



## Modelos Econométricos y de Redes Neuronales

para predecir la Oferta Maderera en México:

**ARIMA vs NAR y ARMAX vs NARX**

Tesis que presenta  
**Claudia Estela Ortiz de Dios**  
para obtener el grado de:  
Maestría en Ciencias  
(Matemáticas Aplicadas e Industriales)

Asesores

Dr. Joaquín Delgado Fernández  
Dr. Julio Goicoechea Moreno

### Jurado Dictaminador

Presidente	Dr. Alberto Castillo Morales
Secretario	Dr. Joaquín Delgado Fernández
Vocal	Dr. Francisco Venegas Martínez
Vocal	Dr. Julio Fernando Goicoechea Moreno

Diciembre, 2012

## Resumen

Se compara el resultado de las predicciones de la oferta de madera en el mercado mexicano, proponiendo un modelo univariado y un multivariado con dos técnicas distintas de predicción.

La econometría fue la primera técnica utilizada mediante la construcción de un modelo ARIMA para el caso univariado y un ARMAX para el multivariado. La segunda técnica de predicción fueron las Redes Neuronales Artificiales (RNA) utilizando un modelo NAR y un NARX respectivamente. Se contrastó la exactitud mediante el mínimo MSE (cuadrado medio del error) en la predicción para determinar qué modelo y técnica obtuvo mejor desempeño.

Se encontró que, en ambos casos no existe diferencia estadísticamente significativa en la predicción entre ambas técnicas de modelación, tanto para el caso univariado como en el multivariado.

Considero que las RNA, ofrecen a la econometría una alternativa de predicción, utilizando series temporales que son las que atañen a este trabajo.

### **Dedicatoria:**

A mis padres: Violeta de Dios Ricardez y Roberto Ortiz Guerrero. †

A mi hermano: José Eduardo Ortiz de Dios.

A mi esposo: Raúl José Santos Bernard.

### **Agradecimientos:**

A mis asesores, Dr. Julio F. Goicoechea Moreno y Dr. Joaquín Delgado Fernández, por su apoyo, orientación, consejos y paciencia. Todos ellos brindados durante el desarrollo de mis estudios y elaboración de la tesis.

A mis sinodales, Dr. Alberto Castillo Morales y Dr. Francisco Venegas Martínez, por todo el apoyo, recomendaciones y consejos brindados durante la revisión de éste trabajo.

A mis profesores de la UAM por compartir sus conocimientos, tiempo y afecto. En especial agradezco al Dr. Felipe Peredo Rodríguez por su orientación. Dr. José Luis Estrada López por su ayuda, amistad y preocupación. Dra. Nora Garro Bordonaro y Dra. Alenka Guzmán por su ayuda y apoyo. Dr. Mario Medina por su paciencia y afecto. Dra. Rosalinda Arriaga Navarrete por su amistad y apoyo moral. Dr. Mario Pineda Ruelas por siempre darme ánimo. A la Dra. María Luisa Sandoval Solís por su apoyo y orientación.

A la MSc. María Iseo González Christen y al Lic. José A. Atenco Sangrador por todo el apoyo, orientación, buena disposición y paciencia casi infinita, muchas gracias.

A mi familia, Violeta de Dios y Eduardo Ortiz por su apoyo y paciencia. A mi esposo por su cariño y apoyo incondicional. A mi primo Rodrigo de Dios por su desvelo en la revisión. A mi tía Irma Urbina por siempre darme ánimo. A mi cuñada Dora Santos, por sus consejos y apoyo. A mis suegros Raúl Santos y Roberta Bernard por su apoyo y ánimo.

A mis compañeros y amigos, Heri Oscar Landa Díaz, Juan Carlos Vázquez Almaráz, Marco Rojo Gutiérrez, Eduardo Moctezuma Navarro y su esposa Erika Sánchez Puebla, Sr. David Martínez y su esposa Sra. Teresa Valdés, Paulo Sergio García Méndez, Guadalupe García Salazar, Gerardo del Muro, David Tonatiu Moreno González, Ximena Salinas, Margarita Macín, Ricardo Porrúa, Enrique Minor, por su amistad, tiempo, consejos, ayuda y experiencias compartidas muchas gracias.

Un agradecimiento especial, a Marisela González Zayas, por su ánimo, amistad y apoyo incondicional gracias infinitas. Igualmente para Cyntia Paredes, Ana Lozada y Ana Elena Martínez por no perder la paciencia, y siempre darme su muy amable apoyo y atención gracias.

# Índice general

Índice general	1
<b>1. Introducción</b>	<b>4</b>
<b>2. Modelos <i>ARIMA</i> y <i>ARMAX</i></b>	<b>9</b>
2.1. Proceso autorregresivo ( <i>AR</i> ) . . . . .	9
2.2. Proceso de media móvil ( <i>MA</i> ) . . . . .	10
2.3. Proceso autorregresivo y de media móvil ( <i>ARMA</i> ) . . . . .	11
2.4. Proceso autorregresivo integrado de media móvil ( <i>ARIMA</i> ) . . . . .	11
2.5. Condiciones y procedimiento de modelado <i>ARIMA</i> . . . . .	12
2.5.1. Condiciones . . . . .	12
2.5.2. Procedimiento . . . . .	13
2.6. Proceso autoregresivo y media móvil con variables exógenas ( <i>ARMAX</i> ) . . . . .	15
2.7. Condiciones y procedimiento de modelado <i>ARMAX</i> . . . . .	16
2.7.1. Condiciones . . . . .	16
2.7.2. Procedimiento . . . . .	17
<b>3. Construcción, especificación, modelado y resultados de los modelos <i>ARIMA</i> y <i>ARMAX</i></b>	<b>19</b>
3.1. Modelos <i>ARIMA</i> y <i>ARMAX</i> . . . . .	19
3.1.1. Selección de las series . . . . .	19
3.1.2. Construcción de la base de datos . . . . .	30
3.1.3. Procedimiento de pruebas de verificación . . . . .	32

3.2.	Modelación . . . . .	41
3.2.1.	Modelo ARIMA univariado . . . . .	41
3.2.2.	Modelo <i>ARMAX</i> multivariado . . . . .	70
<b>4.</b>	<b>Introducción a los Modelos de Redes Neuronales Artificiales (<i>RNA</i>)</b>	<b>78</b>
4.1.	Fundamentos y conceptos básicos . . . . .	78
4.2.	Tipos de redes neuronales . . . . .	80
4.2.1.	El preceptrón . . . . .	81
4.2.2.	El preceptrón multicapa . . . . .	83
4.2.3.	Retropropagación . . . . .	84
4.2.4.	Algoritmo de aprendizaje . . . . .	85
4.3.	Aproximación por redes neuronales . . . . .	86
4.3.1.	Modelo NAR caso univariado . . . . .	87
4.3.2.	Modelo NARX caso multivariado . . . . .	89
<b>5.</b>	<b>Construcción, especificación, modelación y resultados de los modelos <i>RNA</i></b>	<b>91</b>
5.1.	Construcción de un modelo <i>RNA</i> . . . . .	91
5.2.	Especificación de un modelo <i>RNA</i> . . . . .	92
5.3.	Modelación . . . . .	93
5.3.1.	Modelo NAR univariado . . . . .	93
5.3.2.	Modelo NARX multivariado . . . . .	99
<b>6.</b>	<b>Comparación de resultados y conclusiones</b>	<b>104</b>
6.1.	Comparación de resultados . . . . .	104
6.1.1.	Modelos univariados . . . . .	104
6.1.2.	Modelos multivariados . . . . .	106
6.2.	Conclusiones y perspectivas . . . . .	108
6.2.1.	Caso univariado . . . . .	109
6.2.2.	Caso multivariado . . . . .	109
6.2.3.	Conclusión general . . . . .	110
6.2.4.	Perspectivas . . . . .	111

<b>Anexos</b>	<b>112</b>
Anexo 1: Base de datos final . . . . .	113
Anexo 2: Scripts en R de automatización . . . . .	117
Script general: general PAR.R . . . . .	117
Script complementario: klzPAR.R . . . . .	119
Anexo 3: Script en PHP de generación de combinaciones para E-Views . . . . .	126
Anexo 4: Script de E-Views de automatización . . . . .	135
Anexo 5: Script de MATLAB modelo RNA univariado . . . . .	146
Anexo 6: Script de MATLAB modelo RNA multivariado . . . . .	148
<b>Bibliografía</b>	<b>151</b>

# Capítulo 1

## Introducción

Una de las muchas inquietudes del ser humano es predecir el futuro; la ciencia económica no es la excepción.

Para tomar las mejores decisiones sobre el futuro, los agentes económicos necesitan contar con la mejor información posible.

Un sector económico que considero importante es el primario, es decir aquel que provee de materias primas a diversas industrias, para producir bienes con mayor valor agregado.

En su mayoría este sector primario se compone de empresas que explotan recursos naturales renovables, los cuales son escasos y necesitan de una explotación racional y eficiente.

Un ejemplo de este sector, es el productor de madera, la cual se utiliza en diversas industrias para ser transformada en bienes intermedios, que a su vez son empleados como materias primas para una transformación adicional o como bienes finales que llegan al consumidor de bienes completamente transformados. Como ejemplo de estas industrias tenemos a las fábricas de muebles, las de productos para la construcción, la industria del papel, etc.

Para este trabajo éste es el sector de interés, dado que la explotación de la madera no se ha llevado a cabo de forma completamente racional, y a pesar de existir políticas de reforestación, éstas no tienen el impacto que se necesita. La reforestación es más lenta que la deforestación.

De aquí la importancia de conocer el desempeño del sector maderero a través de las expectativas de su oferta.

Una manera de saberlo es analizando su evolución a lo largo del tiempo, y suponiendo que se seguirá comportando de manera similar, para obtener la información esperada utilizando

pronósticos.

En economía se analiza, interpreta y modela la información económica cuantitativa para predecir el comportamiento de la misma y con base en ello tomar las mejores decisiones posibles.

La forma más general de establecer relaciones entre los diversos agentes que intervienen en la economía, es decir, explicar y predecir su comportamiento es utilizando la econometría. Definiendo a ésta como la parte de la economía que se encarga de realizar todo tipo de análisis, principalmente cuantitativos de las relaciones económicas entre los agentes, sin embargo dependiendo del tipo de modelo también se pueden realizar análisis cualitativos.

