[as.character(p)]] = list()
  for(m in seq(p, 12, p)){
    aux = window(sulfa, start = c(1998, 1), end = c(2010, m))
    pred011int[[as.character(p)]] [[as.character(m)]] = preds(aux,
0,1,1, xreg = cbind(int.arima011$regresores, A039, A092, A056)
[1:length(aux),], pasos = p)
    pred111int[[as.character(p)]] [[as.character(m)]] = preds(aux,
1,1,1, xreg = cbind(int.arima011$regresores, A039, A092, A056)
[1:length(aux),], pasos = p)
    pred011int0[[as.character(p)]] [[as.character(m)]] = preds(aux,
0,1,1, xreg = cbind(A029, A068, A0130, A039, A092, A056)
[1:length(aux),], pasos = p)
  }
}

## predicciones varias veces a un paso
pred011int_2 = pred111int_2 = pred011int0_2 = list()
for(p in c(6, 12)){
  pred011int_2[[as.character(p)]] = list()
  pred111int_2[[as.character(p)]] = list()
  pred011int0_2[[as.character(p)]] = list()
  for(m in seq(p, 12, p)){
    aux = window(sulfa, start = c(1998, 1), end = c(2010, m))
    pred011int_2[[as.character(p)]] [[as.character(m)]] = preds(aux,
0,1,1, xreg = cbind(int.arima011$regresores, A039, A092, A056)
[1:length(aux),], pasos = 1, cuantos = p)
    pred111int_2[[as.character(p)]] [[as.character(m)]] = preds(aux,
1,1,1, xreg = cbind(int.arima011$regresores, A039, A092, A056)
[1:length(aux),], pasos = 1, cuantos = p)
    pred011int0_2[[as.character(p)]] [[as.character(m)]] = preds(aux,
0,1,1, xreg = cbind(A029, A068, A0130, A039, A092, A056)
[1:length(aux),], pasos = 1, cuantos = p)
  }
}

```

```

}

#####
##### FERTILIZANTES #####
#####

#####
# 1: SERIE #
#####

## directorio de trabajo
setwd("C:/PASANTIA/VENTAS/SERIES VENTAS")

## bibliotecas
library(tseries); library(TSA); library(urca); library(forecast)

## funciones
source("C:\\PASANTIA\\VENTAS\\log.test") # para testear conv de usar log
source("C:\\PASANTIA\\VENTAS\\ analisis") # estima un modelo y devuelve el diagnos
source("C:\\PASANTIA\\VENTAS\\DC") # estima regresoras comerciales
nolaborables=read.table("nolaborables.txt") # regresora que modela feriados no lab
turismo=read.table("turismo.txt") # regresora que modela turismo
source("C:\\PASANTIA\\VENTAS\\efcal") # prueba la sign. de regres. del ef. calend
source("C:\\PASANTIA\\VENTAS\\atipico.txt") # estima tipo y ubicacion del atipico
source("C:\\PASANTIA\\VENTAS\\preds") # hace todo tipo de predicciones

## serie
fert = read.ts("fert.txt", start = c(1998,1), frequency = 12) # en toneladas
n = length(fert)

## grafico
axis = paste("01", c("98", "99", "00", "01", "02", "03", "04", "05", "06",
"07", "08", "09", "10"), sep = "/")
plot(fert, main = "Ventas de fertilizantes plaza - ISUSA", xlab = "mes/a\\~no",
ylab = "toneladas", col = "seagreen", axes = FALSE, frame.plot = TRUE)
axis(1, 1998:2010, axis); axis(2)
abline(v = 1998:2010, lty = 2, col = "azure3")
abline(h = seq(5000, 40000, 5000), lty = 2, col = "azure3")
abline(h = mean(fert), col = "red")

plot(fert, xlab = "tiempo", ylab = "toneladas",
main = "ventas mensuales de fertilizantes")
points(fert, pch = c("E", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D"))

#####
# 2: ESTACIONARIEDAD y ESTACIONALIDAD #
#####

## ACF y PACF de la serie
par(mfrow = c(2, 1))
acf(fert, lag.max = 50, main = "ACF serie de ventas de fertilizantes - ISUSA",

```

```

axes = F, frame.plot = T, ylab = "autocorrelacion")
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)
pacf(fert, lag.max = 50, main = "PACF serie de ventas de fertilizantes - ISUSA",
axes = F, frame.plot = T, ylab = "autocorrelacion")
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)

# se observa presencia de estacionalidad anual debido a las autocorrelaciones
# significativas en acf en los lags 12, 24 y 36 y el pacf en el lag 12
fert12 = diff(fert, 12); plot(fert12)

par(mfrow = c(2, 1))
acf(fert12, lag.max = 50); pacf(fert12, lag.max = 50)

# prueba para testear conveniencia de usar logaritmo
logtest1 = log.test(fert, plot = T); logtest1 # no hay que transformar la serie

## TESTS DE RAIZ UNITARIA DE DICKEY FULLER ##

#1) modelo general:  $\text{diff}(y) = \phi_2 + \tau_3 y_{\text{lag}1} + \phi_3 t + e_t$ 
test1_fert = ur.df(fert, type = "trend", lags = 12)
summary(test1_fert)

par(mfrow = c(2, 1))
acf(test1_fert@res, lag.max=50); pacf(test1_fert@res, lag.max=50)
# se capta la dinamica

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 12:60){ i[j] = Box.test(test1_fert@res, type =
"Ljung-Box", fitdf = 12, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - residuos test1_fert", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 12:60){ i[j] = Box.test(test1_fert@res, type = "Box-Pierce",
fitdf = 12, lag = j)$p.value }; plot(i, ylim = c(0,1), main =
"p-valor test Box-Pierce - residuos test1_fert", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorrelacion

resumen1_fert = cbind(t(test1_fert@teststat), test1_fert@cval)
resumen1_fert
# testeo:  $H_0) \tau_3=0, H_1) \tau_3<0$ 
# estadistico mayor a valor critico al 5% => no rechazo  $H_0) =>$  testeo sign de ten
# testeo al 1%:  $H_0) \phi_3=0, H_1) \phi_3<0$ 
# estadistico menor a valor critico => no rechazo  $H_0) =>$  tend no es sign

# 2) modelo con constante:  $\text{diff}(y) = \phi_1 + \tau_2 y_{\text{lag}1} + e_t$ 
test2_fert = ur.df(fert, type = "drift", lags = 12)
summary(test2_fert)

par(mfrow = c(2, 1))
acf(test2_fert@res, lag.max = 50)

```

```

pacf(test2_fert@res, lag.max = 50)
# se capta la dinamica

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 12:60){ i[j] = Box.test(test2_fert@res, type = "Ljung-Box",
fitdf = 12, lag = j)$p.value }; plot(i, ylim= c (0, 1), main =
"p-valor test Ljung-Box - residuos test2_fert", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 12:60){ i[j] = Box.test(test2_fert@res, type = "Box-Pierce",
fitdf = 12, lag = j)$p.value }; plot(i, ylim = c(0, 1), main =
"p-valor test Box-Pierce - residuos test2_fert", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorrelacion

resumen2_fert = cbind(t(test2_fert@teststat), test2_fert@cval)
resumen2_fert
# testeo al 1%: Ho) tau2=0, H1) tau2<0
# estadistico mayor a valor critico al 5% => no rechazo Ho) => testeo sign de cte

# testeo: Ho) phi1=0, H1) phi1<0
# estadistico menor a valor critico al 5% => no rechazo Ho) => cte no es sign

# 3) modelo sin nada: diff(y) = tau1*y.lag1 + e_t
test3_fert = ur.df(fert, type = "none", lags = 12)
summary(test3_fert)
plot(test3_fert)

par(mfrow = c(2, 1))
acf(test3_fert@res, lag.max = 50)
pacf(test3_fert@res, lag.max = 50) # lag 35 sign

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0,60); for(j in 12:60){i[j] = Box.test(test3_fert@res, type = "Ljung-Box",
fitdf = 12, lag = j)$p.value }; plot(i, ylim = c(0, 1), main =
"p-valor test Ljung-Box - residuos test3_fert", xlab = "lag",ylab="p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 12:60){ i[j] = Box.test(test3_fert@res, type =
"Box-Pierce", fitdf = 12, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - residuos test3_fert", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorrelacion

resumen3_fert = cbind(t(test3_fert@teststat), test3_fert@cval)
resumen3_fert
# testeo al 1%: Ho) tau1=0, H1) tau1<0
# estadistico mayor a valor critico al 5% => no rechazo Ho) => existe RU

##### investigo si puede ser I(2)
dfert = diff(fert)

```

```

test1_dfert = ur.df(dfert, type = "trend", lags = 12)
summary(test1_dfert)

par(mfrow = c(2, 1))
acf(test1_dfert@res, lag.max = 50)
pacf(test1_dfert@res, lag.max = 50) # lag 35 sign
# se capta totalmente la dinamica

# incorrelacion
par(mfrow = c(2, 1))
i = rep(0, 60); for(j in 12:60){ i[j] = Box.test(test1_dfert@res, type = "Ljung-Box",
fitdf = 12, lag = j)$p.value }; plot(i, ylim = c(0, 1), main =
"p-valor test Ljung-Box - residuos test1_dfert", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
i = rep(0, 60); for(j in 12:60){ i[j] = Box.test(test1_dfert@res, type =
"Box-Pierce", fitdf = 12, lag = j)$p.value }; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - residuos test1_dfert", xlab = "lag", ylab =
"p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorrelacion

resumen1_dfert = cbind(t(test1_dfert@teststat), test1_dfert@cval)
resumen1_dfert
# estad tau3 menor a vc al 1% => rechazo Ho)
# estad phi3 mayor a vc al 1% => rechazo Ho)
# FERT ES I(1)

## grafico dfert
plot(dfert, main = "Ventas de fertilizantes plaza en diferencias - ISUSA", xlab =
"mes/a~no", ylab = "toneladas", col = "seagreen", axes = FALSE, frame.plot = TRUE)
axis(1, 1998:2010, axis); axis(2); abline(v = 1998:2010, lty = 2, col = "azure3")
abline(h = seq(-20000, 20000, 10000), lty = 2, col = "azure3")
abline(h = mean(dfert), col = "red")

plot(dfert, xlab = "tiempo", ylab = "toneladas", main =
"ventas mensuales de fertilizantes en diferencias")
points(dfert, pch = c("E", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D"))

par(mfrow = c(2, 1))
acf(dfert, lag.max = 50); pacf(dfert, lag.max = 50)

# puede haber RU en parte estacional
d12dfert = diff(dfert, 12); plot(d12dfert)

# serie a modelizar: d12dfert

logtest.d12dfert = log.test(d12dfert, plot = T)
logtest.d12dfert # no hay que trabajar en log

par(mfrow = c(2, 1))
acf(d12dfert, lag.max = 50, main =
"ACF serie de ventas de fertilizantes en diferencia regular y estacional - ISUSA",

```

```

axes = F, frame.plot = T, ylab = "autocorrelacion")
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)
pacf(d12dfert, lag.max = 50, main =
"PACF serie de ventas de fertilizantes en diferencia regular y estacional - ISUSA",
axes = F, frame.plot = T, ylab = "autocorrelacion")
axis(1, c(0, 1, 2, 3, 4), as.character(c(0, 12, 24, 36, 48))); axis(2)

# los sugeridos por el analisis de ACF y PACF d12dfert: ARIMA(0,1,1)(0,1,1),
# ARIMA(0,1,1)(1,1,0), ARIMA(0,1,1)(1,1,1), ARIMA(3,1,1)(0,1,1),
# ARIMA(3,1,1)(1,1,0), ARIMA(3,1,1)(1,1,1)

#####
# 4: ESTIMACION, INTERVENCION, VALIDACION Y DIAGNOSTICO #
#####

##### estimacion inicial (sin ninguna regresora) #####

arima011011 = analisis(fert, 0,1,1,0,1,1, plot = "inc")
arima011110 = analisis(fert, 0,1,1,1,1,0, plot = "inc")
arima011111 = analisis(fert, 0,1,1,1,1,1, plot = "inc")
arima012011 = analisis(fert, 0,1,2,0,1,1, plot = "inc")
arima012110 = analisis(fert, 0,1,2,1,1,0, plot = "inc")
arima012111 = analisis(fert, 0,1,2,1,1,1, plot = "inc")
arima112011 = analisis(fert, 1,1,2,0,1,1, plot = "inc")
arima112110 = analisis(fert, 1,1,2,1,1,0, plot = "inc")
arima112111 = analisis(fert, 1,1,2,1,1,1, plot = "inc")

arima011011$sign
arima011110$sign
arima011111$sign
arima012011$sign
arima012110$sign
arima012111$sign # ma2 y sar1 no sign (ya esta estimado)
arima112011$sign # ma2 no sign
arima112110$sign # ma2 no sign
arima112111$sign # ma2 y sar1 no sign

arima111011 = analisis(fert, 1,1,1,0,1,1, plot = "inc")
arima111110 = analisis(fert, 1,1,1,1,1,0, plot = "inc")
arima111111 = analisis(fert, 1,1,1,1,1,1, plot = "hom")

arima111011$sign
arima111110$sign
arima111111$sign # sar1 no sign

arima311011 = analisis(fert, 3,1,1,0,1,1, plot = "inc")
arima311110 = analisis(fert, 3,1,1,1,1,0, plot = "inc")
arima311111 = analisis(fert, 3,1,1,1,1,1, plot = "inc")

arima311011$sign # ma1 no sign
arima311110$sign # ma1 no sign

```

```

arima311111$sign # ma1 y sar1 no sign

arima310011 = analisis(fert, 3,1,0,0,1,1, plot = "hom")
arima310110 = analisis(fert, 3,1,0,1,1,0, plot = "hom")
arima310111 = analisis(fert, 3,1,0,1,1,1, plot = "inc")

arima310011$sign
arima310110$sign
arima310111$sign # sar1 no sign, ya estimado

plot(arima011011$res)
plot(arima011110$res)
plot(arima011111$res)
plot(arima012011$res)
plot(arima012110$res)
plot(arima111011$res)
plot(arima111110$res)
plot(arima310011$res)
plot(arima310110$res)

# deteccion automatica

int.arima011011 = detecta(fert, arima011011$mod)
int.arima011110 = detecta(fert, arima011110$mod)
int.arima011111 = detecta(fert, arima011111$mod, methodCSS=T)
int.arima012011 = detecta(fert, arima012011$mod)
int.arima012110 = detecta(fert, arima012110$mod)
int.arima111011 = detecta(fert, arima111011$mod, methodCSS=T)
int.arima111110 = detecta(fert, arima111110$mod, methodCSS = T)
int.arima310011 = detecta(fert, arima310011$mod)
int.arima310110 = detecta(fert, arima310110$mod)

int.arima011011$significacion
int.arima011110$significacion
int.arima011111$significacion
int.arima012011$significacion
int.arima012110$significacion
int.arima111011$significacion
int.arima111110$significacion
int.arima310011$significacion
int.arima310110$significacion