Dentro de la econometría existen diferentes formas de modelar la información dependiendo del tipo de estudio y del tipo de información con el que se cuenta. Los tres tipos principales de información para modelar que se utilizan son:

1. Información de series de tiempo: En estos modelos se identifican y aíslan componentes relacionados con el tiempo que afectan los valores observados; estos componentes son la tendencia, los ciclos, variaciones estacionales, variaciones irregulares, y son excelente fuente de información para elaborar predicciones. Un ejemplo de un modelo con información de series de tiempo sería explicar y predecir el precio del jitomate, analizando su comportamiento desde el año 1960 al año 2000.

Para el caso de las de series de tiempo, algunos tipos de modelos para realizar los análisis son, por ejemplo:

- a) Modelos *AR*
- b) Modelos *MA*
- c) Modelos *ARMA*
- d) Modelos *ARIMA*
- e) Modelos *ARMAX*
- f) Modelos *VAR*
- g) Modelos *ARCH*
- h) Modelos *GARCH*,

*i)* etc.

Más adelante explicaré en qué consisten cada uno de los primeros cinco modelos mencionados, ya que sólo utilicé los procesos *ARIMA* y *ARMAX*.

2. Información de corte transversal: Los modelos de corte transversal analizan un momento específico en el tiempo, características específicas de un grupo de agentes económicos. Como ejemplo podríamos mencionar el análisis del Producto Interno Bruto, de los países de América Latina en 1985.
3. Información de datos panel: Este tipo de modelos se constuyen a través de la conjunción de los dos modelos anteriores. Como ejemplo podríamos tomar el ejemplo del de corte transversal pero estudiando el período 1960-2000.

Sin embargo, en la última década se ha empezado a trabajar con las redes neuronales artificiales (*RNA*) en economía, como alternativa a la modelización tradicional de series de tiempo, para realizar predicciones.

Éstas, dada su estructura y funcionalidad, obtienen buenos resultados en la modelación y predicción de series de tiempo. Por ello que considero interesante analizar si las *RNA* ofrecen mejores resultados que los modelos econométricos de series de tiempo al predecir la oferta de madera en México.

El objetivo principal de este trabajo, es determinar cuál metodología puede ser mejor para realizar las predicciones de la oferta maderera en México, si alguno de los modelos econométricos de series de tiempo o las redes neuronales artificiales.

Lo anterior se realizará utilizando series de tiempo reales cubriendo el período Ene 1994 - Dic 2008, considerando que el volumen producido de madera está relacionado con el consumo de este bien como materia prima en diversas ramas de la producción en nuestro país.

Para efecto de estimación se utiliza el período 1994-2007, y para comprobación de la predicción el 2008.

Las hipótesis del análisis son:

$H_0$  : *Las RNA ofrecen mejores predicciones que la econometría*

$H_A$  : *Las RNA no ofrecen mejores predicciones que la econometría*

Para la elaboración de esta investigación utilicé:

1. En el caso del modelado econométrico, un modelo univariado de vectores autorregresivos integrados y media móvil (*ARIMA*), y un modelo multivariado de vectores autorregresivos de medias móviles con variables exógenas (*ARMAX*), utilizando datos reales obtenidos del Banco de México y del INEGI. Esta información se presenta en series de tiempo, las cuales conformarán las variables de los modelos para posteriormente realizar las predicciones.

En esta parte del trabajo utilicé los programas *R* e *E-Views* para realizar la modelación econométrica.

2. Para las *RNA* se realizaron dos modelos de predicción: uno univariado y otro multivariado, modelando mediante un sistema de retroalimentación, el cual entrenó la red hasta lograr un grado de error mínimo, para posteriormente realizar la comprobación del entrenamiento de la red y finalizar con las predicciones. En esta sección utilizaré el módulo de redes neuronales del software *MATLAB*.

Este trabajo resulta relevante debido a que, si bien el uso de las *RNA* es relativamente nuevo en economía a nivel mundial ; en el caso de México existen muy pocos estudios realizados con esta metodología en comparación con los modelos econométricos. Otra parte relevante es la experimentación con datos reales y analizar el grado de ajuste entre ambos métodos para la predicción.

Los seis capítulos que comprenden este trabajo son:

1. Introducción.
2. Definición de los modelos econométricos *ARIMA* y *ARMAX*, donde se explicará cómo se conforman y su funcionamiento.

3. Construcción, especificación, modelación y resultados de los modelos *ARIMA* y *ARMAX*.
4. Introducción a las redes neuronales composición y funcionamiento.
5. Construcción, especificación, modelación y resultados de los modelos *RNA*
6. Comparación de resultados y conclusiones.

## Capítulo 2

# Modelos *ARIMA* y *ARMAX*

En éste capítulo explico la composición de los modelos *AR*, *MA* y *ARMA* debido a que fungen como base de los modelos *ARIMA* y *ARMAX*.

Describo también condiciones y procedimientos de los modelos *ARIMA* y *ARMAX* que son los de interés para este estudio.

En economía, para realizar estimaciones y predicciones con series de tiempo existen varios métodos, en una parte de éste trabajo se utilizó un modelo *ARIMA* generalmente conocido como la metodología de Box-Jenkins<sup>1</sup>.

Un modelo *ARIMA* es una variación de un modelo *ARMA* aplicado a series integradas de orden “*d*”. Esto quiere decir que las series han sido necesariamente diferenciadas “*d*” veces para conseguir la eliminación de la tendencia.

Por su parte, un modelo *ARMA* es la conjunción de los modelos *AR* (Autorregresivos) y *MA* (Medias Móviles o Moving Average). A continuación explicaré cómo se conforma un modelo *ARIMA* y su funcionamiento.

### 2.1. Proceso autorregresivo (*AR*)

Sean  $Y_t = X_t$  la variable dependiente del modelo, y  $X_t$  con retardo,  $Z_t$  con retardo las variables explicativas del mismo en el período  $t$ . Si  $X_t$  se modelara como:

---

<sup>1</sup>G. P. E. Box y G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., Holden Day, San Francisco, 1978.

$$(X_t - \delta) = \alpha_1(X_{t-1} - \phi) + \beta_1(Z_{t-1} - \gamma) + u_t \quad (2.1)$$

Donde:

$\alpha$  = Proporción de  $X$  en el periodo  $(t - 1)$

$\beta$  = Proporción de  $Z$  en el periodo  $(t - 1)$

$\delta$  = Media de  $X$

$\phi$  = Media de  $X$  con retardo

$\gamma$  = Media de  $Z$

$u_t$  = Término de error aleatorio no correlacionado con media cero y varianza constante  $\sigma^2$ , es decir “ ruido blanco” .

Se dice entonces que  $X_t$  sigue un proceso estocástico autorregresivo de primer orden o  $AR(1)$ .

Sin embargo, si se considera el siguiente modelo:

$$(X_t - \delta) = \alpha_1(X_{t-1} - \phi) + \alpha_2(X_{t-2} - \phi) + \dots + \alpha_p(X_{t-p} - \phi) + \beta_1(Z_{t-1} - \gamma) + \quad (2.2)$$

$$\beta_2(Z_{t-2} - \gamma) + \dots + \beta_p(Z_{t-p} - \gamma) + u_t \quad (2.3)$$

En este caso se dice que es un proceso estocástico autorregresivo de orden “ $p$ ”, o  $AR(p)$ .<sup>2</sup>

## 2.2. Proceso de media móvil ( $MA$ )

Dado que un proceso  $AR$  no es el único mecanismo que pudo generar la variable dependiente, supongamos ahora que la variable  $Y_t$  se modela de la siguiente manera:

$$Y_t = \mu + \eta_0 u_t + \eta_1 u_{t-1} \quad (2.4)$$

Donde:

$\mu$  = Constante

$u_t$  = Al igual que en la ecuación (1.1), es el término de error estocástico ruido blanco.

---

<sup>2</sup>Damodar N. Gujarati. Econometría. Quinta Edición, McGraw-Hill 2010. pp 775-776

En este modelo, la variable dependiente es igual a una constante más un promedio móvil de los términos de error presente y pasado, de tal manera que  $Y_t$  sigue un proceso de media móvil de primer orden o  $MA(1)$ .

De forma más general,

$$Y_t = \mu + \eta_0 u_t + \eta_1 u_{t-1} + \eta_2 u_{t-2} + \dots + \eta_q u_{t-q} \quad (2.5)$$

es un proceso  $MA(q)$ .

Resumiendo: Un proceso de media móvil simplemente es una combinación lineal de términos de error ruido blanco.<sup>3</sup>

### 2.3. Proceso autorregresivo y de media móvil ( $ARMA$ )

En la modelación cabe la posibilidad de que la variable dependiente cuente con características de  $AR$  y  $MA$  al mismo tiempo, y por ende sea un modelo  $ARMA$ , de tal manera que como ejemplo un proceso  $ARMA(1, 1)$  se pueda escribir como:

$$Y_t = \theta + \alpha_1 X_{t-1} + \eta_0 u_t + \eta_1 u_{t-1} \quad (2.6)$$

Donde  $\theta$  representa un término constante.

Entonces, dado que este ejemplo presenta un término autorregresivo y uno de media móvil, se dice que es un proceso  $ARMA(1, 1)$ .<sup>4</sup>

### 2.4. Proceso autorregresivo integrado de media móvil ( $ARIMA$ )

Las series de tiempo económicas que no son estacionarias, es decir que su media, varianza y autocorrelaciones no son constantes, porque dependen del tiempo, se consideran variables integradas.

Una variable integrada de orden 1, denotada como  $I(1)$  implica que sus primeras diferencias son  $I(0)$ , esto quiere decir estacionaria, por lo que su media, varianza y autocorrelaciones son

---

<sup>3</sup>Damodar N. Gujarati. Econometría. Quinta Edición, McGraw-Hill 2010. pp 776

<sup>4</sup>Damodar N. Gujarati. Econometría. Quinta Edición, McGraw-Hill 2010. pp 776

invariantes en el tiempo. Asimismo, si una serie de tiempo es  $I(2)$ , su segunda diferencia es  $I(0)$ . En general, una serie de tiempo es  $I(d)$ , y después de diferenciarla “ $d$ ” veces se obtiene una serie  $I(0)$  estacionaria.

Entonces, si se debe diferenciar una serie de tiempo “ $d$ ” veces para hacerla estacionaria y luego aplicar a ésta un modelo  $ARMA(p, q)$ , se dice que la serie de tiempo original es una serie de tiempo autorregresiva integrada de media móvil  $ARIMA(p, d, q)$ , donde “ $p$ ” denota el número de términos autorregresivos, “ $d$ ” el número de veces que la serie debe ser diferenciada para hacerse estacionaria y “ $q$ ” el número de términos de media móvil.<sup>5</sup>

El punto importante de ésta metodología es que, para utilizarla, se deben tener series de tiempo estacionarias. La razón obedece a que:

*“El objetivo de Box-Jenkins<sup>6</sup> es identificar y estimar un modelo estadístico que se pueda interpretar como generador de la información muestral.*

*Entonces, si el modelo se va a utilizar para realizar predicciones, debe suponer que sus características son constantes a través del tiempo, y particularmente en períodos de tiempo futuro. Así, la simple razón para requerir información estacionaria es que cualquier modelo inferido a partir de ésta información pueda interpretarse como estacionario o estable, proporcionando entonces una base válida para la predicción”<sup>7</sup>.*

## 2.5. Condiciones y procedimiento de modelado $ARIMA$

### 2.5.1. Condiciones

Para que esta metodología sea efectiva es necesario que las series con las que se realice el modelo, cumplan dos condiciones:

1. Las series deben ser estacionarias, esto es:
  - a) Las series deben *NO* tener raíz unitaria. y además,
  - b) tener varianza constante.

---

<sup>5</sup>Damodar N. Gujarati. Econometría. Quinta Edición, McGraw-Hill 2010. pp 776-777

<sup>6</sup>G. P. E. Box y G. M. Jenkins, *Time Series Analysis: Forecasting and Control, revised ed., Holden Day, San Francisco, 1978.*

<sup>7</sup>Michael Pokorny, *An Introduction to Econometrics*, Basil Blackwell, New York, 1987, p.343

2. Los residuos de la mejor regresión modelada con términos *AR* y *MA* deben ser ruido blanco<sup>8</sup>, es decir cumplir con las condiciones de Gauss-Markov:
  - a) Tener media cero.
  - b) Tener varianza constante.
  - c) Tener covarianza igual a cero.

### 2.5.2. Procedimiento

Los pasos para efectuar la metodología de modelado *ARIMA* son los siguientes:

1. Graficar series para ver su comportamiento.

La graficación de las series nos provee la primera parte de la información para observar la evolución y comportamiento de la serie a lo largo del tiempo. Este análisis visual nos provee una primera idea sobre la estacionariedad, o falta de ella, en una serie. Sin embargo, otras pruebas formales deben ser aplicadas para asegurar que la serie es estacionaria. Dentro de los puntos a observar en la gráfica están la detección de tendencia, estacionalidad y/o fluctuaciones irregulares. Todas estas características influyen sobre la estacionariedad.

Generalmente, las series NO estacionarias muestran un comportamiento errático (como una caminata aleatoria) alrededor de una constante y/o tendencia, mientras que las series estacionarias muestran un comportamiento fluctuante (es decir, oscilan entre un rango de valores) alrededor de una constante y/o tendencia.

El correlograma de la serie también nos provee de una idea visual de si la serie es estacionaria o no, pues si las barras no se salen de las bandas implica que la covarianza es igual a cero ( $Cov = 0$ ), requisito indispensable para ser estacionario.

2. Verificar su estacionariedad.

Para verificar que las tres condiciones fundamentales se cumplan, es necesario aplicar algunas pruebas.

---

<sup>8</sup>Un proceso de ruido blanco representa una variable que: oscila en torno a una media constante, con una volatilidad constante y cuyo pasado no contiene información útil para predecir valores futuros.

a) Prueba de White<sup>9</sup>:

Esta prueba es para determinar si la varianza no es constante y muestra si la serie tiene problemas de heteroscedasticidad. En este mismo paso se determina, mediante una regresión auxiliar de la serie analizada con una constante y su tendencia, si la media es constante. Ejemplificando, la función a regresar sería:

$$\text{Serie} = f(c, @trend(0)) \quad (2.7)$$

donde  $c$  es una constante y  $@trend(0)$  es una variable que  $E - Views$  llama *variable ficticia de tendencia*. Si en el resultado de la regresión la probabilidad (*prob.*) de la variable  $@trend(0) \geq 0,05$ , se considera que la media de la serie es constante, y cumple con uno de los requisitos indispensables para que la variable sea estacionaria.

b) Prueba de raíz unitaria:

Esta prueba es para determinar si una serie es  $I(0)$  integrada de orden cero. Si esta condición no se cumple implica que hay existencia de raíz unitaria y es muy probable que la serie no sea estacionaria.

Existen cuatro casos en el análisis de series de tiempo referentes a estacionariedad o falta de ella:

- 1) Si tiene raíz unitaria  $\implies$  No es estacionaria
  - 2) Si la varianza NO es constante  $\implies$  No es estacionaria
  - 3) Si NO tiene raíz unitaria y su varianza NO es constante  $\implies$  No es estacionaria
  - 4) Si NO tiene raíz unitaria y su varianza SI es constante  $\implies$  Es estacionaria
3. Si las series no son estacionarias, se requiere realizar las transformaciones necesarias para conseguir la estacionariedad. Esto se logra sacando las diferencias  $(Y_t - Y_{t-1})$  necesarias hasta alcanzarla.
4. Una vez siendo estacionarias, se debe correr la regresión del modelo, generar la serie de

---

<sup>9</sup>H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity". *Econometrica*, vol. 48, 1980, pp. 817-818.

los residuos  $E1$ , utilizando el comando  $GENR E1 = resid$  y analizar su correlograma. Esto es para obtener los términos  $AR$  y  $MA$  significativos y generar una tabla de combinaciones involucrando todos los términos  $AR$  y  $MA$  que se obtuvieron.

Los elementos que resultan significativos, son aquellos que salen del área de las bandas punteadas del correlograma de autocorrelación ( $AC$ ) y del de autocorrelación parcial ( $PAC$ ).

El ancho de las bandas se determina por:  $ABanda = \pm \frac{2}{\sqrt{\#Observaciones}}$

La información que proveen los correlogramas una vez teniendo las series estacionarias, nos permite identificar los elementos  $AR$  y  $MA$  del modelo.

Del correlograma de correlación parcial ( $PAC$ ) obtenemos los elementos  $AR$ , del de autocorrelación ( $AC$ ) los elementos  $MA$ .

5. Calcular las regresiones necesarias con las combinaciones de  $AR$  y  $MA$  obtenidas, seleccionar la que tenga el menor valor de Schwarz y coeficientes significativos para considerarlo el mejor modelo.
6. Verificar que los residuos de la mejor regresión sean ruido blanco.
7. Escribir la ecuación teórica  $ARIMA(p, d, q)$
8. Escribir la ecuación estimada.
9. Realizar la predicción.
10. Graficar la predicción junto a la serie original, en el intervalo pronosticado, y observar la cercanía de los pronósticos con los valores reales.
11. Calcular el cuadrado medio del error ( $MSE$ ) y sacar conclusiones.

## 2.6. Proceso autoregresivo y media móvil con variables exógenas ( $ARMAX$ )

Un modelo  $ARMAX$  es una variante del  $ARMA$  en la cual se considera que el término de error de la función tiene su propia estructura de retardos.

A diferencia de los modelos AR, MA, ARMA y ARIMA, los cuales son univariados, y su proceso se realiza sobre la función original, un ARMAX es multivariado y su proceso se realiza sobre los residuos de la función, de tal forma que:

Sea  $Y_t$  la variable dependiente del modelo y  $X_t, Z_t$  las variables explicativas del mismo.

Si  $Y_t$  se modelara como:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + \mu_t \quad (2.8)$$

Si:

$$\mu_t = e_t \quad (2.9)$$

Los residuos se modelarían con un modelo *ARMA* de la siguiente forma:

$$e_t = \rho_i e_{t-i} + \dots + \theta_j e_{t-j} + \dots + \theta_k e_{t-k} + E_t \quad (2.10)$$

Donde los términos que contienen  $\rho$  corresponden a los términos *AR*, los que contienen  $\theta$  a los *MA*, y  $E_t$  es ruido blanco.

Se calcula el pronóstico de  $e_t$ , al cual denominaremos  $EF_t$  para finalmente retomar la primera regresión (ecuación 2.8) e integrar el término de error pronosticado, y calcular el pronóstico de la variable dependiente mediante:

$$Y_{t+i} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+i} + \hat{\beta}_3 Z_{t+i} + EF_{t+i} \quad (2.11)$$

Donde:

$\hat{\beta}_1, \hat{\beta}_2$  y  $\hat{\beta}_3$  son los coeficientes estimados de la regresión original.<sup>10</sup>

## 2.7. Condiciones y procedimiento de modelado *ARMAX*

### 2.7.1. Condiciones

1. Las series que componen tanto a la variable dependiente como las independientes, deben ser **NO** estacionarias.

---

<sup>10</sup>William H. Greene. Análisis Econométrico. Tercera Edición, Prentice Hall 2000. pp 694-703, y notas de clase.

2. La ecuación de regresión  $Y_t$  debe realizarse en niveles, es decir no llevar diferenciación alguna.
3. Los residuos de la regresión de  $E_t$  deben ser ruido blanco, es decir:
  - a) Tener covarianza igual a cero.
  - b) Tener media constante.
  - c) Tener varianza constante.

### 2.7.2. Procedimiento

Los pasos para realizar la modelación *ARMAX* son los siguientes:

1. Seleccionar las variables dependiente y explicativas que compondrán el modelo. Esta selección puede depender del modelador. Para este trabajo propongo:
  - a) Correr una regresión (modelo 1º) con todas las variables seleccionadas
  - b) Correr una segunda regresión (modelo 2º) solo con las variables no significativas del modelo 1º.
  - c) Correr una tercera regresión (modelo 3º) solo con las variables significativas de los modelos 1º y 2º.
2. Generar la variable  $E_t$  es decir generar una variable con los residuos de la regresión del modelo 3º.
3. Analizar los residuos ( $E_t$ ) mediante:
  - a) Prueba de raíz unitaria: Si al 95 % de confianza la serie no tiene raíz unitaria decimos que las series están cointegradas.
  - b) Revisar los correlogramas de residuos para seleccionar los términos *AR* y *MA* que se utilizarán para realizar el modelo *ARMA* de los residuos  $E_t$
4. Correr las regresiones de  $E_t$  con todas las combinaciones posibles de los términos *AR* y *MA* encontrados en el análisis del correlograma.