# intervencion

arima011011_int = analisis(fert, 0,1,1,0,1,1,
xreg = int.arima011011$regresores, plot = "inc")
arima011110_int = analisis(fert, 0,1,1,1,1,0,
xreg = int.arima011110$regresores, plot = "inc")
arima011111_int = analisis(fert, 0,1,1,1,1,1,
xreg = int.arima011111$regresores, plot = "inc")
arima012011_int = analisis(fert, 0,1,2,0,1,1,

```

```

xreg = int.arima012011$regresores, plot = "hom")
arima012110_int = analisis(fert, 0,1,2,1,1,0,
xreg = int.arima012110$regresores, plot = "hom")
arima111011_int = analisis(fert, 1,1,1,0,1,1,
xreg = int.arima111011$regresores, plot = "inc")
arima111110_int = analisis(fert, 1,1,1,1,1,0,
xreg = int.arima111110$regresores, plot = "inc")
arima310011_int = analisis(fert, 3,1,0,0,1,1,
xreg = int.arima310011$regresores, plot = "inc")
arima310110_int = analisis(fert, 3,1,0,1,1,0,
xreg = int.arima310110$regresores, plot = "hom")

arima011011_int$sign
arima011110_int$sign
arima011111_int$sign
arima012011_int$sign # pasa inco
arima012110_int$sign # pasa inco
arima111011_int$sign
arima111110_int$sign
arima310011_int$sign
arima310110_int$sign # pasa inco
# todos sign

# residuos
par(mfrow = c(2, 2))
plot(arima011011_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,1)(0,1,1)")
plot(arima011110_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,1)(1,1,0)")
plot(arima011111_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,1)(1,1,1)")
plot(arima012011_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,2)(0,1,1)")
plot(arima012110_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,2)(1,1,0)")
plot(arima111011_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(1,1,1)(0,1,1)")
plot(arima111110_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(1,1,1)(1,1,0)")
plot(arima310011_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(3,1,0)(0,1,1)")
plot(arima310110_int$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(3,1,0)(1,1,0)")

# faltan intervenciones
plot(arima011011_int$res)
plot(arima011110_int$res)
plot(arima011111_int$res)
plot(arima012011_int$res)
plot(arima012110_int$res)
plot(arima111011_int$res)

```



```

plot(arima111110_int$res)
plot(arima310011_int$res)
plot(arima310110_int$res)

atipicosAO = c(19, 90, 122, 131, 133, 135, 138, 141, 152, 154)
AO = matrix(0, nrow = n, ncol = length(atipicosAO))
colnames(AO) = paste('AO', atipicosAO, sep = '')
for(i in 1:ncol(AO)){ AO[atipicosAO[i], i] = 1 }
AO = as.data.frame(AO); attach(AO)
TC152 = TC(152, n)
TC122=TC(122, n)

arima011011_int1 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, AO138, AO152, AO90, AO154), plot = "inc"); arima011011_int1$sign
arima011110_int1 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, AO138, AO141, AO19, AO90, TC152), plot = "inc"); arima011110_int1$sign
arima011111_int1 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, AO138, AO90, TC152), plot = "inc"); arima011111_int1$sign
arima012011_int1 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, AO138, AO141, AO152, AO19, AO90), plot = "inc")
arima012110_int1 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, AO138, AO141, AO152, AO19, AO90), plot = "inc")
arima111011_int1 = analisis(fert, 1,1,1,0,1,1, xreg = cbind(int.arima111011$
regresores, AO138, AO141, AO90, AO19, AO133, AO135), plot = "inc")
arima111011_int1$sign # ar1 no sign DESCARTADO
arima111110_int1 = analisis(fert, 1,1,1,1,1,0, xreg = cbind(int.arima111110$
regresores, AO138, AO19, AO90, TC152), plot = "inc")
arima111110_int1$sign # ar1 no sign DESCARTADO
arima310011_int1 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, AO138, AO141, AO152, AO90), plot = "inc"); arima310011_int1$sign
IO19_310011 = IO(arima310011_int1$mod, 19, n)
arima310011_int1 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, AO138, AO141, AO152, AO90, IO19_310011), plot = "inc")
arima310011_int1$sign
arima310110_int1 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, AO138, AO152, AO90), plot = "inc"); arima310110_int1$sign
IO19_310110 = IO(arima310110_int1$mod, 19, n)
arima310110_int1 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, AO138, AO152, AO90, IO19_310110), plot = "inc"); arima310110_int1$sign
IO131_310110 = IO(arima310110_int1$mod, 131, n)
arima310110_int1 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, AO138, AO152, AO90, IO19_310110, IO131_310110), plot = "inc")
arima310110_int1$sign

# posibles: arima011011_int1, arima011110_int1, arima011111_int1, arima012011_int1
# arima012110_int1, arima310011_int1, arima310110_int1

# residuos
par(mfrow = c(2, 2))
plot(arima011011_int1$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,1)(0,1,1)")

```

```

plot(arima011110_int1$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,1)(1,1,0)")
plot(arima011111_int1$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,1)(1,1,1)")
plot(arima012011_int1$res, xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,2)(0,1,1)")
plot(arima012110_int1$res,xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(0,1,2)(1,1,0)")
plot(arima310011_int1$res,xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(3,1,0)(0,1,1)")
plot(arima310110_int1$res,xlab = "tiempo", ylab = "toneladas",
main = "ARIMA(3,1,0)(1,1,0)")

# pruebo efecto calendario

op = DC(n); op1 = op$op1; op2 = op$op2; op3 = op$op3; op4 = op$op4
op5 = op$op5; op6 = op$op6; op7 = op$op7
tur = turismo[1:n,]; nolab = nolaborables[1:n,]

arima011011_tur = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, tur), plot = F)
arima011011_nolab = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, nolab), plot = F)
arima011011_op1 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, op1), plot = F)
arima011011_op2 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, op2), plot = F)
arima011011_op3 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, op3), plot = F)
arima011011_op4 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, op4), plot = F)
arima011011_op5 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, op5), plot = F)
arima011011_op6 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, op6), plot = F)
arima011011_op7 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154, op7), plot = F)

arima011011_tur$sign # tur no sign
arima011011_nolab$sign # nolab NAN
arima011011_op1$sign # op1 no sign
arima011011_op2$sign # op2 no sign
arima011011_op3$sign # op3 no sign
arima011011_op4$sign # op4 no sign
arima011011_op5$sign # op5 no sign
arima011011_op6$sign # op6 no sign
arima011011_op7$sign # op7 no sign

arima011110_tur = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, tur), plot = F)
arima011110_nolab = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$

```

```

regresores, A0138, A0141, A019, A090, TC152, nolab), plot = F)
arima011110_op1 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, op1),plot = F)
arima011110_op2 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, op2), plot = F)
arima011110_op3 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, op3), plot = F)
arima011110_op4 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, op4), plot = F)
arima011110_op5 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, op5), plot = F)
arima011110_op6 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, op6), plot = F)
arima011110_op7 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152, op7), plot = F)

```

```

arima011110_tur$sign # tur no sign
arima011110_nolab$sign # nolab NAN
arima011110_op1$sign # op1 no sign
arima011110_op2$sign # op2 no sign
arima011110_op3$sign # op3 no sign
arima011110_op4$sign # op4 no sign
arima011110_op5$sign # op5 no sign
arima011110_op6$sign # op6 no sign
arima011110_op7$sign # op7 no sign

```

```

arima011111_tur = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, tur), plot = F)
arima011111_nolab = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, nolab), plot = F)
arima011111_op1 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, op1), plot = F)
arima011111_op2 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, op2), plot = F)
arima011111_op3 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, op3), plot = F)
arima011111_op4 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, op4), plot = F)
arima011111_op5 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, op5), plot = F)
arima011111_op6 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, op6), plot = F)
arima011111_op7 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152, op7), plot = F)

```

```

arima011111_tur$sign # tur no sign
arima011111_nolab$sign # nolab NAN
arima011111_op1$sign # op1 no sign
arima011111_op2$sign # op2 no sign
arima011111_op3$sign #op3 no sign
arima011111_op4$sign # op4 no sign

```

```
arima011111_op5$sign # op5 no sign
arima011111_op6$sign # op6 no sign
arima011111_op7$sign # op7 no sign
```

```
arima012011_tur = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, tur), plot = F)
arima012011_nolab = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, nolab), plot = F)
arima012011_op1 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, op1), plot = F)
arima012011_op2 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, op2), plot = F)
arima012011_op3 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, op3), plot = F)
arima012011_op4 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, op4), plot = F)
arima012011_op5 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, op5), plot = F)
arima012011_op6 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, op6), plot = F)
arima012011_op7 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090, op7), plot = F)
```

```
arima012011_tur$sign # tur no sign
arima012011_nolab$sign # nolab NAN
arima012011_op1$sign # op1 no sign
arima012011_op2$sign # op2 no sign
arima012011_op3$sign # op3 no sign
arima012011_op4$sign # op4 no sign
arima012011_op5$sign # op5 no sign
arima012011_op6$sign # op6 no sign
arima012011_op7$sign # op7 no sign
```

```
arima012110_tur = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, tur), plot = F)
arima012110_nolab = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, nolab), plot = F)
arima012110_op1 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, op1), plot = F)
arima012110_op2 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, op2), plot = F)
arima012110_op3 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, op3), plot = F)
arima012110_op4 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, op4), plot = F)
arima012110_op5 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, op5), plot = F)
arima012110_op6 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, op6), plot = F)
arima012110_op7 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090, op7), plot = F)
```

```
arima012110_tur$sign # tur no sign
arima012110_nolab$sign # nolab no sign
arima012110_op1$sign # op1 no sign
arima012110_op2$sign # op2 no sign
arima012110_op3$sign # op3 no sign
arima012110_op4$sign # op4 no sign
arima012110_op5$sign # op5 no sign
arima012110_op6$sign # op6 no sign
arima012110_op7$sign # op7 no sign
```

```
arima310011_tur = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, tur), plot = F)
arima310011_nolab = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, nolab), plot = F)
arima310011_op1 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, op1), plot = F)
arima310011_op2 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, op2), plot = F)
arima310011_op3 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, op3), plot = F)
arima310011_op4 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, op4), plot = F)
arima310011_op5 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, op5), plot = F)
arima310011_op6 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, op6), plot = F)
arima310011_op7 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011, op7), plot = F)
```

```
arima310011_tur$sign # tur no sign
arima310011_nolab$sign # nolab no sign
arima310011_op1$sign # op1 no sign
arima310011_op2$sign # op2 no sign
arima310011_op3$sign # op3 no sign
arima310011_op4$sign # op4 no sign
arima310011_op5$sign # op5 no sign
arima310011_op6$sign # op6 no sign
arima310011_op7$sign # op7 no sign
```

```
arima310110_tur = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, tur), plot = F)
arima310110_nolab = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, nolab), plot = F)
arima310110_op1 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, op1), plot = F)
arima310110_op2 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, op2), plot = F)
arima310110_op3 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, op3), plot = F)
arima310110_op4 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
```

```

regresores, A0138, A0152, A090, I019_310110, I0131_310110, op4), plot = F)
arima310110_op5 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, op5), plot = F)
arima310110_op6 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, op6), plot = F)
arima310110_op7 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110, op7), plot = F)

arima310110_tur$sign # tur no sign
arima310110_nolab$sign # nolab NAN
arima310110_op1$sign # op1 no sign
arima310110_op2$sign # op2 no sign
arima310110_op3$sign # op3 no sign
arima310110_op4$sign # op4 no sign
arima310110_op5$sign # op5 no sign
arima310110_op6$sign # op6 no sign
arima310110_op7$sign # op7 no sign

# incorrelacion #

arima011011_int1 = analisis(fert, 0,1,1,0,1,1, xreg = cbind(int.arima011011$
regresores, A0138, A0152, A090, A0154), plot = "inc")
arima011110_int1 = analisis(fert, 0,1,1,1,1,0, xreg = cbind(int.arima011110$
regresores, A0138, A0141, A019, A090, TC152), plot = "inc")
arima011111_int1 = analisis(fert, 0,1,1,1,1,1, xreg = cbind(int.arima011111$
regresores, A0138, A090, TC152), plot = "inc")
arima012011_int1 = analisis(fert, 0,1,2,0,1,1, xreg = cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090), plot = "inc") # pasa
arima012110_int1 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090), plot = "inc") # pasa
arima310011_int1 = analisis(fert, 3,1,0,0,1,1, xreg = cbind(int.arima310011$
regresores, A0138, A0141, A0152, A090, I019_310011), plot = "inc")
arima310110_int1 = analisis(fert, 3,1,0,1,1,0, xreg = cbind(int.arima310110$
regresores, A0138, A0152, A090, I019_310110, I0131_310110), plot = "inc")
# pasan arima012011_int1 y arima012110_int1

# media cero #
med = mean(arima012011_int1$res)
sd = sqrt(sum((arima012011_int1$res - med)^2)/(156-12-1-3))
t = sqrt(156-12-1)*med/sd; pv.arima012011_int1 = (1 - pnorm(abs(t)))*2
# pasa

med = mean(arima012110_int1$res)
sd = sqrt(sum((arima012110_int1$res - med)^2)/(156-12-1-3))
t = sqrt(156-12-1)*med/sd; pv.arima012110_int1 = (1 - pnorm(abs(t)))*2
# pasa

# homoscedasticidad #

arima012011_int1 = analisis(fert, 0,1,2,0,1,1, xreg=cbind(int.arima012011$
regresores, A0138, A0141, A0152, A019, A090), plot = "hom") # pasa

```

```

arima012110_int1 = analisis(fert, 0,1,2,1,1,0, xreg = cbind(int.arima012110$
regresores, A0138, A0141, A0152, A019, A090), plot = "hom") # pasa

# normalidad #

arima012011_int1$norm
arima012110_int1$norm

par(mfrow = c(1, 2))
qqnorm(arima012011_int1$res); qqline(arima012011_int1$res)
hist(arima012011_int1$res, freq = FALSE, col = "lavender",
main = "Residuos", xlab = "residuos", ylim = c(0,0.00015))
lines(density(arima012011_int1$res), col = "red")

par(mfrow = c(1, 2))
qqnorm(arima012110_int1$res); qqline(arima012110_int1$res)
hist(arima012110_int1$res, freq = FALSE, col = "lavender",
main = "Residuos", xlab = "residuos", ylim = c(0,0.00015))
lines(density(arima012110_int1$res), col = "red")