5. Seleccionar como mejor modelo el que cumpla las siguientes condiciones:
  - a) Menor valor de Schwarz y/o Akaike.
  - b) Coeficientes significativos.
  - c) Residuos de regresión ruido blanco.
6. Generar la serie  $EF_t$  con las estimaciones de las medias de  $E_t$  predichas.
7. Regresar a la ecuación original estimada del modelo definitivo añadiendo la variable  $EF_t$  y calcular el pronóstico de la variable dependiente.
8. Graficar la variable dependiente original con la estimada para analizar qué tan precisa resultó
9. Sacar conclusiones.

A modo de conclusión, en este capítulo se explicó de forma detallada la conformación, requisitos y procedimiento a seguir de los modelos ARIMA y ARMAX, los cuales se desarrollarán paso a paso en el siguiente capítulo a partir del punto 3.1.3.

## Capítulo 3

# Construcción, especificación, modelado y resultados de los modelos *ARIMA* y *ARMAX*

### 3.1. Modelos *ARIMA* y *ARMAX*

En este capítulo se muestra la forma en que se realizó la selección de variables, la definición del modelo propuesto de oferta de madera y su modelación econométrica.

#### 3.1.1. Selección de las series

La fuente principal de información para seleccionar las series utilizadas en este trabajo fueron las “Matrices de Insumo Producto” de los años 1960<sup>1</sup> y 1970<sup>2</sup>. Se utilizan sólo estos años debido a que no pude consultar ni de forma impresa ni electrónica la matriz de 1980, y la de 1990 no se había publicado todavía con cifras definitivas al momento de iniciar este trabajo.

Por otra parte, también se utilizó la información de la “Encuesta Industrial Mensual” del

---

<sup>1</sup> *La Estructura industrial de México en 1960*. Banco de México, S. A. Investigaciones industriales. Oficina de economía industrial.

Departamento de investigaciones industriales.  
Oficina editorial 1967.

<sup>2</sup> *Matriz de insumo producto de México año 1970*. Tomo 3 Actividades primarias y servicios. Secretaría de programación y presupuesto.

INEGI para la segunda parte de la selección de series.

Los bienes producidos en una economía pueden ser de dos tipos, según el destino de demanda cubierta.

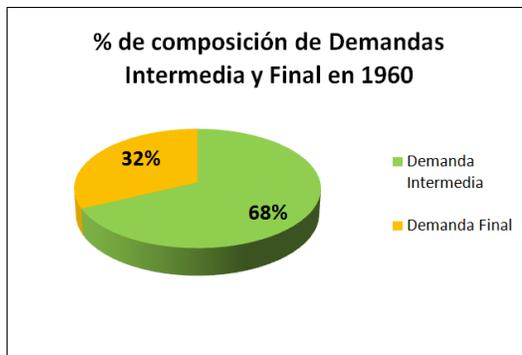
El primero es el consumo intermedio o demanda intermedia, y el segundo el consumo final o demanda final.

El consumo intermedio se refiere a aquellos bienes producidos que sirven como materia prima para otras industrias. Como ejemplo del sector tratado en este trabajo, la madera aserrada sirve como materia prima para la industria del mueble.

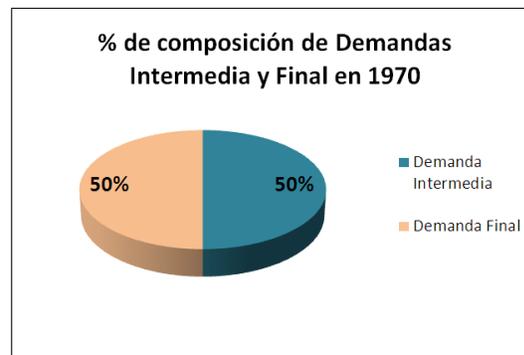
Por otra parte, el consumo final sería la compra de madera aserrada al usuario final sin llevar una transformación adicional para su uso. Es decir, la madera se compra y se usa tal cual fue adquirida y no sufrirá transformación alguna.

La Matriz de Insumo Producto es un reporte que indica de dónde proviene y hacia dónde se dirige la producción de cada sector involucrado en la economía. En otras palabras, a qué compradores va a parar la producción de cada sector.

Los siguientes gráficos nos muestran la composición de la demanda del sector maderero en 1960 y 1970.



Porcentaje de la composición de las demandas intermedia y final de la producción maderera en 1960



Porcentaje de la composición de las demandas intermedia y final de la producción maderera en 1970

Entre 1960 y 1970 se observa una variación del 18% en la composición de las demandas. Suponiendo que la proporción entre demandas intermedia y final no varía de forma considerable en los siguientes periodos, y promediando los tipos de demanda de 1960 y 1970, el sector silvícola

envía el 59% de su producción al abasto de la demanda intermedia y el 41% a la demanda final.

A continuación se hace un análisis de la composición de las demandas intermedia y final.

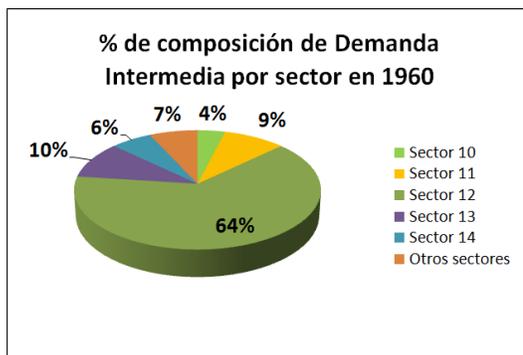
### **Demanda Intermedia**

En la información económica que se genera, ésta se separa en clasificaciones según el tipo de actividad a la que se dedica. Cada tipo de actividad se divide en sectores o ramas, y esta información a su vez se subdivide en clases y las clases en subclases.

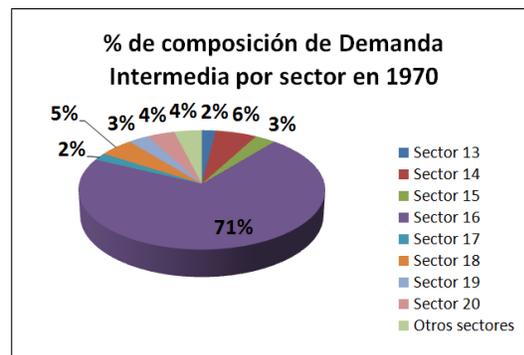
En este trabajo analicé una sección del primer sector que es la actividad agropecuaria, en su tercer rama que es la producción silvícola, considerando únicamente la producción de productos maderables. Sin embargo, dado que el punto de partida es la matriz de insumo producto, se analizó información que involucra a otros sectores.

Debo aclarar que las clasificaciones sectoriales, identificadas con un número, han sufrido modificaciones a lo largo del tiempo, por lo que un sector puede tener un nombre numérico en una matriz, y en otra posterior tener otro o subdividirse en varios.

Esta aclaración sirve para comprender con mayor facilidad los gráficos que a continuación se presentan y nos muestran la composición de la producción del sector maderero, dirigido a cubrir la demanda intermedia.



Porcentaje de la composición de la demanda intermedia por sector productivo en 1960 (ver tabla página 24)



Porcentaje de la composición de la demanda intermedia por sector productivo en 1970 (ver tabla página 24)

Considerando el promedio de la información que nos arrojan los gráficos presentados, y

seleccionando los porcentajes más significativos, se deduce en resumen lo que la tabla muestra a continuación.

Nombre del sector	Año	Identificador del sector	Porcentaje promedio de demanda
Madera y corcho	1960	16	68%
	1970	29 y 30	
Papel y cartón	1960	17	7.5%
	1970	31	
Otras industrias textiles	1960	14	9%
	1970	25	
Molienda de nixtamal y productos de maíz	1970	14	6%
Resto de los sectores	1960	10, 21 y otros	9.50%
	1970	13, 38, 40 y otros	
<b>TOTAL=</b>			<b>100%</b>

Porcentaje promedio de la composición de la demanda intermedia del sector silvícola 1960-1970

De tal forma que, analizando la información de la tabla anterior, los sectores o ramas que se llevan la mayor participación de la producción del sector maderero son la industria de la madera y el corcho, así como la industria del papel y cartón que son los sectores que considero importantes, por utilizar la madera como materia prima principal de transformación, y por tanto fueron los que consideré para calcular los modelos.

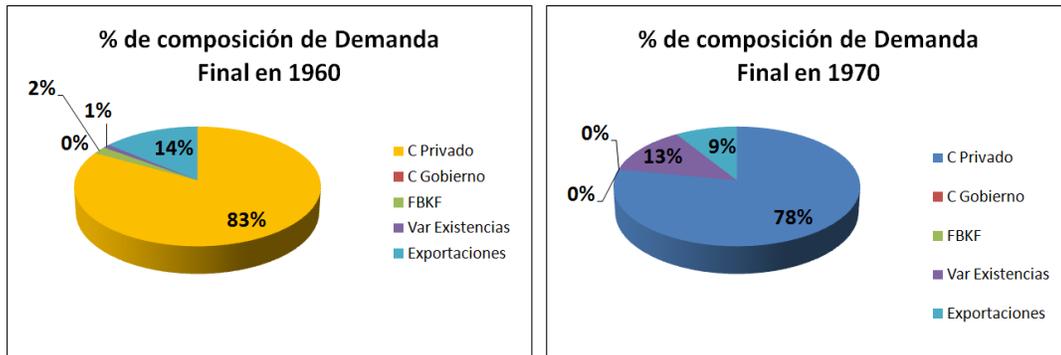
### **Demanda final**

Para el análisis de la demanda final de madera, veremos cuál es el porcentaje de la composición de ésta en cada uno de los períodos analizados.

La demanda final de la producción puede dirigirse a:

1. Consumo privado.
2. Consumo de gobierno.
3. Formación bruta de capital fijo.
4. Variación de existencias.

## 5. Exportaciones.



Porcentajes de la composición de la demanda final de madera en 1960

Porcentajes de la composición de la demanda final de madera en 1970

En ambas gráficas, se observa que en promedio el 80.5 % de la demanda final es para el consumo privado. El 11.5 % son exportaciones madereras, la variación de existencias se incrementa en 12 % de 1960 a 1970, el consumo del gobierno es de 0 % y la formación bruta de capital fijo es prácticamente nula en ambos casos.