# pasan

arima012011_int1$sign
arima012110_int1$sign

# AICC

arima012011_int1$mod
arima012110_int1$mod

# residuos #

plot(arima012011_int1$res, xlab = "tiempo", ylab = "toneladas",
main = "residuos ARIMA(0,1,2)(0,1,1)")
plot(arima012110_int1$res, xlab = "tiempo", ylab = "toneladas",
main = "residuos ARIMA(0,1,2)(1,1,0)")

# prediccion #

# p es la cantidad de pasos
# m es el mes de 2010 en que se corta la serie

## predicciones a varios pasos
pred012011 = pred012110 = list()
for(p in c(2:4, 6, 12)){
  pred012011[[as.character(p)]] = list()
  pred012110[[as.character(p)]] = list()
  for(m in seq(p, 12, p)){
    aux = window(fert, start = c(1998, 1), end = c(2010, m))
    xreg = cbind(int.arima012011$regresores, A0138, A0141,
A0152, A019, A090)[1:length(aux),]

```

```

    xreg = xreg[,colSums(xreg[1:(nrow(xreg) - p),]) > 0]
    pred012011[[as.character(p)]] [[as.character(m)]] = preds(aux,
    0,1,2,0,1,1, xreg = xreg, pasos = p)
    xreg = cbind(int.arima012110$regresores, A0138, A0141,
    A0152, A019, A090)[1:length(aux),]
    xreg = xreg[,colSums(xreg[1:(nrow(xreg) - p),]) > 0]
    pred012110[[as.character(p)]] [[as.character(m)]] = preds(aux,
    0,1,2,1,1,0, xreg = xreg, pasos = p)
  }
}

## predicciones varias veces a un paso
pred012011_2 = pred012110_2 = list()
for(p in c(6, 12)){
  pred012011_2[[as.character(p)]] = list()
  pred012110_2[[as.character(p)]] = list()
  for(m in seq(p, 12, p)){
    aux = window(fert, start = c(1998, 1), end = c(2010, m))
    xreg = cbind(int.arima012011$regresores, A0138, A0141,
    A0152, A019, A090)[1:length(aux),]
    xreg = xreg[,colSums(xreg[1:(nrow(xreg) - p),]) > 0]
    pred012011_2[[as.character(p)]] [[as.character(m)]] = preds(aux,
    0,1,2,0,1,1, xreg = xreg, pasos = 1, cuantos = p)
    xreg = cbind(int.arima012110$regresores, A0138, A0141,
    A0152, A019, A090)[1:length(aux),]
    xreg = xreg[,colSums(xreg[1:(nrow(xreg) - p),]) > 0]
    pred012110_2[[as.character(p)]] [[as.character(m)]] = preds(aux,
    0,1,2,1,1,0, xreg = xreg, pasos = 1, cuantos = p)
  }
}

#####
##### FUNCIONES #####
#####

#funcion para testear transformacion logaritimica
log.test = function(serie1, alfa = NULL, s= NULL, plot = F){
  if(is.null(alfa)) alfa = 0.05; if(is.null(s)) s = 12
  if(class(serie1) != 'ts') serie1 = ts(serie1, frequency = s)
  if(alfa < 0 | alfa > 1) stop('alfa tiene que estar entre 0 y 1')
  n = floor(length(serie1)/s)
  Z = matrix(serie1[1:(n*s)], ncol = s, byrow = T)
  max = apply(Z, 1, max); min = apply(Z, 1, min)
  rango = max - min
  media = apply(Z, 1, mean); sd = apply(Z, 1, sd)
  ml = lm(log(rango) ~ log(media))
  est = summary(ml)$coefficients[2, 1]; desv = summary(ml)$coefficients[2, 2];
  if(plot == T) plot(log(media), log(rango))
  lambda = 1 - as.numeric(ml$coef[2])
  if(1 - pnorm((est - 1)/desv) < alfa) decision = 'TRANSFORMAR'
  else decision = ('NO TRANSFORMAR')
}

```



```

    cosas = list(BOXCOX = lambda, summary(ml), Decision = decision)
}

#### funcion para crear regresoras que modelicen los dias comerciales
DC = function(largo){
a1 = rep(seq(1:12), 50) # indica el mes
# indica la cant de dias del mes
a2 = rep(c(31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31), 50)
i = 3 + 4*0:100 # indica que a\~nos son biciestos
# indica los febreros de los anios biciestos
a3 = rep(0, 12*50); a3[12*(i[i <= 50] - 1) + 2] = 1
d = sum(a2[1:largo] + a3[1:largo])
dia = rep(1:7, round(d/7) + 1); dia = dia[4:(d + 3)]
# 1=lunes, 2=martes ... 7=domingo, 1 de enero del 98 fue jueves
mes = a2[1:largo] + a3[1:largo]
# dia esta entre 1 y 7 indica que dia fue, mes indica cant de dias en cada mes
# indica el mes al que corresponde cada dia (entre 1 y 152)
ind_mes = rep(1:largo, mes); ind_mes = ind_mes[1:d]
# cantidad de dias de la semana en cada mes, a partir de enero 1998
cal = tapply(rep(1, d), list(ind_mes, dia), sum)

## algunas cantidades de necesito:
lav = rowSums(cal[, 1:5]) # cantidad de dias de lunes a viernes por mes
syd = rowSums(cal[, 6:7]) # cantidad de sabados y domingos por mes
las = rowSums(cal[, 1:6]) # cantidad de dias de lunes a sabado por mes
lun = cal[,1] # cantidad de dias lunes por mes
mar = cal[,2] # cantidad de dias martes por mes
mie = cal[,3] # cantidad de dias miercoles por mes
jue = cal[,4] # cantidad de dias jueves por mes
vie = cal[,5] # cantidad de dias viernes por mes
sab = cal[,6] # cantidad de dias sabado por mes
dom = cal[,7] # cantidad de dias domingo por mes
# op 1: suma de dias de lun a vier en el mes - 5/2*(cant de sab y dom en el mes)
# op 2: 2 vars; una igual a op 1 y la otra es cant de dias en el mes menos 365.25/12
# op 3: suma de dias de lun a sab - 6*(cant de dom en el mes)
# op 4: 2 variables, una igual a op 3, otra es la cant de dias en el mes - 365.25/12
# op 5: suma de dias entre lun y sab en el mes
# op 6: 6 vars que consideran la cant de dias hab menos la cant de dom en el mes
# (ej #lunes - #domingos)
# op 7: igual que el ant mas una var que es la cantidad de dias en el mes 365.25/12

op1 = data.frame(op1 = lav - 5/2*syd)
op2.1 = op1; op2.2 = mes - 365.25/12; op2 = data.frame(op2.1, op2.2)
names(op2) = c("op2.1", "op2.2")
op3 = data.frame(op3 = las - 6*dom)
op4.1 = op3; op4.2 = op2.2; op4 = data.frame(op4.1, op4.2)
names(op4) = c("op4.1", "op4.2")
op5 = data.frame(op5 = las)
op6.1 = lun - dom; op6.2 = mar - dom; op6.3 = mie - dom; op6.4 = jue - dom
op6.5 = vie - dom; op6.6 = sab - dom
op6 = data.frame(op6.1, op6.2, op6.3, op6.4, op6.5, op6.6)

```

```

names(op6) = c("op6.1", "op6.2", "op6.3", "op6.4", "op6.5", "op6.6")
op7.1 = op6; names(op7.1) = c("op7.1", "op7.2", "op7.3", "op7.4", "op7.5", "op7.6")
op7.2 = op2.2; op7 = data.frame(op7.1, op7.7 = op7.2)

return(list(op1 = op1, op2 = op2, op3 = op3, op4 = op4, op5 = op5,
op6 = op6, op7 = op7)) }

# funcion para probar significacion de variables regresoras por efecto calendario
efcal = function(serie, p = 0, d = 0, q = 0, P = 0, D = 0, Q = 0, period =
frequency(serie), xreg = NULL, alfa = 0.05, todo = FALSE, inicio = NULL,
include.mean = TRUE){
source("C:\\PASANTIA\\VENTAS\\DC"); op = DC(length(serie))
if(!is.null(inicio)){
op = DC(length(serie) + inicio - 1); op1 = op
for(i in 1:length(names(op))){
op[[i]] = data.frame(op[[i]][inicio:nrow(op[[i]]),])
names(op[[i]]) = names(op1[[i]]) } }
op1 = op$op1; op2 = op$op2; op3 = op$op3; op4 = op$op4
op5 = op$op5; op6 = op$op6; op7 = op$op7
xreg1 = xreg
if(!is.null(xreg1)){ xreg1 = as.matrix(xreg1) }
setwd("C:/PASANTIA/VENTAS/SERIES VENTAS")
nolaborables = read.table("nolaborables.txt")
turismo = read.table("turismo.txt")
if(is.null(inicio)){ tur = data.frame(tur = turismo[1:length(serie),])
nolab = data.frame(nolab = nolaborables[1:length(serie),])
} else{ tur = data.frame(tur = turismo[inicio:(length(serie) + inicio - 1),])
nolab = data.frame(nolab = nolaborables[inicio:(length(serie) + inicio - 1),])
}
lista = list("ninguna opcion" = NULL, op1 = op1, op2 = op2, op3 = op3, op4 = op4
op5 = op5, op6 = op6, op7 = op7)
listita = list("ni turismo ni dias no laborables" = NULL, turismo = tur,
"no laborables" = nolab, "turismo + no laborables" = cbind(tur, nolab))
nombres = rep(0, length(lista)*length(listita))
lista1 = list(a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL,
a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL,
a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL,
a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL, a = NULL,
a = NULL)
haysign = rep(0, length(lista)*length(listita))
for(i in 1:length(lista)){
for(j in 1: length(listita)){
if(!is.null(xreg1)){
if(i == 1 & j== 1){ xreg = xreg1 }
if(i == 1 & j != 1){ xreg = cbind(xreg1, listita[[j]]) }
if(i != 1 & j == 1){ xreg = cbind(xreg1, lista[i][[1]]) }
if(i > 1 & j > 1){ xreg = cbind(xreg1, lista[i][[1]], listita[[j]])}
mod = analisis(serie, p = p, d = d, q = q, P = P, D = D, Q = Q, period =
period, xreg = xreg, plot = FALSE, include.mean = include.mean)
} if(is.null(xreg1)){
if(i == 1 & j == 1){ xreg = NULL }

```

```

    if(i == 1 & j != 1){ xreg = listita[[j]] }
    if(i != 1 & j == 1){ xreg = lista[i][[1]] }
    if(i > 1 & j > 1){ xreg = cbind(lista[i][[1]], listita[[j]])}
    mod = analisis(serie, p = p, d = d, q = q, P = P, D = D, Q = Q, period =
    period, xreg = xreg, plot = FALSE, include.mean = include.mean) }
    sign = mod$sign; sign$sign = "NO"; sign$sign[sign[, 4] < alfa] = "SI"
    nombres[(i - 1)*4 + j] = paste(names(lista[i]), names(listita[j]), sep = " + ")
    lista1[[i - 1)*4 + j]] = sign; c = p + q + P + Q
    if(d == 0){ c = c + 1 }; if(!is.null(xreg1)){ c = c + ncol(xreg1) }
    hay = ifelse("SI" %in% sign[-seq(c), 5], 1, 0); haysign[[i - 1)*4 + j]] = hay
  } }
names(lista1) = nombres
if(1 %in% haysign){ con.sign = names(lista1)[1:32*haysign]
} else{ con.sign = "ninguno" }
if(1 %in% haysign){ mod.con.sign = lista1[1:32*haysign]
} else{ mod.con.sign = "no corresponde" }
if(todo == FALSE){ return(list("modelos con regresoras significativas" =
    con.sign, sign = mod.con.sign)) }
if(todo == TRUE){ return(lista1) } }

#### funcion para crear regresoras que modelicen los dias comerciales
DC = function(largo){
a1 = rep(seq(1:12), 50) # indica el mes
a2 = rep(c(31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31), 50) # cant dias del mes
i = 3 + 4*0:100 # indica que a\ nos son biciestos
a3 = rep(0, 12*50); a3[12*(i[i <= 50] - 1) + 2] = 1 # feb de a\ nos biciestos
d = sum(a2[1:largo] + a3[1:largo])
dia = rep(1:7, round(d/7) + 1); dia = dia[4:(d + 3)]
mes = a2[1:largo] + a3[1:largo]
ind_mes = rep(1:largo, mes); ind_mes = ind_mes[1:d]
cal = tapply(rep(1, d), list(ind_mes, dia), sum)
lav = rowSums(cal[, 1:5]); syd = rowSums(cal[, 6:7])
las = rowSums(cal[, 1:6]); lun = cal[,1]
mar = cal[,2]; mie = cal[,3]; jue = cal[,4]
vie = cal[,5]; sab = cal[,6]; dom = cal[,7]
# op 1: suma de dias de lun a vier en el mes - 5/2*(cant de sab y dom en el mes)
# op 2: 2 vars; una igual a op 1 y la otra es cant de dias en el mes menos 365.25/12
# op 3: suma de dias de lun a sab - 6*(cant de dom en el mes)
# op 4: 2 variables, una igual a op 3, otra es la cant de dias en el mes - 365.25/12
# op 5: suma de dias entre lun y sab en el mes
# op 6: 6 vars que consideran la cant de dias hab menos la cant de dom en el mes
# (ej #lunes - #domingos)
# op 7: igual que el ant mas una var que es la cant de dias en el mes 365.25/12
op1 = data.frame(op1 = lav - 5/2*syd)
op2.1 = op1; op2.2 = mes - 365.25/12; op2 = data.frame(op2.1, op2.2)
names(op2) = c("op2.1", "op2.2")
op3 = data.frame(op3 = las - 6*dom)
op4.1 = op3; op4.2 = op2.2; op4 = data.frame(op4.1, op4.2)
names(op4) = c("op4.1", "op4.2")
op5 = data.frame(op5 = las)
op6.1 = lun - dom; op6.2 = mar - dom; op6.3 = mie - dom; op6.4 = jue - dom

```

```

op6.5 = vie - dom; op6.6 = sab - dom
op6 = data.frame(op6.1, op6.2, op6.3, op6.4, op6.5, op6.6)
names(op6) = c("op6.1", "op6.2", "op6.3", "op6.4", "op6.5", "op6.6")
op7.1 = op6; names(op7.1) = c("op7.1", "op7.2", "op7.3", "op7.4", "op7.5", "op7.6")
op7.2 = op2.2; op7 = data.frame(op7.1, op7.2 = op7.2)
return(list(op1 = op1, op2 = op2, op3 = op3, op4 = op4, op5 = op5,
op6 = op6, op7 = op7)) }

# funcion para hacer predicciones
preds = function(serie, p = 0, d = 0, q = 0, P = 0, D = 0, Q = 0, period =
frequency(serie), xreg = NULL, include.mean = TRUE, pasos = NULL, cuantos = NULL,
log = FALSE, fuera = FALSE, xregfuera = NULL, plot = TRUE, main = "Prediccion"){
if(fuera == TRUE & !is.null(xreg) & is.null(xregfuera)){
  stop("DEBE ESPECIFICAR LAS REGRESORAS PARA PREDECIR FUERA DE LA MUESTRA")}
if(is.null(serie)){ stop("DEBE PROPORCIONAR UNA SERIE") }
if(identical(c(p, d, q, P, D, Q), rep(0, 6))){
  stop("DEBE ESPECIFICAR UN MODELO ARIMA") }
if(is.null(pasos)){ stop("DEBE ESPECIFICAR LOS PASOS") }
if(pasos == 1 & is.null(cuantos)){ stop("DEBE ESPECIFICAR CUANTAS VECES A UN PASO") }
library(forecast); n = length(serie)
if(pasos != 1){
  serie1 = ts(serie[1:(n - pasos)], start = start(serie), frequency =
frequency(serie)); corte = end(serie1)
  if(corte[2] == 12){ corte1 = c(corte[1] + 1, 1) }
  if(corte[2] < 12){ corte1 = c(corte[1], corte[2] + 1) }
  reales = ts(serie[(n - pasos + 1):n], start = corte1, frequency =
frequency(serie))
}
if(pasos == 1){
  serie1 = ts(serie[1:(n - cuantos)], start = start(serie), frequency =
frequency(serie)); corte = end(serie1)
  if(corte[2] == 12){ corte1 = c(corte[1] + 1, 1) }
  if(corte[2] < 12){ corte1 = c(corte[1], corte[2] + 1) }
  reales = ts(serie[(n - cuantos + 1):n], start = corte1, frequency =
frequency(serie))
}
if(!is.null(xreg)){ xreg = as.matrix(xreg) }
if(is.null(xreg)){ xreg1 = NULL }
if(fuera == FALSE){
if(pasos > 1){
if(!is.null(xreg)){
if(ncol(xreg) > 1) xreg1 = xreg[1:(n - pasos),]
else xreg1 = xreg[1:(n - pasos)]
xreg1 = as.matrix(xreg1)
}
mod = arima(serie1, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = period), xreg = xreg1, include.mean = include.mean)
if(is.null(xreg)){ newxreg1 = NULL }
if(!is.null(xreg)){
if(ncol(xreg) > 1) newxreg1 = xreg[(n - pasos + 1):nrow(xreg),]
else newxreg1 = xreg[(n - pasos + 1):length(xreg)]

```

```

}
pred = predict(mod, n.ahead = pasos, newxreg = newxreg1)
predic = pred$pred
se = pred$se
e = reales - predic
er = 100*e/reales
ecm = mean(e^2)
U = predic + 1.96*se
L = predic - 1.96*se
if(log == FALSE){
  respred = data.frame(reales, predic, e, er, L, U)
  names(respred) = c("observado", "predicho", "error", "error relat",
    "lim inf IC", "lim sup IC")
}
if(log == TRUE){
  predic = exp(pred$pred)
  se = exp(pred$se)
  e = exp(reales) - predic
  er = 100*e/exp(reales)
  ecm = mean(e^2)
  U = exp(U); L = exp(L)
  serie = exp(serie)
  respred = data.frame(exp(reales), predic, e, er, L, U)
  names(respred) = c("observado", "predicho", "error",
    "error relat", "lim inf IC", "lim sup IC")
}
if(plot == TRUE){
  month = (n - 50):n
  plot(month, serie[month], type = "o", main = main, xlab = "mes", ylab =
    "toneladas", xlim = c(n - 55, n + 5), ylim = c(min(c(serie[month], L)),
    max(c(serie[month], U))))
  lines((n - pasos + 1):n, predic, col = "red", type = "o")
  lines((n - pasos + 1):n, U, col = "blue", lty = "dashed")
  lines((n - pasos + 1):n, L, col = "blue", lty = "dashed")
}
return(list(mod = mod, ecm = ecm, errores = respred))
}
if(pasos == 1){
if(is.null(xreg)){ aux = newxreg1 = NULL }
est = rep(0, cuantos)
sd = rep(0, cuantos)
for(i in 0:(cuantos - 1)){
  cut = ts(serie[1:(n - cuantos + i)], start = start(serie),
    frequency = frequency(serie))
  if(!is.null(xreg)){
    if(ncol(xreg) > 1){ aux = xreg[1:length(cut),]; newxreg1 = matrix(
      xreg[length(cut) + 1,], ncol = ncol(xreg)) }
    if(ncol(xreg) == 1){ aux = xreg[1:length(cut)]; newxreg1 = matrix(
      xreg[length(cut) + 1], ncol = 1) }
  }
}
m = arima(cut, order = c(p, d, q), seasonal = list(order = c(P, D, Q),

```

```

    period = 12), xreg = aux, include.mean = include.mean)
pred = predict(m, n.ahead = 1, newxreg = newxreg1)
est[i + 1] = pred$pred
sd[i + 1] = pred$se
ic_sup = est + 1.96*sd
ic_inf = est - 1.96*sd
}

if(log == FALSE){
  tab = data.frame(reales, est,
  e = reales - est,
  erel = (reales - est)/reales*100,
  infIC = (est - 1.96*sd), supIC = (est + 1.96*sd))
  names(tab) = c("observado", "predicho", "error", "error relat",
  "lim inf IC", "lim sup IC")
  ecm = mean((reales - est)^2)
  nombrey = "toneladas"
}
if(log == TRUE){
  tab = data.frame(reales = exp(reales), est = exp(est),
  e = exp(reales) - exp(est),
  erel = (exp(reales) - exp(est))/exp(reales)*100,
  infIC = exp(est - 1.96*sd), supIC = exp(est + 1.96*sd))
  names(tab) = c("observado", "predicho", "error", "error relat",
  "lim inf IC", "lim sup IC")
  ecm = mean((exp(reales) - exp(est))^2)
  ic_inf = exp(est - 1.96*sd); ic_sup = exp(est + 1.96*sd)
  serie = exp(serie); est = exp(est); sd = exp(sd)
}
if(plot == TRUE){
  month = (n - 50):n
  plot(month, serie[month], type = "o", main = main, xlab = "mes", ylab =
  "toneladas", xlim = c(n - 55, n + 5), ylim = c(min(c(serie[month], ic_inf)),
  max(c(serie[month], ic_sup))))
  lines((n - cuantos + 1):n, est, col = "red", type = "o")
  lines((n - cuantos + 1):n, ic_inf, col = "blue", lty = "dashed")
  lines((n - cuantos + 1):n, ic_sup, col = "blue", lty = "dashed")
}
return(list(tab = tab, ecm = ecm))
}
}
if(fuera == TRUE){
mod = arima(serie, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period
= period), xreg = xreg, include.mean = include.mean)
if(log == FALSE){
  pred = forecast(mod, h = pasos, level = c(80, 95), xreg = xregfuera)
  plot(pred, main = main); return(pred)
}
if(log == TRUE){
  pred = predict(mod, n.ahead = pasos, newxreg = xregfuera)
  U = exp(pred$pred + 1.96*pred$se)
}
}

```

```

L = exp(pred$pred - 1.96*pred$se)
serie = exp(serie)
min = min(serie)
max = max(serie)
tab = data.frame(exp(pred$pred), L, U)
names(tab) = c("estimacion", "inf IC 95%", "sup IC 95%")
if(plot == TRUE){
  month = (n - 50):n
  plot(month, serie[month], type = "o", main = main, xlab = "mes", ylab =
    "toneladas", xlim = c(n - 55, n + 5), ylim = c(min, max))
  lines((n + 1):(n + pasos),exp(pred$pred), col = "red", type = "o")
  lines((n + 1):(n + pasos), L, col = "blue", lty = "dashed")
  lines((n + 1):(n + pasos), U, col = "blue", lty = "dashed")
}
return(tab)
}
}
}

#####
## FUNCIONES CREADAS POR FERNANDO MASSA, UTILIZADAS EN MI CODIGO ##
#####

# funcion que combina los polinomios estacionales y regulares en un solo polinomio
# rv son los parametros regulares y Sv son los parametros estacionales
# insumo para identificar y estimar los IO
comb.coef = function(rv, Sv, s) {
  rn = length(rv); Sn = length(Sv)
  if (rn == 1 & 0 %in% rv) { v = rep(0, s*Sn)
    for (i in 1:Sn) v[s*i] = Sv[i]: return(v) }
  if (rn >= 1 & Sn >= 1 & sum(rv != 0) != 0 & sum(Sv != 0) != 0) {
    v = rep(0, (s*Sn + rn + 1)); for (i in 1:(Sn + 1)){ for (j in 1:(rn + 1)) {
      v[j + s*(i - 1)] = c(1, rv)[j]*c(1, Sv)[i] } }
    v = v[-1]; return(v) } }

#funcion que crea la indicatriz para los atipicos innovativos
IO = function(modelo, donde, n) {
  d = modelo$arma[6]; D = modelo$arma[7]; s = modelo$arma[5]
  p = modelo$arma[1]; q = modelo$arma[2]; P = modelo$arma[3]; Q = modelo$arma[4]
  if(any(names(modelo$coef) %in% paste('ma', 1:q, sep = ''))) {
    ma = modelo$coef[which(names(modelo$coef) %in% paste('ma', 1:q, sep = ''))]
  } else{ ma = 0 }
  if(any(names(modelo$coef) %in% paste('sma', 1:Q, sep = ''))) {
    sma = modelo$coef[which(names(modelo$coef) %in% paste('sma', 1:Q, sep = ''))]
  } else{ sma = 0 }
  if(any(names(modelo$coef) %in% paste('ar', 1:p, sep = ''))) {
    ar = modelo$coef[which(names(modelo$coef) %in% paste('ar', 1:p, sep = ''))]
  } else{ ar = 0 }
  if(any(names(modelo$coef) %in% paste('sar', 1:P, sep = ''))) {
    sar = modelo$coef[which(names(modelo$coef) %in% paste('sar', 1:P, sep = ''))]
  } else{ sar = 0 }
}

```

```

if(sum(0 %in% sma) != 0) MA = ma; if(sum(0 %in% sar) != 0) AR = ar
if(sum(0 %in% ma) != 0) MA = comb.coef(0, sma, s)
if(sum(0 %in% ar) != 0) AR = comb.coef(0, sar, s)
if(sum(ma != 0) != 0 & sum(sma != 0) != 0) MA = comb.coef(ma, sma, s)
if(sum(ar != 0) != 0 & sum(sar != 0) != 0) AR = comb.coef(ar, sar, s)
if(donde > n) stop('Los parametros introducidos no son correctos ')
io = rep(0, n)
if(sum(MA) != 0 & sum(AR) != 0) psi = ARMAtoMA(ma = MA, ar = AR, lag.max = n)
if(sum(MA) == 0 & sum(AR) != 0) psi = ARMAtoMA(ar = AR, lag.max = n)
if(sum(MA) != 0 & sum(AR) == 0) psi = ARMAtoMA(ma = MA, lag.max = n)
psi = c(1, psi)
if(d == 0 & D == 0) { for (i in donde:n){ io[i] = psi[i - donde + 1] } }
if( d >= 1 & D == 0) { for (i in donde:n){ io[i] = diffinv(psi, lag = 1,
differences = d)[- seq(D*s + d)][i - donde + 1] } }
if(d == 0 & D >= 1) { for (i in donde:n){ io[i] = diffinv(psi, lag = s,
differences = D)[- seq(D*s + d)][i - donde + 1] } }
if (d >= 1 & D >= 1){ for (i in donde:n){ io[i] = (diffinv(diffinv(psi, lag = s,
differences = D), lag = 1, differences = d)[- seq(D*s + d))][i - donde + 1] } }
return(io) }

# funcion que crea la indicatriz para los cambios transitorios
# cuanto*s es el largo de la progresion geometrica
TC = function(donde, n, cuanto = NULL, s = NULL, p = NULL) {
  if (is.null(s)) s = 12; if (is.null(p)) p = 0.7; if (is.null(cuanto)) cuanto = 3
  tc = rep(0, n); tc[donde] = 1; final = (donde + cuanto*s - 1)
  if (donde + cuanto*s > n) final = n; for (i in (donde + 1):final)
    tc[i] = round(p*tc[i - 1], 5); return(tc) }

# funcion que crea la indicatriz para los cambios de nivel
LS = function(donde, n) {
  ls0 = rep(0, (donde - 1)); ls1 = rep(1, (n - donde + 1))
  ls = c(ls0, ls1); return(ls) }

# funcion que crea la indicatriz para los cambios de nivel estacionales
SLS = function(donde, n, s = NULL) {
  if (is.null(s)) s = 12; sls = rep(0, n); sls[donde] = 1
  for (i in donde:n) if (((i - donde)/s) == round((i - donde)/s)) sls[i] = 1
  return(sls) }

# funcion para crear la indicatriz de un atpico estacional (a un mes)
SOa = function(donde, n, s = NULL) {
  if (is.null(s)) s = 12
  soa1 = ARMAtoMA(ar = c(-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1), ma = 0, n)
  soa = rep(0, n); soa[donde] = 1
  for (i in donde:n) soa[i+1] = soa1[i - donde + 1]; return(soa) }

# funcion para crear la indicatriz de un atipico estacional (a varios meses)
SOB = function(donde, n, s = NULL) {
  if (is.null(s)) s = 12; soa = ARMAtoMA(ar = c(rep(0, s - 1), 1), ma = 0, n)
  sob = -(1/(s - 1))*rep(1, n); soa = c(1, soa)[- (n + 1)]
  soa2 = rep(0, n); for (i in donde:n) { soa2[i] = soa[i - donde + 1]

```