Este trabajo se realizó con base en la información correspondiente a la demanda intermedia, y los sectores industriales que se consideraron fueron:

- Industria de la Madera y corcho (Clase III según la clasificación del INEGI).
- Industria del Papel y Cartón (Clase IV según la clasificación del INEGI).

La razón de lo anterior obedece a que dichas industrias utilizan a la madera y algunos de sus derivados como insumo principal para la producción de bienes finales que van al consumo.

Una vez analizando la dirección global de la producción de madera, analicé hacia qué subsectores de las industrias se dirige el insumo maderero, analizando la información de la Encuesta Industrial Mensual proporcionada por el INEGI.

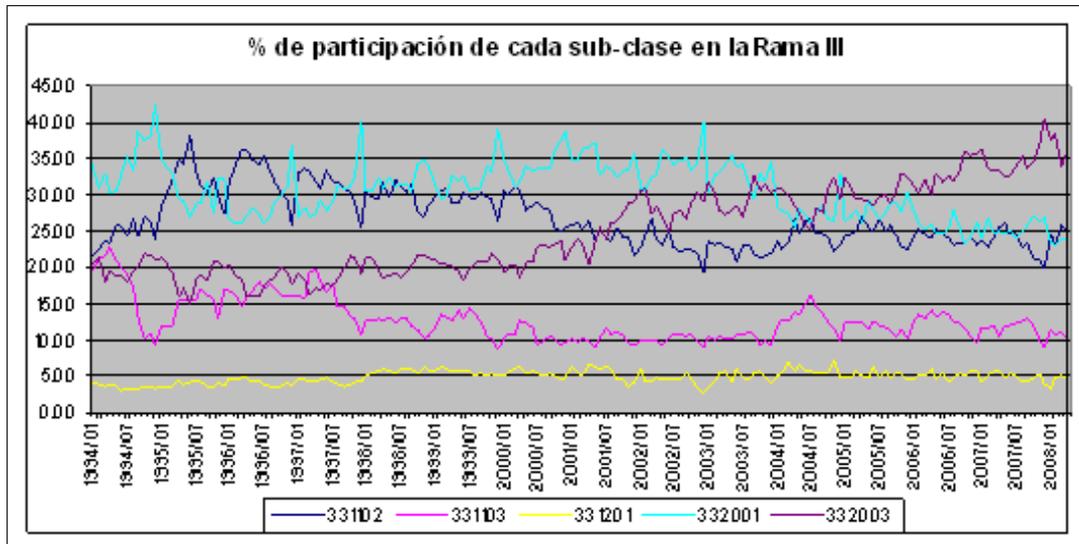
**Clase III: Industria de la madera y productos de madera** Analizando la información de la Encuesta Industrial Mensual (EIM), vemos que la clase III (Industria de la madera y productos de madera) se compone por 5 sub-clases en el siguiente cuadro:

<b>III Industria de la madera y productos de madera</b>	
<b>331102</b>	Fabricación de triplay, tableros aglomerados y de fibra de madera
<b>331103</b>	Fabricación de productos de madera para la construcción
<b>331201</b>	Fabricación de envases de madera
<b>332001</b>	Fabricación y reparación de muebles principalmente de madera
<b>332003</b>	Fabricación de colchones

Subclases que comprenden a la industria de la madera y sus productos.

Cada sub-clase detalla el tipo de productos que manufactura en volumen y valor.

Dado que esta información aparece en precios corrientes, se pueden comparar las sub-clases para determinar cuáles son las que producen un mayor valor en el sector. Esto lo observaremos en el gráfico a continuación:



Participación productiva de las subclases de la Rama III 1994-2008

De ello se puede resumir el orden de importancia de las subclases por su participación en la siguiente tabla:

Lugar obtenido por participación en rama III	Sub-Clase	% promedio de participación en Rama III
1	332001	30.41
2	331102	26.60
3	332003	24.92
4	331103	12.65
5	331201	4.84

Orden decreciente de participación y porcentaje promedio de la misma de las subclases de la Rama III 1994-2008

De la información obtenida de las gráficas anteriores, consideraremos para una primera aproximación del modelo econométrico, a la clase 331102 (en volumen producido) como variable dependiente, y a las clases 331103 , 332001, 332003 y la misma dependiente 331102 unas de las independientes (utilizando los índices de precios de cada una).

Esto debido a que la clase: 301201 envases de madera, tiene una participación mínima y por ello la dejaré fuera del modelo.

Para el caso del modelo ARIMA, al ser univariado, sólo consideraré una variable para el modelo y éste se definiría como:

$$VOLTRIP = f(AR, MA)$$

donde:

VOLTRIP= Es la cantidad de madera producida por la subclase 331102, fabricación de triplay, tableros aglomerados y de fibra de madera en miles de metros cuadrados.

En el caso del modelo multivariado ARMAX, una primera aproximación del modelo sería:

$$VOLTRIP = f(VALTRIPDEF, VALMUEBDEF, VALCONSTDEF, VALCOLCHDEF, IMWDEF, CETE28DEF)$$

donde:

VOLTRIP= Es la cantidad producida por la subclase 331102, fabricación de triplay, tableros aglomerados y de fibra de madera en miles de metros cuadrados.

VALTRIPDEF= Es el valor de la producción de la subclase 331102, fabricación de triplay,

tableros aglomerados y fibra de madera en miles de metros cuadrados.

VALMUEBDEF= Es el valor de la producción de la subclase 332001, fabricación y reparación de muebles principalmente de madera.

VALCONDEF= Es índice de precios de la subclase 331103, fabricación de productos de madera para la construcción.

VALCOLCHDEF= Es índice de precios de la subclase 331103, fabricación de productos de madera para la construcción.

IMWDEF = Es el índice de masa salarial del sector maderero.

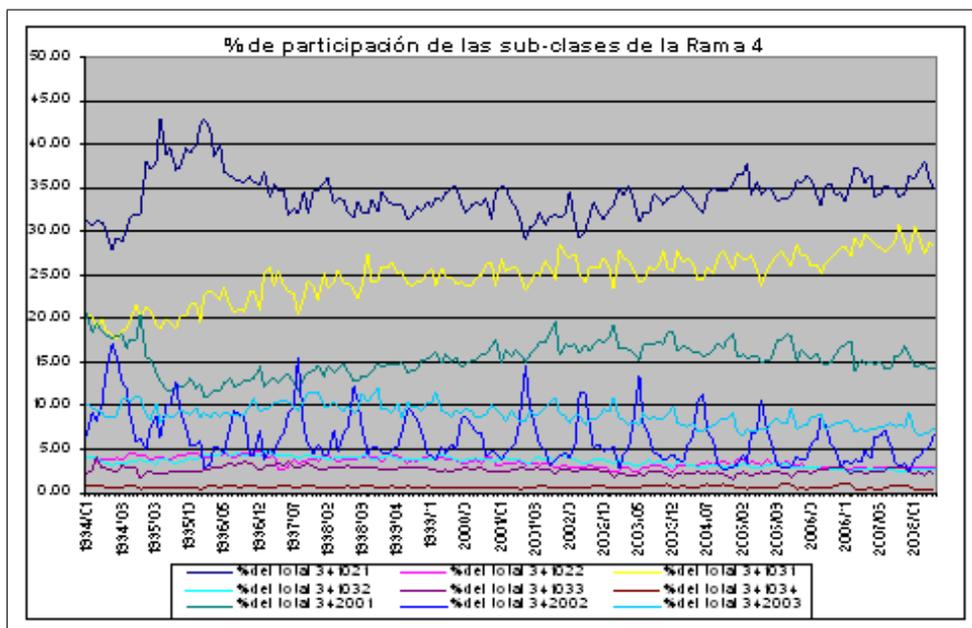
CETE28DEF = Tasa de interés pasiva.

**Clase IV: Industria del papel** La industria del papel se compone de 9 subclases, como se muestra en el siguiente cuadro:

<b>IV Industria del papel</b>
<b>341021</b> Fabricación de celulosa y papel
<b>341022</b> Fabricación de cartón y cartoncillo
<b>341031</b> Fabricación de envases de cartón
<b>341032</b> Fabricación de envases de papel
<b>341033</b> Fabricación de productos de papelería
<b>341034</b> Fabricación de otros productos de papel, cartón y pasta de celulosa, no mencionados anteriormente
<b>342001</b> Edición de periódicos y revistas
<b>342002</b> Edición de libros y similares
<b>342003</b> Impresión y encuadernación

Subclases que componen a la industria del papel.

De la misma manera que analicé la Clase III, los porcentajes de participación promedio de cada subclase en la industria los observamos en la siguiente gráfica:



Participación productiva de las subclases de la Rama IV 1994-2008

De igual manera que en el caso de la Rama III, el cuadro de a continuación muestra en orden descendente la importancia que ocupa cada una de las subclases que componen la Rama IV.

Lugar obtenido por participación en rama IV	Sub-Clase	% de participación en Rama IV
1	341021	34.02
2	341031	24.65
3	342001	15.29
4	342003	8.90
5	342002	6.28
6	341032	3.55
7	341022	3.49
8	341033	2.59
9	341034	0.66

Orden decreciente de participación y porcentaje promedio de la misma de las subclases de la Rama IV 1994-2008

Analizando la tabla anterior podemos concluir que las primeras cinco clases son las de mayor importancia. Sin embargo, la séptima que es la 341022 es la fabricación de cartón y cartoncillo, que si bien no tiene una participación tan alta, es proveedora de insumos a otras sub-clases al igual que la 341021, que es la fabricación de celulosa y papel.

Con lo anterior puedo proponer finalmente un modelo más completo (no necesariamente el definitivo) de la función de oferta de madera, el cual quedaría como a continuación se especifica. Menciono que no es el definitivo dado que puede resultar sobreespecificado el modelo al momento de iniciar las regresiones del mismo en la modelación *ARMAX*:

$$VOLTRIP = f(VALTRIPDEF, VALMUEBDEF, VALCONSTDEF, VALCOLCHDEF, \\ VALPAPDEF, VALCARTDEF, VALENCARTDEF, VALPERDEF, \\ VALEDLIBDEF, VALIMPDEF, IMWDEF, CETE28DEF)$$

Donde:

*VOLTRIP* = Es la cantidad producida por la subclase 331102, fabricación de triplay, tableros aglomerados y de fibra de madera.

*VALTRIPDEF* = Es el valor de la producción de la subclase 331102, fabricación de triplay, tableros aglomerados y fibra de madera en miles de metros cuadrados.