```

if (soa2[i] != 1) soa2[i] = sob[i] }; return(soa2) }

# funcion para poner las fechas al data.frame de los outliers
fechas = function(reg1, regr, serie, s = NULL) {
if (class(serie) != 'ts') stop('Debes introducir una serie de tiempo')
principio = start(serie)[1]; mes1 = start(serie)[2]
s = frequency(serie); largo = nrow(reg1); fin = end(serie)[1]
mes2 = end(serie)[2]; completos = (length(serie) - mes2 - (s - mes1 + 1))/s
a\ no = c(rep(principio, s - mes1 + 1), principio +
as.numeric((gl((fin - principio - 1), s))), rep(fin, mes2))
mes = c(seq(mes1, s), rep(seq(s), completos), seq(mes2))); mes = as.factor(mes)
if(s == 12) levels(mes) = c('Ene', 'Feb', 'Mar', 'Abr', 'May',
'Jun', 'Jul', 'Ago', 'Set', 'Oct', 'Nov', 'Dic')
if(s == 6) levels(mes) = c('1Â° bim', '2Â° bim', '3Â° bim', '4Â° bim',
'5Â° bim', '6Â° bim')
if(s == 4) levels(mes) = c('1Â° trim', '2Â° trim', '3Â° trim', '4Â° trim')
if(s == 3) levels(mes) = c('1Â° cuat', '2Â° cuat', '3Â° cuat')
if(s == 2) levels(mes) = c('1Â° sem', '2Â° sem')
fechas = data.frame(a\ no, mes)
if (s==6) colnames(fechas) = c('a\ no', 'bimestre')
if (s==4) colnames(fechas) = c('a\ no', 'trimestre')
if (s==3) colnames(fechas) = c('a\ no', 'cuatrimestre')
if (s==2) colnames(fechas) = c('a\ no', 'semestre')
cuantos = ncol(reg1) - regr; cual = rep(0, cuantos)
for (k in 1:cuantos) cual[k] = min(which(reg1[, (k + regr)] != 0))
final = fechas[cual,] }

# funcion para hacer un data.frame con la significacion
# de los coeficientes de un arima
significacion = function(mod, print = T){
if (class(mod) == 'Arima') {
valores = mod$coef; desvios = sqrt(diag(mod$var.coef)) }
if (class(mod) == 'ar') {
valores = mod$ar; desvios = sqrt(diag(mod$asy.var.coef)) }
pval = 1 - pnorm(abs(valores/desvios)); sign = pr(pval)
signif = data.frame(valores, desvios, valores/desvios, pval, sign)
colnames(signif) = c('Estimacion', 'Desvio', 't-valor', 'Pr(>|t|)', 'sig')
rownames(signif) = names(mod$coef)
a = '*** significativo al 0.1%'; b = '** significativo al 1%'
c = '* significativo al 5%'; d = '. significativo al 10%'
nota = rbind(a, b, c, d); colnames(nota) = 'Nota'
signif = list(Significacion = signif, nota)
if (print == T) print(signif) invisible(signif[[1]]) }

#funcion para detectar outliers
detecta = function(serie, modelo, Sf = NULL, C = NULL, dummy = NULL,
clase = NULL, maxiter = 3, methodCSS = FALSE, methodCSS_ML = FALSE) {
if(is.null(dummy)) dummy = 0; if(!dummy %in% c(0, 1)) dummy = 1
if(is.null(Sf)) Sf = 0; if(is.null(clase)) clase = rep(1, 7)
if(sum(clase) == 0) C = 100000; c = sum(modelo$ar[1:4])
if(all(modelo$ar[6:7] == 0)) c = c + 1

```

```

w = length(serie); s = modelo$arima[5]
p = modelo$arima[1]; d = modelo$arima[6]; q = modelo$arima[2];
P = modelo$arima[3]; D = modelo$arima[7]; Q = modelo$arima[4];
if(is.null(C)) C = 3 + 0.0025*(w - 50)
son = 1; j = 0; maxiter = maxiter; xdet = NULL; son1 = NULL
while(sum(son) != 0 & j <= maxiter) {
j = j + 1; residuos = residuals(modelo)
atipicos1 = which(abs(residuos) > median(residuos) + 2*sd(residuos))
testeOLS = TRUE; if(1 %in% atipicos1) testeOLS = FALSE
  # SI EL ATIPICO ES LA 1RA OBS DE LA SERIE QUE NO TESTEE CAMBIO DE NIVEL
  # SI NO QUEDA TODO UNOS Y DA ERROR
if(j > 1) { algunos = any(atipicos1 %in% todos)
ninguno = all(atipicos1 %in% todos) if(algunos == T) atipicos1 = atipicos1
  [-which(atipicos1 %in% todos)]; if (ninguno == T) atipicos1 = NULL }
cuantos = length(atipicos1)
if(j > 1) todos = sort(c(todos, atipicos1)) else todos = atipicos1
son = rep(0, cuantos)
if(cuantos > 0){ A1d = matrix(0, ncol = cuantos, nrow = w + Sf)
  colnames(A1d) = paste('IO', 1:cuantos, sep = '')
A1a = matrix(0, ncol = cuantos, nrow = w + Sf)
  colnames(A1a) = paste('AO', 1:cuantos, sep = '')
A1c = matrix(0, ncol = cuantos, nrow = w + Sf)
  colnames(A1c) = paste('LS', 1:cuantos, sep = '')
A1b = matrix(0, ncol = cuantos, nrow = w + Sf)
  colnames(A1b) = paste('TC', 1:cuantos, sep = '')
A1e = matrix(0, ncol = cuantos, nrow = w + Sf)
  colnames(A1e) = paste('SLS', 1:cuantos, sep = '')
A1f = matrix(0, ncol = cuantos, nrow = w + Sf)
  colnames(A1f) = paste('SOa', 1:cuantos, sep = '')
A1g = matrix(0, ncol = cuantos, nrow = w + Sf)
  colnames(A1g) = paste('SOB', 1:cuantos, sep = '')
if(j == 1){ tc = TC(1, w + Sf); ls = LS(1, w + Sf)
io = IO(modelo, 1, w + Sf); sls = SLS(1, (w + Sf), s)
soa = SOa(1, (w + Sf), s); sob = SOB(1, (w + Sf), s) }
for(i in 1:cuantos) { A1a[atipicos1[i], i] = 1
A1b[(atipicos1[i):(w + Sf)), i] = tc[1:(w + Sf - atipicos1[i] + 1)]
A1c[(atipicos1[i):(w + Sf)), i] = ls[1:(w + Sf - atipicos1[i] + 1)]
A1d[(atipicos1[i):(w + Sf)), i] = io[1:(w + Sf - atipicos1[i] + 1)]
A1e[(atipicos1[i):(w + Sf)), i] = sls[1:(w + Sf - atipicos1[i] + 1)]
A1f[(atipicos1[i):(w + Sf)), i] = soa[1:(w + Sf - atipicos1[i] + 1)]
A1g[(atipicos1[i):(w + Sf)), i] = sob[1:(w + Sf - atipicos1[i] + 1)] }
for (i in 1:cuantos){ x1 = A1a[(1:w), i, drop = F]
  x2 = A1b[(1:w), i, drop = F], x3 = A1c[(1:w), i, drop = F]
  x4 = A1d[(1:w), i, drop = F]; x5 = A1e[(1:w), i, drop = F]
  x6 = A1f[(1:w), i, drop = F]; x7 = A1g[(1:w), i, drop = F]
if(methodCSS == FALSE & methodCSS_ML == FALSE) {
if(clase[1] != 0) modAO = arima(serie, order = c(p, d, q), seasonal =
  list(order = c(P, D, Q), period = s), xreg = x1, method = 'ML')
if(clase[2] != 0) modTC = arima(serie, order = c(p, d, q), seasonal =
  list(order = c(P, D, Q), period = s), xreg = x2, method = 'ML')
if(testeOLS == TRUE & clase[3] != 0) modLS = arima(serie, order = c(p, d, q),

```

```

    seasonal = list(order = c(P, D, Q), period = s), xreg = x3, method = 'ML') }
if(methodCSS == TRUE & methodCSS_ML == FALSE) {
if(clase[1] != 0) modAO = arima(serie, order = c(p, d, q), seasonal =
    list(order = c(P, D, Q), period = s), xreg = x1, method = 'CSS')
if(clase[2] != 0) modTC = arima(serie, order = c(p, d, q), seasonal =
    list(order = c(P, D, Q), period = s), xreg = x2, method = 'CSS')
if(testeoLS == TRUE & clase[3] != 0) modLS = arima(serie, order = c(p, d, q),
    seasonal = list(order = c(P, D, Q), period = s), xreg = x3, method = 'CSS') }
if(methodCSS == FALSE & methodCSS_ML == TRUE) { if(clase[1] != 0) modAO =
    arima(serie, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period =
    s), xreg = x1, method = 'CSS-ML'); if(clase[2] != 0) modTC = arima(serie,
    order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = s), xreg = x2,
    method = 'CSS-ML'); if(testeoLS == TRUE & clase[3] != 0) modLS = arima(serie,
    order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = s), xreg = x3,
    method = 'CSS-ML') }; if(clase[4] != 0){ conv = 1; k = 0; nIO = w + 100
while(conv > 0.0001 & k <= 50) { k = k + 1
if(methodCSS == FALSE & methodCSS_ML == FALSE){ modIO = arima(serie,
    order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = s),
    xreg = x4, method = 'ML') }
if(methodCSS == TRUE & methodCSS_ML == FALSE){ modIO = arima(serie,
    order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = s),
    xreg = x4, method = 'CSS') }
if(methodCSS == FALSE & methodCSS_ML == TRUE){ modIO = arima(serie,
    order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = s),
    xreg = x4, method = 'CSS-ML') }
io2 = IO(modIO, atipicos1[i], nIO); A1d[, i] = io2[1:(w + Sf)]
if(k > 1) conv = abs(sqrt(sum(as.numeric(modIO$coef)^2)) -
    sqrt(sum(as.numeric(par)^2))); par = as.numeric(modIO$coef)
x4 = matrix(io2[1:w], ncol = 1) } }
if (sum(clase[5:7]) > 0) {
    if (dummy == 0 & atipicos1[i] %in% ((s + 1):(w - s - 1))) {
if (clase[5] != 0) modSLS = arima(serie, order = c(p, d, q), seasonal =
    list(order = c(P, D, Q), period = s), xreg = x5, method = 'ML')
if (clase[6] != 0) modSOa = arima(serie, order = c(p, d, q), seasonal =
    list(order = c(P, D, Q), period = s), xreg = x6, method = 'ML')
if (clase[7] != 0) modSOb = arima(serie, order = c(p, d, q), seasonal =
    list(order = c(P, D, Q), period = s), xreg = x7, method = 'ML') } }
if(clase[1] != 0) gao = ((modAO$coef)/sqrt(diag(modAO$var.coef)))
    [length(modAO$coef)] else gao = 0
if(clase[2] != 0) gtc = ((modTC$coef)/sqrt(diag(modTC$var.coef)))
    [length(modTC$coef)] else gtc = 0
if(testeoLS == TRUE & clase[3] != 0) gls = ((modLS$coef)/sqrt(diag(
    modLS$var.coef)))[length(modLS$coef)] else gls = 0
if(clase[4] != 0) gio = ((modIO$coef)/sqrt(diag(modIO$var.coef)))
    [length(modIO$coef)] else gio = 0
if(sum(clase[5:7]) > 0) {
if(dummy == 0 & atipicos1[i] %in% ((s + 1):(w - s - 1))) {
if(clase[5] != 0) gsls = ((modSLS$coef)/sqrt(diag(modSLS$var.coef)))
    [length(modSLS$coef)] else gsls = 0
if(clase[6] != 0) gsoa = ((modSOa$coef)/sqrt(diag(modSOa$var.coef)))
    [length(modSOa$coef)] else gsoa = 0

```

```

if(clase[7] != 0) gsob = ((modSOB$coef)/sqrt(diag(modSOB$var.coef)))
  [length(modSOB$coef)] else gsob = 0 }
if(dummy != 0 | !atipicos1[i] %in% ((s + 1):(w - s - 1)))
  gsols = gsoa = gsob = 0 } else gsols = gsoa = gsob = 0
posibles = c(abs(gao), abs(gtc), abs(gls), abs(gio), abs(gsols),
  abs(gsoa), abs(gsob)); candidato = which.max(c(abs(gao), abs(gtc),
  abs(gls), abs(gio), abs(gsols), abs(gsoa), abs(gsob)))
if(posibles[candidato] > C) son[i] = candidato }
if(j == 1) xdet = data.frame(rep(1, length(serie)))
if(sum(son) != 0) xdet = cbind(xdet, A1a[(1:w), which(son == 1), drop = F],
  A1b[(1:w), which(son == 2), drop = F], A1c[(1:w), which(son == 3), drop = F],
  A1d[(1:w), which(son == 4), drop = F], A1e[(1:w), which(son == 5), drop = F],
  A1f[(1:w), which(son == 6), drop = F], A1g[(1:w), which(son == 7), drop = F])
if(Sf > 0) { if (j == 1) xfdet = data.frame(rep(1, Sf))
if(sum(son) != 0) xfdet = cbind(xfdet, A1a[(w + 1):(w + Sf), which(son == 1),
  drop = F], A1b[(w + 1):(w + Sf), which(son == 2), drop = F], A1c[(w + 1):(w + Sf),
  which(son == 3), drop = F], A1d[(w + 1):(w + Sf), which(son == 4), drop = F],
  A1e[(w + 1):(w + Sf), which(son == 5), drop = F], A1f[(w + 1):(w + Sf),
  which(son == 6), drop = F], A1g[(w + 1):(w + Sf), which(son == 7), drop = F])}
  if(j == 1){ xdet = xdet[, -1, drop = F]
  if(Sf != 0) xfdet = xfdet[, -1, drop = F] }
if(sum(son) != 0){ sale = 1; son1 = c(son1, sort(son[son != 0]))
while (sum(sale) > 0) {
  modelodet = arima(serie, order = c(p, d, q), seasonal = list(order =
  c(P, D, Q), period = s), xreg = xdet, method = 'ML')
t_est = (modelodet$coef/sqrt(diag(modelodet$var.coef)))[-seq(c)]
if(any(abs(t_est) < C)){ sale = which.min(abs(t_est))
if(t_est[sale] < C) sale1 = sale else break
son = son[-sale]; son1 = son1[-sale]; xdet = xdet[, -sale, drop = F]
if(Sf != 0) xfdet = xfdet[, -sale, drop = F] }
if(all(abs(t_est) >= C)) sale = 0; if (ncol(xdet) == 0) break }
Nout = ncol(xdet); modelo = modelodet } } else son = 0 }
nombres1 = NULL
  if(sum(son1 == 1) != 0) nombres1[son1 == 1] = paste('A0', 1:sum(son1 == 1),
  sep = '')
  if(sum(son1 == 2) != 0) nombres1[son1 == 2] = paste('TC', 1:sum(son1 == 2),
  sep = '')
  if(sum(son1 == 3) != 0) nombres1[son1 == 3] = paste('LS', 1:sum(son1 == 3),
  sep = '')
  if(sum(son1 == 4) != 0) nombres1[son1 == 4] = paste('IO', 1:sum(son1 == 4),
  sep = '')
if(sum(son1 == 5) != 0) nombres1[son1 == 5] = paste('SLS', 1:sum(son1 == 5),
  sep = '')
if(sum(son1 == 6) != 0) nombres1[son1 == 6] = paste('SOa', 1:sum(son1 == 6),
  sep = '')
if(sum(son1 == 7) != 0) nombres1[son1 == 7] = paste('SOB', 1:sum(son1 == 7),
  sep = ''); colnames(xdet) = nombres1; if(Sf > 0) colnames(xfdet) = nombres1
if(ncol(xdet) > 0) { if (!is.null(xdet) & any(xdet == 1)) {
cuantos = ncol(xdet); cuando = fechas(xdet, regr = 0, serie)
if(any(paste('A0', 1:cuantos, sep = '') %in% colnames(xdet))) {
  ad = rep('A0', sum(paste('A0', 1:cuantos, sep = '') %in% colnames(xdet)))

```

```

else ad = NULL }
if(any(paste('TC', 1:cuantos, sep = '') %in% colnames(xdet))) {
  ct = rep('TC', sum(paste('TC', 1:cuantos, sep = '') %in% colnames(xdet)))
  else ct = NULL }
if(any(paste('LS', 1:cuantos, sep = '') %in% colnames(xdet))) {
  cn = rep('LS', sum(paste('LS', 1:cuantos, sep = '') %in% colnames(xdet)))
  else cn = NULL }
if(any(paste('IO', 1:cuantos, sep = '') %in% colnames(xdet))) {
  ai = rep('IO', sum(paste('IO', 1:cuantos, sep = '') %in% colnames(xdet)))
  else ai = NULL }
if(any(paste('SLS', 1:cuantos, sep = '') %in% colnames(xdet))) {
  cne = rep('SLS', sum(paste('SLS', 1:cuantos, sep = '') %in% colnames(xdet)))
  else cne = NULL }
if(any(paste('SOa', 1:cuantos, sep = '') %in% colnames(xdet))) {
  aea = rep('SOa', sum(paste('SOa', 1:cuantos, sep = '') %in% colnames(xdet)))
  else aea = NULL }
if(any(paste('SOB', 1:cuantos, sep = '') %in% colnames(xdet))) {
  aeb = rep('SOB', sum(paste('SOB', 1:cuantos, sep = '') %in% colnames(xdet)))
  else aeb = NULL }
tipo = c(ad, ct, cn, ai, cne, aea, aeb)
magnitud = as.numeric(modelodet$coef[-seq(c)])
desvio = sqrt(diag(modelodet$var.coef)[-seq(c)])
  outliers = data.frame(cuando, tipo, magnitud, desvio) } }
  else { outliers = c('No se encontraron outliers') }
if(Sf != 0) salida = list(matriz = rbind(xdet, xfdet), significacion =
  outliers, valorC = C) else salida = list(matriz = xdet,
  significacion = outliers, valorC = C)
names(salida) = c('regresores', 'significacion', 'valor critico')
if (is.null(xdet)) invisible(NULL)
if (ncol(xdet) > 0) { if (all(xdet == 1) & !is.null(xdet)) invisible(NULL) }
if (!is.null(xdet)){ if (ncol(xdet) == 0) invisible(NULL)
if (ncol(xdet) > 0){ if (any(xdet != 1)) invisible(salida) } } }

```

```
#####
### MODELIZACION MULTIVARIADA SERIE DE VENTA DE FERTILIZANTES EN PLAZA ###
#####

setwd("C:/PASANTIA/VENTAS/SERIES VENTAS")
library(urca); library(vars)

# variables
data = read.csv("variables_multi3.csv", dec = ",")

# paso a series
for(i in c("fert", "IVFPBI_agro", "lluvia_pais", "FPI", "p_fert")){
  data[,i] = ts(data[,i], start = c(1998,1), frequency = 12)
}

# grafico de las variables
r1 = c(1,1,2,2,3,3); r2 = c(0,4,4,5,5,0); r = as.matrix(rbind(r1, r2)); layout(r)
plot(data$fert, xlab = "tiempo", ylab = "ton", main = "ventas de fertilizantes")
plot(data$p_fert, xlab = "tiempo", ylab = "US$/ton",main = "precio de fertilizantes")
plot(data$lluvia_pais, xlab = "tiempo", ylab = "mm", main = "lluvia promedio Uruguay")
plot(data$IVFPBI_agro, xlab = "tiempo", ylab = "", main = "IVF PBI agropecuario")
plot(data$FPI, xlab = "tiempo", ylab = "", main = "Indice de precios de alimentos")

#####
### Pruebas de Integracion de Dickey Fuller ###
#####

### FERTILIZANTES ###

plot(data$fert)
par(mfrow = c(2, 1))
acf(data$fert, lag.max = 60)
pacf(data$fert, lag.max = 60)

test1_fert = ur.df(data$fert, type="trend", lags=12)
plot(test1_fert) # capta dinamica

par(mfrow = c(1, 2))
i=rep(0,60); for(j in 12:60){i[j] = Box.test(test1_fert@res,
type = "Ljung-Box", fitdf = 12,lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_fert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 12:60){i[j] = Box.test(test1_fert@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test1_fert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorr

resumen1_fert = cbind(t(test1_fert@teststat), test1_fert@cval)
resumen1_fert # no rechazo que tau3=0 ni que phi3=0
```

```

test2_fert = ur.df(data$fert, type = "drift", lags = 12)
plot(test2_fert)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test2_fert@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test2_fert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test2_fert@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test2_fert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorrelacion

resumen2_fert = cbind(t(test2_fert@teststat), test2_fert@cval)
resumen2_fert # no rechazo tau2=0 ni phi1=0

test3_fert = ur.df(data$fert, type = "none", lags = 12)
plot(test3_fert)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test3_fert@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test3_fert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 12:60){i[j] = Box.test(test3_fert@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test3_fert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorrelacion

resumen3_fert = cbind(t(test3_fert@teststat), test3_fert@cval)
resumen3_fert # no rechazo tau1 = 0 => FERT ES INTEGRADO

dfert = diff(data$fert)
plot(dfert)

test1_dfert = ur.df(dfert, type = "none", lags = 12)
plot(test1_dfert)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test1_dfert@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_dfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test1_dfert@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test1_dfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorr

```

```

resumen1_dfert = cbind(t(test1_dfert@teststat), test1_dfert@cval)
resumen1_dfert
# rechazo que tau3=0 y que phi3=0 => dfert es estacionaria

### => FERT ES I(1)

#####

### PBI AGROPECUARIO ###

plot(data$IVFPBI_agro)

par(mfrow = c(2, 1))
acf(data$IVFPBI_agro, lag.max = 50)
pacf(data$IVFPBI_agro, lag.max = 50)
# tiene estacionalidad

test1_IVFPBI_agro = ur.df(data$IVFPBI_agro, type = "trend", lags = 12)
plot(test1_IVFPBI_agro) # capta dinamica

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test1_IVFPBI_agro@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_IVFPBI_agro", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 12:60){i[j] = Box.test(
test1_IVFPBI_agro@res, type = "Box-Pierce", fitdf = 12, lag = j)$p.value}
plot(i, ylim = c(0, 1), main = "p-valor test Box-Pierce -
residuos test1_IVFPBI_agro", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorr

resumen1_IVFPBI_agro = cbind(t(test1_IVFPBI_agro@teststat),
test1_IVFPBI_agro@cval); resumen1_IVFPBI_agro
# no rechazo que tau3=0 ni que phi3=0

test2_IVFPBI_agro = ur.df(data$IVFPBI_agro, type = "drift", lags = 12)
plot(test2_IVFPBI_agro)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test2_IVFPBI_agro@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test2_IVFPBI_agro", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 12:60){i[j] = Box.test(
test2_IVFPBI_agro@res, type = "Box-Pierce", fitdf = 12, lag = j)$p.value}
plot(i, ylim = c(0, 1), main = "p-valor test Box-Pierce -
residuos test2_IVFPBI_agro", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorrelacion

```



```

resumen2_IVFPBI_agro = cbind(t(test2_IVFPBI_agro@teststat),
test2_IVFPBI_agro@cval); resumen2_IVFPBI_agro # no rechazo tau2=0 ni phi1=0

test3_IVFPBI_agro = ur.df(data$IVFPBI_agro, type = "none", lags = 12)
plot(test3_IVFPBI_agro)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test3_IVFPBI_agro@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test3_IVFPBI_agro", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 12:60){i[j] = Box.test(
test3_IVFPBI_agro@res, type = "Box-Pierce", fitdf = 12, lag = j)$p.value}
plot(i, ylim = c(0, 1), main = "p-valor test Box-Pierce -
residuos test3_IVFPBI_agro", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05" ,col = "red")
# se cumple incorrelacion

resumen3_IVFPBI_agro = cbind(t(test3_IVFPBI_agro@teststat),
test3_IVFPBI_agro@cval); resumen3_IVFPBI_agro
# no rechazo tau1 = 0 => PBI AGROPECUARIO ES INTEGRADO

dIVFPBI_agro = diff(data$IVFPBI_agro)
plot(dIVFPBI_agro)

test1_dIVFPBI_agro = ur.df(dIVFPBI_agro, type = "none", lags = 12)
plot(test1_dIVFPBI_agro) # capta la dinamica

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test1_dIVFPBI_agro@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_dIVFPBI_agro", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 12:60){i[j] = Box.test(
test1_dIVFPBI_agro@res, type = "Box-Pierce", fitdf = 12, lag = j)$p.value}
plot(i, ylim = c(0, 1), main = "p-valor test Box-Pierce -
residuos test1_dIVFPBI_agro", xlab = "lag", ylab = "p-valor")
abline(h = 0.05, col = "red"); text(1, 0.075, "0.05", col = "red")
# se cumple incorr

resumen1_dIVFPBI_agro = cbind(t(test1_dIVFPBI_agro@teststat),
test1_dIVFPBI_agro@cval); resumen1_dIVFPBI_agro
# rechazo tau1 = 0 => dIVFPBI_agro no es integrada

### => PBI AGROPECUARIO ES I(1)

### INDICE DE PRECIOS DE ALIMENTOS (FPI) ###

plot(data$FPI)
par(mfrow = c(2, 1))

```

```

acf(data$FPI, lag.max = 60)
pacf(data$FPI, lag.max = 60)

test1_FPI = ur.df(data$FPI, type="trend", lags=12)
plot(test1_FPI) # capta dinamica

par(mfrow = c(1, 2))
i=rep(0,60); for(j in 12:60){i[j] = Box.test(test1_FPI@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_FPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 12:60){i[j] = Box.test(test1_FPI@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test1_FPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorr

resumen1_FPI = cbind(t(test1_FPI@teststat), test1_FPI@cval)
resumen1_FPI # no rechazo que tau3=0 ni que phi3=0

test2_FPI = ur.df(data$FPI, type = "drift", lags = 12)
plot(test2_FPI)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test2_FPI@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test2_FPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test2_FPI@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test2_FPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorrelacion

resumen2_FPI = cbind(t(test2_FPI@teststat), test2_FPI@cval)
resumen2_FPI # no rechazo tau2=0 ni phi1=0

test3_FPI = ur.df(data$FPI, type = "none", lags = 12)
plot(test3_FPI)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test3_FPI@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test3_FPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 12:60){i[j] = Box.test(test3_FPI@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test3_FPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorrelacion

```

```

resumen3_fert = cbind(t(test3_FPI@teststat), test3_FPI@cval)
resumen3_fert # no rechazo tau1 = 0 => FERT ES INTEGRADO

dFPI = diff(data$FPI)
plot(dFPI)

test1_dFPI = ur.df(dFPI, type = "none", lags = 12)
plot(test1_dFPI)

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 12:60){i[j] = Box.test(test1_dFPI@res,
type = "Ljung-Box", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_dFPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 12:60){i[j]=Box.test(test1_dFPI@res,
type = "Box-Pierce", fitdf = 12, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test1_dFPI", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorr

resumen1_dFPI = cbind(t(test1_dFPI@teststat), test1_dFPI@cval)
resumen1_dFPI # rechazo que tau1 = 0 => dFPI no es integrada

### => FPI ES I(1)

### PRECIO DE FERTILIZANTES (PFERT) ###

plot(data$p_fert)
par(mfrow = c(2, 1))
acf(data$p_fert, lag.max = 60)
pacf(data$p_fert, lag.max = 60)

test1_pfert = ur.df(data$p_fert, type = "trend", lags = 1)
plot(test1_pfert) # capta dinamica

par(mfrow = c(1, 2))
i=rep(0,60); for(j in 1:60){i[j] = Box.test(test1_pfert@res,
type = "Ljung-Box", fitdf = 1,lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_pfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 1:60){i[j] = Box.test(test1_pfert@res,
type = "Box-Pierce", fitdf = 1, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test1_pfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorr

resumen1_pfert = cbind(t(test1_pfert@teststat), test1_pfert@cval)
resumen1_pfert # no rechazo que tau3=0 ni que phi3=0

test2_pfert = ur.df(data$p_fert, type = "drift", lags = 1)
plot(test2_pfert)

```

```

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 1:60){i[j] = Box.test(test2_pfert@res,
type = "Ljung-Box", fitdf = 1, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test2_pfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 1:60){i[j] = Box.test(test2_pfert@res,
type = "Box-Pierce", fitdf = 1, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test2_pfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorrelacion

```

```

resumen2_pfert = cbind(t(test2_pfert@teststat), test2_pfert@cval)
resumen2_pfert # no rechazo tau2=0 ni phi1=0

```

```

test3_pfert = ur.df(data$p_fert, type = "none", lags = 1)
plot(test3_pfert)

```

```

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 1:60){i[j] = Box.test(test3_pfert@res,
type = "Ljung-Box", fitdf = 1, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test3_pfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 1:60){i[j] = Box.test(test3_pfert@res,
type = "Box-Pierce", fitdf = 1, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test3_pfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorrelacion

```

```

resumen3_pfert = cbind(t(test3_pfert@teststat), test3_pfert@cval)
resumen3_pfert # no rechazo tau1 = 0 => FERT ES INTEGRADO

```

```

dpfert = diff(data$p_fert)
plot(dpfert)

```

```

test1_dpfert = ur.df(dpfert, type = "none", lags = 1)
plot(test1_dpfert)

```

```

par(mfrow = c(1, 2))
i = rep(0, 60); for(j in 1:60){i[j] = Box.test(test1_dpfert@res,
type = "Ljung-Box", fitdf = 1, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_dpfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0, 60); for(j in 1:60){i[j]=Box.test(test1_dpfert@res,
type = "Box-Pierce", fitdf = 1, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test1_dpfert", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorr

```

```

resumen1_dpfert = cbind(t(test1_dpfert@teststat), test1_dpfert@cval)
resumen1_dpfert # rechazo que tau1 = 0 => dFPI no es integrada

```