*VALMUEBDEF* = Valor de la producción de la subclase 332001, fabricación y reparación de muebles principalmente de madera.

*VALCONSTDEF* = Valor de la producción de la subclase 331103, fabricación de productos de madera para la construcción.

*VALCOLCHDEF* = Valor de la producción de la subclase 332003 fabricación de colchones.

*VALPAPDEF* = Valor de la producción de la subclase 341021, fabricación de celulosa y papel.

*VALCARTDEF* = Valor de la producción de la subclase 341022, fabricación de cartón y cartoncillo.

*VALENCARTDEF* = Valor de la producción de la subclase 341031, fabricación de envases de cartón.

*VALPERDEF* = Valor de la producción de la subclase 342001, edición de periódicos y revistas.

VALEDLIBDEF = Valor de la producción de la subclase 342002, edición de libros y similares.

VALIMPDEF = Valor de la producción de la subclase 342003, impresión y encuadernación.

IMWDEF = Índice de masa salarial del sector.

CETE28DEF = Tasa de interés pasiva.

La propuesta funcional busca entonces analizar y explicar la relación existente entre los valores de las producciones de las subclases, y el volumen producido del triplay como una función de tipo oferta-demanda.

En todas la series utilicé logaritmos para hacer interpretaciones de impacto porcentual de los coeficientes respecto de la variable dependiente, excepto en la variable CETE28DEF a la que no se puede aplicar debido a cifras negativas al momento de deflactar.

**Modelación econométrica** En economía, los modelos de oferta y demanda, postulan relaciones determinísticas precisas.

Las variables dependientes e independientes están identificadas. Una forma funcional está especificada también, y en la mayoría de los casos se especifica, al menos, una afirmación cualitativa acerca de los efectos que tienen lugar cuando cambian las variables independientes en el modelo.

Dado que un modelo es una simplificación de la realidad, éste incluirá rasgos sobresalientes de la relación de interés, pero no tomando en cuenta todas las influencias que podrían estar presentes.

Naturalmente se puede diferir acerca de hasta qué punto estas influencias pueden considerarse importantes.

El proceso de análisis econométrico parte de la especificación de una relación teórica. Inicialmente procedemos con la hipótesis optimista de que podemos obtener medidas precisas sobre todas las variables de nuestro modelo correctamente especificado. Si las condiciones ideales se cumplen en cada paso, el análisis posterior será, probablemente rutinario. Desafortunadamente eso raramente ocurre.

En este trabajo se formula un modelo econométrico para determinar la oferta de la madera.

A partir de éste modelo se busca experimentar qué metodología es más precisa para pronosticar, si la econométrica o las redes neuronales.

Dado que los datos a emplear son series de tiempo, y se buscará realizar un pronóstico con ambos métodos, para el caso econométrico se utilizará la metodología Box-Jenkins de modelos *ARIMA* (Proceso autorregresivo integrados de medias móviles) y *ARMAX* (Proceso autorregresivo y medias móviles con variables exógenas).

Estas metodologías permiten realizar modelos de pronóstico económico con una considerable exactitud.

Sin embargo, la aptitud de las *RNA* para buscar y reconocer patrones pueden ofrecer una alternativa para pronosticar, esperando obtener un mejor resultado en los pronósticos con *RNA*, que con los modelos *ARIMA* y *ARMAX*.

### 3.1.2. Construcción de la base de datos

Las series de información se obtuvieron del INEGI y del Banco de México para construir la base de datos definitiva que se utilizó en este trabajo. Esta base inicia con 29 series originales, que al realizar las transformaciones correspondientes se obtuvo:

Una variable dependiente, 12 variables independientes o explicativas, con 180 observaciones que comprenden el período Enero 1994 – diciembre 2008 en frecuencia mensual. El análisis no se presenta más actualizado debido a que INEGI cambió la metodología de medición de las variables y éstas ya no son compatibles con el método utilizado.

En las siguientes tablas muestro los nombres de las series originales que se utilizan en este trabajo:

SERIES DE INEGI POR CLASES PRODUCTIVAS											
Clasificación y Nombre	331102 Fabricación de triplay, tableros aglomerados y de fibra de madera	331102 Fabricación de triplay, tableros aglomerados y de fibra de madera	331103 Fabricación de productos de madera para la construcción	332001 Fabricación y reparación de muebles principalmente de madera	332003 Fabricación de colchones	341021 Fabricación de celulosa y papel	341022 Fabricación de cartón y cartoncillo	341031 Fabricación de envases de cartón	342001 Edición de periódicos y revistas	342002 Edición de libros y similares	342003 Impresión y encuadernación
Descripción	Volumen total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/	Valor total de la clasep1/
Unidad de medida	(Miles de metros cuadrados)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)	(Miles de pesos a precios corrientes)

Series de INEGI, obtenidas de la Encuesta Industrial Mensual

SERIES BANXICO													
Clasificación	SL2538	SP9234	SP9236	SP9237	SP9238	SP9239	SP9241	SP9243	SP9244	SP9908	SF4830	SP1	SP4
Nombre	Salario Mínimo General	Indices de Precios Productor y de Comercio Exterior 331103	Indices de Precios Productor y de Comercio Exterior 332001	Indices de Precios Productor y de Comercio Exterior 332003	Indices de Precios Productor y de Comercio Exterior 341021	Indices de Precios Productor y de Comercio Exterior 341022	Indices de Precios Productor y de Comercio Exterior 341031	Indices de Precios Productor y de Comercio Exterior 342001	Indices de Precios Productor y de Comercio Exterior 342002	Indices de Precios Productor y de Comercio Exterior 342003	SF4830 Cetes a 28 días	IPC Por objeto del gasto Nacional, Índice General	IPP Mercancías y servicios finales, (Por destino y origen de la producción), Mercancías finales excluyendo petróleo
Unidad de medida	(Pesos por día)	Índice, base Dic-03	Tasa de rendimiento promedio ponderado, en por ciento anual	Base 2Q Jun 2002	Índice, base Dic-03								

Series de BANXICO, utilizadas algunas para deflactar y otras como variables en el modelo econométrico

Las series de tiempo finales que se utilizaron en el trabajo fueron deflactadas para quitar los efectos de inflación.

La base de datos final transformada se encuentra en el Anexo 1 de este trabajo.

Las series obtenidas fueron:

El volumen en metros cuadrados (variable dependiente), así como el valor en precios corrientes del resto de las series. Y el índice de la masa salarial del sector maderero; para la tasa de interés se tomaron los Cetes a 28 días.

Los índices de precios al productor (IPP) de cada una de las clases que constituyen las series explicativas, y el índice de precios al productor (sin petróleo) para deflactar los Cetes, están en base Diciembre-2003. El (IMS) índice de masa salarial se encontraba en base enero de 2004 y se pasó a base Diciembre-2003 mediante las siguientes fórmulas:

$$IMS_{mes} = \left( \frac{Obs_{mes}}{Obs_{Dic2003Original}} \right) * 100 \quad (3.1)$$

Los índices de precios se utilizarán para deflactar las series y obtenerlas en pesos constantes para quitar el efecto de la inflación.

La fórmula general para deflactar las series de la categoría de producción es la siguiente:

$$Obs_{mes} = \frac{Obs_{precios\ corrientes,\ mes}}{IPP_{de\ su\ clase\ correspondiente,\ mes}} \quad (3.2)$$

Para el caso de la deflactación de la tasa Cetes a 28 días:

$$Cete28_{Def, t} = \left( \frac{1 + \left( \frac{\left( \frac{Obs\ Cete28_t}{12} \right)}{100} \right)}{1 + \frac{ippsp_t - ippsp_{t-1}}{ipp_{t-1}}} \right) - 1 \quad (3.3)$$

donde  $ippsp_t$  = Índice de precios al productor sin petróleo en el mes  $t$ .

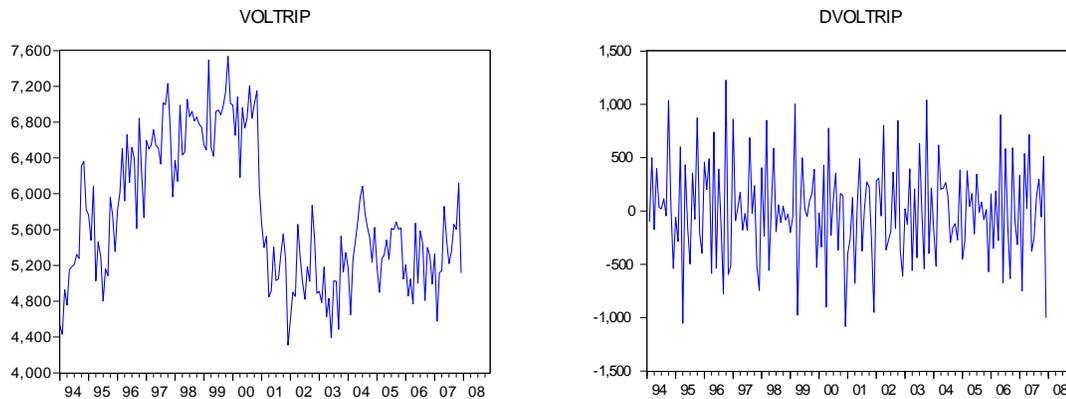
### 3.1.3. Procedimiento de pruebas de verificación

En esta sección iré mostrando paso a paso, como se indicó en la descripción del procedimiento de las modelaciones *ARIMA* y *ARMAX*, las pruebas y procesos realizados para efectuar los modelos.

Los análisis se aplicarán a la serie *VOLTRIP* y eventualmente serán aplicados a las series de los residuos de las regresiones de los modelos obtenidos, con la finalidad de determinar si cumplen las especificaciones necesarias para modelar y para ser el mejor modelo.

#### Pruebas de ruido blanco y estacionariedad

**Gráficas de las series** El primer paso de la modelación consiste en graficar la serie que interviene en este modelo, el prefijo *D* en el nombre de la serie implica que ésta ha sido diferenciada una vez.



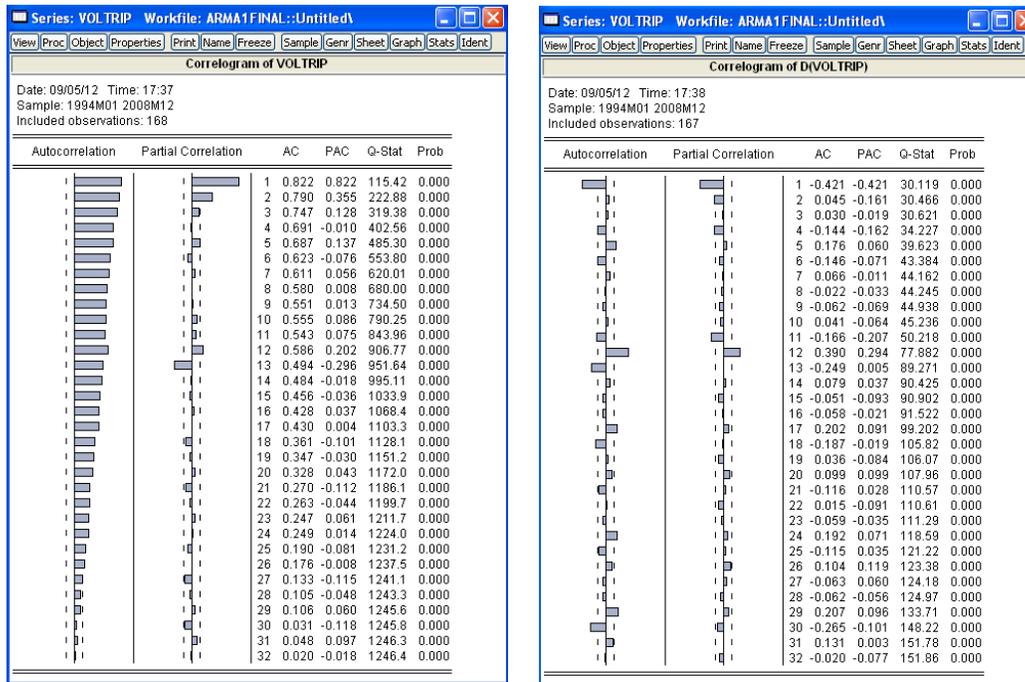
Comportamiento NO estacionario

Comportamiento posiblemente estacionario

A simple vista parece ser que la serie no es estacionaria en su forma original. Por ello,

primero se deberán hacer las pruebas necesarias para asegurar su no estacionariedad y en su caso, realizar las transformaciones convenientes para convertirlas en estacionarias.

**Correlograma** Otra prueba visual es el análisis del correlograma. Una regla sencilla de observación sería: Los valores del correlograma de una serie no estacionaria, descienden muy suavemente hacia el cero, mientras que en el caso de las estacionarias la caída es exponencial. Como ejemplo se muestra la misma serie sin estacionariedad y con estacionariedad.



Comportamiento NO estacionario con la serie en niveles

Comportamiento posiblemente estacionario

Así, de una manera gráfica se puede deducir en qué orden diferencial una serie es estacionaria, pues si la mayor parte de las observaciones no salen de las bandas implica que la covarianza es  $Cov = 0$ ; sin embargo, hay que comprobarlo con el resto de las pruebas formales.

En esta sección se verán las tres pruebas que se aplican a las series de tiempo para determinar su estacionariedad y ruido blanco mediante: la Prueba de Media Constante, la Prueba de White y la Prueba de Raíz Unitaria.

**Prueba de White**<sup>3</sup> Esta prueba se realiza en dos partes: la primera es determinar si la media de la serie es constante, la segunda determina si la varianza es constante.

Para ver cómo funciona, consideremos el siguiente modelo (la generalización al modelo con  $k$  variables es directa):

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (3.4)$$

Para realizar la prueba de White se procede como a continuación se indica:

1. Se estima el modelo anterior obteniendo los residuales  $\hat{u}_i$ .
2. Se estima la regresión auxiliar:

$$\hat{u}_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i \quad (3.5)$$

y obtener  $R^2$  de esta regresión auxiliar<sup>4</sup>.

3. Bajo el supuesto de la inexistencia de heteroscedasticidad, puede demostrarse que el tamaño de la muestra ( $n$ ) multiplicado por el  $R^2$  obtenido de la regresión auxiliar, asintóticamente sigue la distribución  $Ji^2$ , con grados de libertad igual al número de regresores (excluyendo el término constante si lo hay) en la regresión auxiliar.

Es decir:

$$n \cdot R^2_{asy} \widetilde{X}_{\# \text{ regresores}}^2 \quad (3.6)$$

donde los grados de libertad igualan a los definidos anteriormente.

El criterio de aceptación práctico de la prueba de White se da, partiendo de la información obtenida de  $E - Views$ :

---

<sup>3</sup>H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity". *Econometrica*, vol. 48, 1980, pp. 817-818.

<sup>4</sup>El supuesto de que la varianza del error  $u_i$ ,  $\sigma_i^2$  se relaciona funcionalmente con los regresores, con sus valores al cuadrado y con sus productos cruzados, está implícito en este procedimiento. Si todos los coeficientes parciales de pendiente de esta regresión son simultáneamente iguales a cero, entonces la varianza del error es homoscedástica e igual a la constante  $\alpha_1$ . (Damodar N. Gujarati. *Econometría*. Quinta Edición, McGraw-Hill 2010. Pie de página 25, pp 387)

Se corre una regresión de mínimos cuadrados de la serie a analizar, con una constante y su tendencia ( $y = \alpha_1 + \alpha_2 t + u_t$ ). Del reporte de la regresión, se obtiene el  $\alpha^*$  de “t” esto es, la probabilidad del coeficiente  $\alpha_2$  como se muestra en la siguiente salida de *E – Views*:

Dependent Variable: VOLTRIP				
Method: Least Squares				
Date: 09/05/12 Time: 17:43				
Sample: 1994M01 2007M12				
Included observations: 168				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5828.651	58.21156	100.1287	0.000000
@TREND(0)	-5.766751	1.167936	-4.937559	0.000000
R-squared	0.128057	Mean dependent var	5762.333	
Adjusted R-squared	0.122805	S.D. dependent var	783.8561	
S.E. of regression	734.1495	Akaike info criterion	16.04714	
Sum squared resid	89469927	Schwarz criterion	16.08433	
Log likelihood	-1345.959	Hannan-Quinn criter.	16.06223	
F-statistic	24.37949	Durbin-Watson stat	0.387929	
Prob(F-statistic)	0.000002			

Media NO constante

Dependent Variable: DVOLTRIP				
Method: Least Squares				
Date: 09/05/12 Time: 17:49				
Sample (adjusted): 1994M02 2007M12				
Included observations: 167 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.407041	36.53234	0.230126	0.818300
@TREND(0)	-0.406176	0.735368	-0.552343	0.581500
R-squared	0.001846	Mean dependent var	3.532934	
Adjusted R-squared	-0.004204	S.D. dependent var	457.1622	
S.E. of regression	458.1221	Akaike info criterion	15.10405	
Sum squared resid	34629512	Schwarz criterion	15.14139	
Log likelihood	-1259.188	Hannan-Quinn criter.	15.11921	
F-statistic	0.305083	Durbin-Watson stat	2.817751	
Prob(F-statistic)	0.581461			

Media constante

Una vez realizada la regresión, se sigue el siguiente criterio para determinar la constancia de la media:

1. Si  $\alpha^* < 0.05 \implies$  que el coeficiente  $\alpha_2$  es significativo  $\therefore$  la media de  $Y_t$  no es constante (Cuadro Rojo o sombreado de la izquierda).
2. Si  $\alpha^* > 0.05 \implies$  que el coeficiente  $\alpha_2$  no es significativo  $\therefore$  la media de  $Y_t$  sí es constante (Cuadro Verde o sombreado de la derecha).

Aplicando el criterio anterior, se observa que la serie del volumen del triplay tiene media constante en la primera diferencia.

Sobre las mismas regresiones se realiza la prueba de White. Un ejemplo del reporte que se obtiene de dicha prueba es el siguiente:

Heteroskedasticity Test: White				
F-statistic	14.74468	Prob. F(2,165)	0.0000	
Obs*R-squared	25.47292	Prob. Chi-Square(2)	0.0000	
Scaled explained SS	17.61872	Prob. Chi-Square(2)	0.0001	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/05/12 Time: 17:56				
Sample: 1994M01 2007M12				
Included observations: 168				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	651685.3	66955.48	9.733114	0.0000
@TREND(0)	-4296.706	1060.91	-4.05002	0.0001
(@TREND(0))^2	-28.06338	21.60926	-1.298674	0.1959
R-squared	0.151625	Mean dependent var	532559.1	
Adjusted R-squared	0.141341	S.D. dependent var	635811.1	
S.E. of regression	589167	Akaike info criterion	29.42850	
Sum squared resid	5.73E+13	Schwarz criterion	29.48429	
Log likelihood	-2468.994	Hannan-Quinn criter.	29.45114	
F-statistic	14.74468	Durbin-Watson stat	0.989687	
Prob(F-statistic)	0.000001			

Varianza NO constante *VOLTRIP*

Heteroskedasticity Test: White				
F-statistic	0.364104	Prob. F(2,164)	0.6954	
Obs*R-squared	0.738251	Prob. Chi-Square(2)	0.6913	
Scaled explained SS	0.672199	Prob. Chi-Square(2)	0.7146	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/05/12 Time: 17:55				
Sample: 1994M02 2007M12				
Included observations: 167				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	203915.9	32446.33	6.284715	0.0000
@TREND(0)	-444.6252	523.8803	-0.848715	0.3973
(@TREND(0))^2	3.558321	10.61653	0.335168	0.7379
R-squared	0.004421	Mean dependent var	207362.3	
Adjusted R-squared	-0.007721	S.D. dependent var	284072.0	
S.E. of regression	285166.5	Akaike info criterion	27.97733	
Sum squared resid	1.33E+13	Schwarz criterion	28.03335	
Log likelihood	-2333.107	Hannan-Quinn criter.	28.00007	
F-statistic	0.364104	Durbin-Watson stat	1.773283	
Prob(F-statistic)	0.695379			

Varianza constante *DVOLTRIP*

Del reporte anterior se obtiene el  $\alpha^*$  (valor  $Ji^2$ ) de *Obs \* R - squared*, que es igual a 0.00000 y se aplica el siguiente criterio práctico:

1. Si  $\alpha^* \geq 0.01 \implies$  que la varianza es constante (Cuadro Verde o sombreado de la derecha).
2. Si  $\alpha^* < 0.01 \implies$  que la varianza no es constante (Cuadro Rojo o sombreado de la izquierda).

El resultado de esta prueba sobre las series *VOLTRIP* y *DVOLTRIP* indica entonces que la varianza es constante en la primera diferencia.