```

### => PFERT ES I(1)

### NIVEL DE LUVIA (LLUVIA) ###

plot(data$lluvia_pais)
par(mfrow = c(2, 1))
acf(data$lluvia_pais, lag.max = 60)
pacf(data$lluvia_pais, lag.max = 60)

test1_lluvia = ur.df(data$lluvia_pais, type = "drift", lags = 0)
plot(test1_lluvia) # capta dinamica

par(mfrow = c(1, 2))
i=rep(0,60); for(j in 1:60){i[j] = Box.test(test1_lluvia@res,
type = "Ljung-Box", fitdf = 0,lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Ljung-Box - \n residuos test1_lluvia", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red"); i = rep(0,60); for(j in 1:60){i[j] = Box.test(test1_lluvia@res,
type = "Box-Pierce", fitdf = 0, lag = j)$p.value}; plot(i, ylim = c(0, 1),
main = "p-valor test Box-Pierce - \n residuos test1_lluvia", xlab = "lag",
ylab = "p-valor"); abline(h = 0.05, col = "red"); text(1, 0.075, "0.05",
col = "red") # se cumple incorr

resumen1_lluvia = cbind(t(test1_lluvia@teststat), test1_lluvia@cval)
resumen1_lluvia # rechazo que tau3=0

### => LLUVIA ES I(0)

#####
### Causalidad de Granger ###
#####

# grangertest(y ~ x, orden), si rechazo x causa a y

cg = matrix(0, nrow = 12, ncol = 20)
for(i in c("fert", "IVFPBI_agro", "FPI", "p_fert")){
  data[, paste('d', i, sep = '')] = c(NA, diff(data[,i]))
}
vars = c("dfert", "dIVFPBI_agro", "dFPI", "dp_fert", "lluvia_pais")
h = 1; aux = NULL
for(i in vars){
  varaux = vars[vars != i]
  for(j in varaux){
    for(k in 1:12){
      cg[k,h] = round(100*grangertest(data[-1,j] ~ data[-1,i], order = k)[2,4], 2)
    }
    h = h + 1
    aux = c(aux, paste(i, j, sep = "."))
  }
}
}

```

```

rownames(cg) = 1:12; colnames(cg) = aux

#####
### Prueba de Cointegracion de Johansen ###
#####

## funciones
# modificacion en funcion ca.jo: saco la cte en los datos si ecdet es none, trend
source("C:/PASANTIA/VENTAS/cajo2.R")
# modificacion en funcion ca.jo: pongo la cte en coint si ecdet es trend
source("C:/PASANTIA/VENTAS/cajo3.R")

# series en logaritmos
data$logfert = log(data$fert)
data$logpbi = log(data$IVFPBI_agro)

# intervenciones
for(i in c(13, 42, 132, 140)){
  data[, paste('ao', i, sep = "")] = 0; data[i, paste('ao', i, sep = "")] = 1
}

# tendencia lineal
data$tend = 1:nrow(data)

## especificaciones
cajos = nom = noms = list()
casos = c("cteytend.no.restr", "cte.no.restr", "tend.no.restr",
"cteytend.restr", "cte.restr", "tend.restr", "nada")
season = 12 # si considero o no indicatrices estacionales por defecto
dumfijas = c(paste('DFPI', 1:2, sep = '_'),
paste("ao", c(13, 42, 132, 140), sep = "")) # dummies fijas
ks = 2:13 # cant de rezagos en el VAR
z = 1

for(i in ks){
  for(k in casos){

    dumvar = data[, dumfijas]
    if(i == 13 & ("ao13" %in% dumfijas | "ao12" %in% dumfijas)){
      dumvar$ao13 = NULL; dumvar$ao12 = NULL} # sino da sist singular
    if(i == 12 & "ao12" %in% dumfijas){
      dumvar$ao12 = NULL} # sino da sistema singular
    if(k == "cteytend.no.restr"){
      aux = ca.jo(data[,c("logfert", "logpbi")], type = "trace",
      ecdet = "none", K = i, spec = "longrun", dumvar = cbind(tend =
      1:nrow(data), dumvar), season = season)
    }
    if(k == "cte.no.restr"){
      aux = ca.jo(data[,c("logfert", "logpbi")], type = "trace",
      ecdet = "none", K = i, spec = "longrun", dumvar =
      dumvar, season = season)
    }
  }
}

```

```

}
if(k == "tend.no.restr"){
  aux = cajo.scte(data[,c("logfert", "logpbi")], type = "trace",
    ecdet = "none", K = i, spec = "longrun", dumvar =
    cbind(data$tend, dumvar), season = season)
}
if(k == "cteytend.restr"){
  aux = cajo.tendycteR(data[,c("logfert", "logpbi")], type = "trace",
    ecdet = "trend", K = i, spec = "longrun", dumvar =
    dumvar, season = season)
}
if(k == "cte.restr"){
  aux = ca.jo(data[,c("logfert", "logpbi")], type = "trace", ecdet =
    "const", K = i, spec = "longrun", dumvar = dumvar, season = season)
}
if(k == "tend.restr"){
  aux = cajo.scte(data[,c("logfert", "logpbi")], type = "trace", ecdet =
    "trend", K = i, spec = "longrun", dumvar = dumvar, season = season)
}
if(k == "cte.restr.tend.no.restr"){
  aux = ca.jo(data[,c("logfert", "logpbi")], type = "trace", ecdet =
    "const", K = i, spec = "longrun", dumvar = cbind(data$tend,
    dumvar), season = season)
}
if(k == "cte.no.restr.tend.restr"){
  aux = ca.jo(data[,c("logfert", "logpbi")], type = "trace", ecdet =
    "trend", K = i, spec = "longrun", dumvar = dumvar, season = season)
}
if(k == "nada"){
  aux = cajo.scte(data[,c("logfert", "logpbi")], type = "trace", ecdet
    = "none", K = i, spec = "longrun", dumvar = dumvar, season = season)
}
nom[[z]] = paste(i, k); noms[[z]] = c(i, k); cajos[[z]] = aux; z = z + 1
}
}

nom = do.call(c, nom); noms = as.data.frame(do.call(rbind, noms))
names(cajos) = nom; rm(nom)

## ajustes

# otras intervenciones
for(i in c(6, 7, 14, 29, 31, 34, 43, 49, 76, 90, 108, 112, 116, 119, 123, 133,
136, 141, 146, 155)){
  data[, paste('ao', i, sep = "")] = 0; data[i, paste('ao', i, sep = "")] = 1
}

# se descartan modelos 1, 4, 8, 50 y 71 por dar siempre RR

cajos[[2]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',

```

```
ecdet = 'none', K = 2, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao31', 'ao34', 'ao42',
'ao43', 'ao90', 'ao116', 'ao119', 'ao132', 'ao133', 'ao141', 'ao146',
'ao155')], season = 12)
```

```
cajos[[3]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 2, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29',
'ao42', 'ao43', 'ao90', 'ao119', 'ao133', 'ao146')]), season = 12)
```

```
cajos[[5]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'const', K = 2, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao90',
'ao119', 'ao133', 'ao146')], season = 12)
```

```
cajos[[6]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 2, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao90',
'ao119', 'ao133', 'ao146')], season = 12)
```

```
cajos[[7]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 2, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao90',
'ao119', 'ao133', 'ao146')], season = 12)
```

```
cajos[[9]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 3, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep
= '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao49', 'ao90',
'ao119', 'ao133', 'ao146')], season = 12)
```

```
cajos[[10]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 3, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42',
'ao43', 'ao49', 'ao90', 'ao119', 'ao133', 'ao146')]), season = 12)
```

```
cajos[[11]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 3, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao49', 'ao90',
'ao119', 'ao133', 'ao146')], season = 12)
```

```
cajos[[12]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'const', K = 3, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep =
'_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao49', 'ao90', 'ao119',
'ao133', 'ao146')], season = 12)
```

```
cajos[[13]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'trend', K = 3, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep =
'_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao90', 'ao133', 'ao146')],
season = 12)
```

```
cajos[[14]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
```



```

'none', K = 3, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep = '_'),
'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133', 'ao146')],
season = 12)

cajos[[15]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 4, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data), data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao49', 'ao90',
'ao133', 'ao146')]))), season = 12)

cajos[[16]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 4, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep =
'_'), 'ao6', 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao49', 'ao90',
'ao133', 'ao146')]), season = 12)

cajos[[17]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 4, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao6', 'ao7', 'ao13', 'ao14', 'ao29',
'ao42', 'ao43', 'ao49', 'ao90', 'ao133', 'ao146')])), season = 12)

cajos[[18]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 4, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao49', 'ao90',
'ao133', 'ao146')]), season = 12)

cajos[[19]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'const', K = 4, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133', 'ao146')]), season = 12)

cajos[[20]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 4, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao43', 'ao49', 'ao90',
'ao133', 'ao146')]), season = 12)

cajos[[21]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 4, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1, sep = '_'), 'ao6', 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao116', 'ao133', 'ao141', 'ao146')]), season = 12)

cajos[[22]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 5, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42',
'ao49', 'ao90', 'ao133', 'ao146')])), season = 12)

cajos[[23]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 5, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133', 'ao146')]), season = 12)

cajos[[24]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 5, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),

```

```

data[, c(paste('DFPI', 1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42',
'ao49', 'ao90', 'ao133', 'ao146'))], season = 12)

cajos[[25]] = cajo.tendycter(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 5, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2,
sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133',
'ao146'))], season = 12)

cajos[[26]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'const', K = 5, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133', 'ao146'))], season = 12)

cajos[[27]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 5, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133', 'ao146'))], season = 12)

cajos[[28]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 5, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133', 'ao146'))], season = 12)

cajos[[29]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 6, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42',
'ao49', 'ao90', 'ao133'))]), season = 12)

cajos[[30]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 6, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133'))], season = 12)

cajos[[31]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 6, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42',
'ao49', 'ao90', 'ao133'))]), season = 12)

cajos[[32]] = cajo.tendycter(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 6, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1,
sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133',
'ao141'))], season = 12)

cajos[[33]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'const', K = 6, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133'))], season = 12)

cajos[[34]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 6, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2,
sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',

```

```

'ao133']), season = 12)

cajos[[35]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 6, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao7', 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133')], season = 12)

cajos[[36]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 7, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao13', 'ao14', 'ao29', 'ao42',
'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[37]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 7, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133')], season = 12)

cajos[[38]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 7, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao13', 'ao14', 'ao29', 'ao42',
'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[39]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 7, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133')], season = 12)

cajos[[40]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'const', K = 7, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133')], season = 12)

cajos[[41]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 7, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')],
season = 12)

cajos[[42]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 7, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao14', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')],
season = 12)

cajos[[43]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 8, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133')]), season = 12)

cajos[[44]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 8, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao76', 'ao90', 'ao133')],
season = 12)

```

```

cajos[[45]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 8, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49',
'ao90', 'ao133')]), season = 12)

cajos[[46]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 8, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[47]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'const', K = 8, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[48]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 8, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[49]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 8, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[50]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 9, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90',
'ao108', 'ao133')]), season = 12)

cajos[[52]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'none', K = 9, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data),
data[, c(paste('DFPI', 1, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90',
'ao133')]), season = 12)

cajos[[53]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 9, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1,
sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[54]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'const', K = 9, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[55]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'trend', K = 9, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[56]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 9, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[57]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 10, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data), data[, c(paste('DFPI',
1, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

```

```

cajos[[58]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 10, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep = '_'), 'ao13',
'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[59]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 10, spec = 'longrun', dumvar = cbind(trend = 1:nrow(data),
data[, c(paste('DFPI', 1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90',
'ao108', 'ao133')]), season = 12)

cajos[[60]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'trend', K = 10, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[61]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'const',
K = 10, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao108', 'ao133')], season = 12)

cajos[[62]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'trend',
K = 10, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'), 'ao13',
'ao29', 'ao42', 'ao49', 'ao90', 'ao108', 'ao133')], season = 12)

cajos[[63]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 10, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep = '_'), 'ao13',
'ao29', 'ao42', 'ao49', 'ao90', 'ao108', 'ao133')], season = 12)

cajos[[64]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 11, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data), data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao112', 'ao133')]), season = NULL)

cajos[[65]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 11, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep = '_'), 'ao13',
'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[66]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 11, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data), data[, c(paste('DFPI',
1:2, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]), season = 12)

cajos[[67]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 11, spec = 'longrun', dumvar = data[, c(paste('DFPI',
1, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[68]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'const', K = 11, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2,
sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[69]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'trend', K = 11, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep =
'_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[70]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =

```

```

'none', K = 11, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1:2, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[72]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 12, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'), 'ao13',
'ao29', 'ao42', 'ao90', 'ao133')], season = 12)

cajos[[73]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 12, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data), data[,
c(paste('DFPI', 1, sep = '_'), 'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')]),
season = 12)

cajos[[74]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'trend', K = 12, spec = 'longrun', dumvar = data[,c(paste('DFPI', 1, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[75]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'const',
K = 12, spec = 'longrun', dumvar = data[,c(paste('DFPI', 1, sep = '_'), 'ao13',
'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[76]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'trend', K = 12, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'),
'ao29', 'ao42', 'ao13', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[77]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 12, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'),
'ao13', 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[78]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 13, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'), 'ao29',
'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

cajos[[79]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet = 'none',
K = 13, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'), 'ao42',
'ao49', 'ao90', 'ao133')], season = 12)

cajos[[80]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 13, spec = 'longrun', dumvar = cbind(tend = 1:nrow(data), data[,
c(paste('DFPI', 1, sep = '_'), 'ao90', 'ao42', 'ao133')]), season = 12)

cajos[[81]] = cajo.tendycteR(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 13, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep
= '_'), 'ao29', 'ao42', 'ao90', 'ao133')], season = 12)

cajos[[82]] = ca.jo(data[,c('logfert', 'logpbi')], type = 'trace', ecdet =
'const', K = 13, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep =
'_'), 'ao29', 'ao42', 'ao90', 'ao133')], season = 12)

cajos[[83]] = cajo.scte(data[,c('logfert', 'logpbi')], type = 'trace',
ecdet = 'trend', K = 13, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep
= '_'), 'ao29', 'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

```

```

cajos[[84]] = cajo.scte(data[, c('logfert', 'logpbi')], type = 'trace', ecdet =
'none', K = 13, spec = 'longrun', dumvar = data[, c(paste('DFPI', 1, sep = '_'),
'ao42', 'ao49', 'ao90', 'ao133')], season = 12)

## resultados
resultados = list()
for(i in 1:length(cajos)){
  if(!is.complex(cajos[[i]]@teststat[1]) & !is.complex(cajos[[i]]@teststat[2])){
    res = ifelse(cajos[[i]]@teststat[1] > cajos[[i]]@cval[1,2] &
cajos[[i]]@teststat[2] > cajos[[i]]@cval[2,2], 'RR',
ifelse(cajos[[i]]@teststat[1] > cajos[[i]]@cval[1,2] &
cajos[[i]]@teststat[2] < cajos[[i]]@cval[2,2], 'RN',
ifelse(cajos[[i]]@teststat[1] < cajos[[i]]@cval[1,2] &
cajos[[i]]@teststat[2] > cajos[[i]]@cval[2,2], 'NR',
ifelse(cajos[[i]]@teststat[1] < cajos[[i]]@cval[1,2] &
cajos[[i]]@teststat[2] < cajos[[i]]@cval[2,2], 'NN', '0'))))
  } else { res = 'Complejo' }
  resultados[[i]] = data.frame(noms[i,], res)
  names(resultados[[i]]) = c('K', 'caso', 'res')
}
resultados = do.call(rbind, resultados)
resultados$res = as.character(resultados$res)
table(resultados$res)

# completo espec de cada modelo
resultados$nrezDFPI = resultados$inter = NA
for(i in 1:nrow(resultados)){
  a = colnames(cajos[[i]]@dumvar)
  resultados[i, c('nrezDFPI', 'inter')] =
c(length(grep('DFPI', a)), length(grep('ao', a)))
}
resultados = resultados[, c('K', 'caso', 'nrezDFPI', 'inter', 'res')]

# descarto 9 modelos por resultado prueba coint 'RR' y uno por
# resultado complejo quedan 73 modelos NR y uno NN

#####
### Estimacion, diagnostico y seleccion ###
#####

mods = list()
for(i in 1:nrow(resultados)){
  if(resultados[i,'res'] == 'NR'){
    mods[[i]] = cajorls(cajos[[i]], 1)
  } else if(resultados[i,'res'] == 'NN'){
    tipo = ifelse(resultados$caso[i] %in% c('cte.no.restr',
'cte.no.restr.tend.restr'), 'const',
ifelse(resultados$caso[i] %in% c('tend.no.restr',
'cte.restr.tend.no.restr'), 'trend',
ifelse(resultados$caso[i] %in% c('cteytend.no.restr'), 'both',

```

```

    ifelse(resultados$caso[i] %in% c('cteytend.restr', 'cte.restr',
    'tend.restr', 'nada'), 'none'))))
    exo = data.frame(cajos[[i]]@Z1[, colnames(cajos[[i]]@dumvar)])
    mods[[i]] = VAR(cajos[[i]]@Z1[,c('logfert.dl1', 'logpbi.dl1')],
    cajos[[i]]@lag - 1, type = tipo, season = 12, exogen = exo)
  } else mods[[i]] = NULL
}

## p.valor en VECM
summary(mods[[i]]$rlm)

## p.valor en VAR (83)
summary(mods[[i]]$varresult$logfert.dl1)$coefficients
summary(mods[[i]]$varresult$logpbi.dl1)$coefficients

## seleccion: modelos con parametros sign: 17, 21, 32, 44:49, 51:55, 72:77
resultados$valido = 'no'; resultados$valido[c(17, 21, 32, 44:49, 51:55,
72:77)] = 'si'

cand = list(); h = 1
for(i in c(17, 21, 32, 44:49, 51:55, 72:77)){
  cand[[h]] = list(ca.jo = cajos[[i]], vecm = mods[[i]])
  h = h + 1
}
names(cand) = as.character(c(17, 21, 32, 44:49, 51:55, 72:77))

# estimo como VAR
source('C:/PASANTIA/VENTAS/vec2var2.R')
source('C:/PASANTIA/VENTAS/vec2var3.R')

for(i in names(cand)){
  if(i %in% c(17, 21, 45, 48, 49, 52, 55, 72, 73, 76, 77)){
    cand[[i]][[3]] = vec2var.scte(cand[[i]][[1]], r = 1)
  } else if(i %in% c(32, 46, 53, 74)){
    cand[[i]][[3]] = vec2var.tendycteR(cand[[i]][[1]], r = 1)
  } else if(i %in% c(44, 47, 51, 54, 72, 75)){
    cand[[i]][[3]] = vec2var(cand[[i]][[1]], r = 1)
  }
}

# normality.test, si rechazo no hay normalidad
norm = tab.norm = list()
for(i in 1:length(cand)){
  norm[[i]] = normality.test(cand[[i]][[3]], multivariate.only = FALSE)
  tab.norm[[i]] = data.frame(fert.uni = norm[[i]]$jb.uni[[1]]$p.value,
  pbi.uni = norm[[i]]$jb.uni[[2]]$p.value,
  multi = norm[[i]]$jb.mul$JB$p.value,
  sim = norm[[i]]$jb.mul$Skewness$p.value,
  kur = norm[[i]]$jb.mul$Kurtosis$p.value)
}
names(norm) = names(cand)

```



```

tab.norm = do.call(rbind, tab.norm)
rownames(tab.norm) = names(cand)
tab.norm = round(100*tab.norm, 2)
# rechazo normalidad para 51

for(i in names(cand)){
  pdf(paste('C:/PASANTIA/VENTAS/graficosLog/qnorm_cand', i, '.pdf', sep = ''))
  par(mfrow=c(1,2))
  qqnorm(residuals(cand[[i]][[3]])[,1], main = 'Normal \n Q-Q plot logfert')
  qqline(residuals(cand[[i]][[3]])[,1])
  qqnorm(residuals(cand[[i]][[3]])[,2], main = 'Normal \n Q-Q plot logpbi')
  qqline(residuals(cand[[i]][[3]])[,2])
  dev.off()
}

# graficos de los residuos
for(i in names(cand)){
  postscript(paste('C:/PASANTIA/VENTAS/graficosLog/cand', i,
'_fert.eps', sep = ''), horizontal = FALSE)
  h = hist(residuals(cand[[i]][[3]])[,1], plot = FALSE)
  plot(norm[[i]], names = 'resids of logfert',
main.resid = 'Residuals of logfert',
ylim.hist = c(0, max(h$density)*1.25))
  title(main = paste('MODELO', i, '- fert'), cex.main = 1.2
outer = TRUE, line = -1); dev.off()
  postscript(paste('C:/PASANTIA/VENTAS/graficosLog/cand', i,
'_pbi.eps', sep = ''), horizontal = FALSE)
  h = hist(residuals(cand[[i]][[3]])[,2], plot = FALSE)
  plot(norm[[i]], names = 'resids of logpbi',
main.resid = 'Residuals of logpbi',
ylim.hist = c(0, max(h$density)*1.25))
  title(main = paste('MODELO', i, '- PBI'), cex.main = 1.2,
outer = TRUE, line = -1); dev.off()
}

# hay autocorrelacion en los modelos 17, 21, 32, 44, 45, 72:76

# serial.test: si rechazo hay autocorrelacion
serial = list()
for(i in names(cand)){
  serial[[i]] = data.frame(
  PT.asymptotic = serial.test(cand[[i]][[3]], lags.pt = 12,
lags.bg = 12,type = 'PT.asymptotic')$serial$p.value,
  PT.adjusted = serial.test(cand[[i]][[3]], lags.pt = 12,
lags.bg = 12, type = 'PT.adjusted')$serial$p.value,
  BG = serial.test(cand[[i]][[3]], type = 'BG', lags.pt = 12,
lags.bg = 12)$serial$p.value,
  ES = serial.test(cand[[i]][[3]], type = 'ES', lags.pt = 12,
lags.bg = 12)$serial$p.value)
}
serial = do.call(rbind, serial)
serial = round(100*serial, 2)

```

```

rownames(serial) = names(cand)
# rechazo incorrelacion para modelo 21

# arch.test: si rechazo hay heteroscedasticidad
heter = list()
for(i in names(cand)){
  aux = arch.test(cand[[i]][[3]], multivariate.only = FALSE)
  heter[[i]] = data.frame(fert.uni = aux$arch.uni[[1]]$p.value,
    pbi.uni = aux$arch.uni[[2]]$p.value,
    multi = aux$arch.mul$p.value)
}
heter = do.call(rbind, heter)
heter = round(100*heter, 2)
# rechazo H0 para 17, 21 y 72

round(cbind(tab.norm, serial, heter), 1)

## preseleccion: 32, 44:49, 52:55, 73:77

# ACF y PACF
for(i in names(cand)){
  postscript(paste('C:/PASANTIA/VENTAS/graficosLog/ACFyPACF',
    i, '.eps', sep = ''), horizontal = FALSE)
  par(mfrow = c(2, 2))
  acf(cand[[i]][[3]]$resid[,1], main = 'ACF logfert')
  pacf(cand[[i]][[3]]$resid[,1], main = 'PACF logfert')
  acf(cand[[i]][[3]]$resid[,2], main = 'ACF logpbi')
  pacf(cand[[i]][[3]]$resid[,2], main = 'PACF logpbi')
  title(main = paste('MODELO', i), cex.main = 1.2, outer = TRUE, line = -1)
  dev.off()
}

# CCF
for(i in names(cand)){
  postscript(paste('C:/PASANTIA/VENTAS/graficosLog/CCF',
    i, '.eps', sep = ''), horizontal = FALSE)
  par(mfrow = c(2, 1))
  ccf(cand[[i]][[3]]$resid[,1], cand[[i]][[3]]$resid[,2],
    main = 'CCF logfert - logpbi')
  ccf(cand[[i]][[3]]$resid[,2], cand[[i]][[3]]$resid[,1],
    main = 'CCF logpbi - logfert')
  title(main = paste('MODELO', i), cex.main = 1.2, outer = TRUE, line = -1)
  dev.off()
}

#####
### Prediccion dentro de la muestra ###
#####

# funciones
source('C:/PASANTIA/VENTAS/predict1.R'); source('C:/PASANTIA/VENTAS/predict2.R')

```

```

source('C:/PASANTIA/VENTAS/predict3.R'); source('C:/PASANTIA/VENTAS/internal.R')

n = 12
data1 = data[1:(nrow(data)-n),]; data2 = data[nrow(data) - (n-1):0,]

pred = list()
FPIreal = TRUE # si FALSE uso pred de FPI
predFPI = # predicciones de FPI

# for(i in names(cand)){
for(i in as.character(47:49)){

  pred[[i]] = list()
  x = data1[,colnames(cand[[i]][[1]]@x)]
  ecdet = cand[[i]][[1]]@ecdet
  K = cand[[i]][[1]]@lag
  if(FPIreal){
    pred1.dumvar = data1[, colnames(cand[[i]][[1]]@dumvar)]
    pred1.dumvar = pred1.dumvar[, colSums(pred1.dumvar) > 0]
    pred2.dumvar = data2[, names(pred1.dumvar)]
  } else{
    pred1.dumvar = data1[, colnames(cand[[i]][[1]]@dumvar)]
    pred1.dumvar = pred1.dumvar[, colSums(pred.dumvar) > 0 &
!grepl('FPI', names(pred1.dumvar))]
    pred1.dumvar = cbind(pred1.dumvar, predFPI[grepl('FPI',
colnames(cand[[i]][[1]]@dumvar), value = 'TRUE')])
  }

  if(i %in% c(17, 21, 45, 48, 49, 52, 55, 72, 73, 76, 77)){
    pred[[i]][[1]] = cajo.scte(x, type = 'trace', ecdet = ecdet,
K = K, spec = 'longrun', dumvar = pred1.dumvar, season = 12)
    if(!is.complex(pred[[i]][[1]]@teststat)){
      pred[[i]][[2]] = cajorls(pred[[i]][[1]])
      pred[[i]][[3]] = vec2var.scte(pred[[i]][[1]], r = 1)
      pred[[i]][[4]] = predict.scte(pred[[i]][[3]], n.ahead = n,
dumvar = pred2.dumvar)
    }
  }

  if(i %in% c(32, 46, 53, 74)){
    pred[[i]][[1]] = cajo.tendycteR(x, type = 'trace', ecdet = ecdet,
K = K, spec = 'longrun', dumvar = pred1.dumvar, season = 12)
    if(!is.complex(pred[[i]][[1]]@teststat)){
      pred[[i]][[2]] = cajorls(pred[[i]][[1]])
      pred[[i]][[3]] = vec2var.tendycteR(pred[[i]][[1]], r = 1)
      pred[[i]][[4]] = predict.tendycteR(pred[[i]][[3]], n.ahead = n,
dumvar = pred2.dumvar)
    }
  }

  if(i %in% c(44, 47, 51, 54, 72, 75)){

```

```

pred[[i]][[1]] = ca.jo(x, type = 'trace', ecdet = ecdet,
K = K, spec = 'longrun', dumvar = pred1.dumvar, season = 12)
if(!is.complex(pred[[i]][[1]]@teststat)){
  pred[[i]][[2]] = cajorls(pred[[i]][[1]])
  pred[[i]][[3]] = vec2var(pred[[i]][[1]], r = 1)
  pred[[i]][[4]] = predict(pred[[i]][[3]], n.ahead = n,
dumvar = pred2.dumvar)
}
}
}

# errores relativos y ECM
ERfert = ERpbi = ECMfert = ECMpbi = list()
for(i in 1:length(pred)){
  if(length(pred[[i]]) > 1){
    ERfert[[i]] = (data2$fert - exp(pred[[i]][[4]]$fcst$logfert[,1]))/data2$fert
    ERpbi[[i]] = (data2$IVFPBI_agro - exp(pred[[i]][[4]]$fcst$logpbi[,1]))/
data2$IVFPBI_agro
    ECMfert[[i]] = round(mean((data2$fert - exp(pred[[i]][[4]]$fcst$logfert[,1])
)^2), 2)
    ECMpbi[[i]] = round(mean((data2$IVFPBI_agro - exp(pred[[i]][[4]]$fcst$
logpbi[,1]))^2), 2)
  } else{ ERfert[[i]] = ERpbi[[i]] = rep('complejo', n)
    ECMfert[[i]] = ECMpbi[[i]] = 'complejo'
  }
}
}
ERfert = as.data.frame.matrix(do.call(cbind, ERfert), stringsAsFactors = FALSE)
ERpbi = as.data.frame.matrix(do.call(cbind, ERpbi), stringsAsFactors = FALSE)
colnames(ERfert) = colnames(ERpbi) = names(pred)
rownames(ERfert) = rownames(ERpbi) = 1:n
for(i in which(ERfert[1,] != 'complejo')){
  ERfert[,i] = round(100*as.numeric(ERfert[,i]), 2)
  ERpbi[,i] = round(100*as.numeric(ERpbi[,i]), 2)
}
}
ECMfert = do.call(rbind, ECMfert)
ECMpbi = do.call(rbind, ECMpbi)

```