Si  $\mu = Cte$ ,  $Cov = 0$ , y  $Varianza = Cte \implies$  ruido blanco en las series, esta prueba será de utilidad para el análisis de los errores en las regresiones de los modelos tanto *ARIMA* como *ARMAX*.

El ruido blanco implica que la serie analizada o el estudio de los errores de la regresión (si las pruebas son aplicadas en ellos) son bien comportados. Esto es que las estimaciones son confiables.

Como precaución hay que recordar que con una de las condiciones que no se cumpla, ya no es ruido blanco.

**Prueba de raíz unitaria** Una serie de tiempo que presenta raíz unitaria se conoce como caminata aleatoria, y a su vez una caminata aleatoria es un ejemplo de una serie no estacionaria.

La prueba de raíz unitaria nos permite verificar si existe autocorrelación, así como la existencia de raíz unitaria.

A la prueba de raíz unitaria también se le conoce como la prueba Dickey Fuller<sup>4</sup> (*DF*). De esta prueba existen dos versiones; La *DF* simple y la *DF* aumentada (*ADF*), la cual se utilizará en este trabajo. Esta se compone de tres partes:

1. Caminata aleatoria con intercepto.
2. Caminata aleatoria con tendencia e intercepto.
3. Caminata aleatoria.

Las hipótesis que maneja la prueba son:

$H_0 = \text{Tiene raíz unitaria.}$

$H_1 = \text{No tiene raíz unitaria.}$

Consideremos el siguiente modelo para ilustrar la prueba Dickey Fuller simple:

$$y_t = \rho y_{t-1} + \delta x_t + \epsilon_t \quad (3.7)$$

Donde  $x_t$  son regresores exógenos opcionales que pueden ser constantes o constantes y tendencias,  $\rho$  y  $\delta$  son parámetros que se estimarán y  $\epsilon_t$  el término de error de la regresión, el cual debe ser ruido blanco.

Si  $|\rho| \geq 1$ , “ $y_t$ ” es una serie no estacionaria, su varianza se incrementará con el tiempo y tiende al infinito.

Si  $|\rho| < 1$ , “ $y_t$ ” es una serie con tendencia estacionaria. La estacionariedad de la tendencia queda comprobada si el valor absoluto de  $\rho$  es estrictamente menor a uno, pues de lo contrario se demuestra que existe raíz unitaria.

En el caso de la Dickey Fuller aumentada (*ADF*), partimos de la misma ecuación anterior, pero restaremos  $y_{t-1}$  a ambos lados de la misma, quedando entonces:

$$\Delta y_t = \alpha y_{t-1} + x'_t \delta + \epsilon_t$$

Donde,  $\alpha = \rho - 1$ , y  $\Delta y_t = y_t - y_{t-1}$  (llamaremos  $\Delta$  al término diferenciado).

---

<sup>4</sup>Dickey, D.A. and W.A. Fuller (1979), “*Distribution of the Estimators for Autoregressive Time Series with a Unit Root*” *Journal of the American Statistical Association*, 74, p. 427–431.

Las hipótesis nula y alternativa son:

H0:  $\alpha = 0$

H1:  $\alpha < 0$

Evaluadas usando la proporción  $t$  para  $\alpha$ :

$$t_\alpha = \hat{\alpha} / (se(\hat{\alpha}))$$

Donde  $\hat{\alpha}$  es el estimado de  $\alpha$ , y  $se(\hat{\alpha})$  es el coeficiente del error estándar.

La prueba simple Dickey Fuller sólo es válida para procesos correlacionados de primer orden  $AR(1)$ . Si la serie está correlacionada en un orden mayor, el supuesto de ruido blanco de los errores es inválido.

Con la Dickey Fuller aumentada se lleva a cabo una corrección paramétrica para una correlación mayor a 1, asumiendo que las series siguen un proceso  $AR(p)$ , al añadir términos diferenciados rezagados de la variable dependiente al lado derecho de la regresión de prueba, quedando:

$$\Delta y_t = \alpha y_{t-1} + x_t \delta + \beta_1 \Delta y_{t-1} + \beta_2 y_{t-1} + \dots + \beta_p \Delta y_{t-p} + v_t$$

Esta especificación aumentada se utiliza para probar las hipótesis utilizando la proporción  $t$ .

Un resultado importante obtenido por Fuller es que la distribución asintótica de la proporción  $t$  para  $\alpha$  es independiente del número rezagado de las primeras diferencias incluidas en la regresión  $ADF$ . Más aún, mientras el supuesto de que  $y$  sigue un proceso autorregresivo ( $AR$ ) puede parecer restrictivo, Said y Dickey<sup>4</sup> demostraron en 1984 que la prueba  $ADF$  es asintóticamente válida en la presencia de componentes de media móvil ( $MA$ ), siempre que sean incluidos suficientes términos diferenciados rezagados en la regresión de prueba.

Una muestra del reporte que se obtiene en E-Views es el siguiente:

---

<sup>4</sup>Said E. and David A. Dickey (1984), "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order", *Biometrika*, 71, 599–607.

Augmented Dickey-Fuller Unit Root Test on VOLTRIP				
Null Hypothesis: VOLTRIP has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)				
			<b>t-Statistic</b>	<b>Prob.*</b>
<b>Augmented Dickey-Fuller test statistic</b>			-4.049728	0.0015
<b>Test critical values:</b>	<b>1% level</b>		-3.469691	
	<b>5% level</b>		-2.878723	
	<b>10% level</b>		-2.576010	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(VOLTRIP)				
Method: Least Squares				
Date: 09/05/12 Time: 18:08				
Sample (adjusted): 1994M02 2007M12				
Included observations: 167 after adjustments				
<b>Variable</b>	<b>Coefficient</b>	<b>Std. Error</b>	<b>t-Statistic</b>	<b>Prob.</b>
VOLTRIP(-1)	-0.175194	0.043261	-4.049728	0.0001
C	1013.735	251.7343	4.027002	0.0001
R-squared	0.090409	Mean dependent var		3.532934
Adjusted R-squared	0.084897	S.D. dependent var		457.1622
S.E. of regression	437.3260	Akaike info criterion		15.01114
Sum squared resid	31556918	Schwarz criterion		15.04848
Log likelihood	-1251.430	Hannan-Quinn criter.		15.02629
F-statistic	16.40030	Durbin-Watson stat		2.588658
Prob(F-statistic)	0.000079			

Prueba de raíz unitaria con constante en  
VOLTRIP

Augmented Dickey-Fuller Unit Root Test on DVOLTRIP				
Null Hypothesis: DVOLTRIP has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)				
			<b>t-Statistic</b>	<b>Prob.*</b>
<b>Augmented Dickey-Fuller test statistic</b>			-20.00996	0.0000
<b>Test critical values:</b>	<b>1% level</b>		-3.469933	
	<b>5% level</b>		-2.878829	
	<b>10% level</b>		-2.576067	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DVOLTRIP)				
Method: Least Squares				
Date: 09/05/12 Time: 18:12				
Sample (adjusted): 1994M03 2007M12				
Included observations: 166 after adjustments				
<b>Variable</b>	<b>Coefficient</b>	<b>Std. Error</b>	<b>t-Statistic</b>	<b>Prob.</b>
DVOLTRIP(-1)	-1.43352	0.071640	-20.00996	0.0000
C	8.297773	32.27871	0.257066	0.7974
R-squared	0.709425	Mean dependent var		-5.415663
Adjusted R-squared	0.707653	S.D. dependent var		768.9945
S.E. of regression	415.7884	Akaike info criterion		14.91020
Sum squared resid	28352316	Schwarz criterion		14.94770
Log likelihood	-1235.547	Hannan-Quinn criter.		14.92542
F-statistic	400.3983	Durbin-Watson stat		2.107011
Prob(F-statistic)	0.000000			

Prueba de raíz unitaria con constante en  
DVOLTRIP

El criterio de decisión para determinar la existencia de raíz unitaria de las tres pruebas ADF a partir del reporte obtenido, es:

1. Buscar el valor Durbin-Watson stat ( $D$ ) para verificar que no haya autocorrelación, ver las tablas  $D - W$  y obtener  $d_l$  y  $d_u$ . En las tablas “ $n$ ” es el número de observaciones y “ $k$ ” el número de variables explicativas sin considerar el término constante.

Los criterios de evaluación de autocorrelación son los siguientes:

- a) Si  $0 < D < d_l \implies$  Hay autocorrelación.
- b) Si  $d_u < D < 4 - d_u \implies$  No hay autocorrelación.
- c) Si  $4 - d_u < D < 4 \implies$  Hay autocorrelación.

Si en los tres casos se cumple que no hay autocorrelación se puede proseguir, en caso contrario hay que incrementar el número de rezagos hasta que ésta desaparezca.

Dado que en la tercera prueba de raíz unitaria aplicada a  $DVOLTRIP$  había ine-

xistencia de raíz unitaria y de autocorrelación, no fue necesario aumentar rezagos.

A continuación presentaré un cuadro resumen con las conclusiones obtenidas tras aplicar el análisis de autocorrelación especialmente en el segundo criterio ya que éste excluye a los otros dos. Sin embargo es recomendable revisar el resultado de los tres criterios, para el caso de VOLTRIP y DVOLTRIP, los criterios 1 y 3 no se cumplieron:

Los valores  $d_L$  y  $d_U$  se obtienen de la tabla del estadístico  $d$  de Durbin-Watson, con  $n = 150$  y  $k' = 1$  al nivel de significancia del 0.05.

Prueba Durbin Watson de Autocorrelación		
VOLTRIP CON INTERCEPT		# Rezagos= 0
DURBIN STATS	CRITERIO	CONCLUSIÓN
D=	2.588658	El valor D=2.5886 no se encuentra dentro del intervalo [1.746 , 2.254] por lo que <b>Hay Autocorrelación</b>
DU=	1.746	
4-DU=	2.254	
VOLTRIP CON TREND & INTERCEPT		# Rezagos= 0
DURBIN STATS	CRITERIO	CONCLUSIÓN
D=	2.567448	El valor D=2.5674 no se encuentra dentro del intervalo [1.746 , 2.254] por lo que <b>Hay Autocorrelación</b>
DU=	1.746	
4-DU=	2.254	
VOLTRIP CON NONE		# Rezagos= 0
DURBIN STATS	CRITERIO	CONCLUSIÓN
D=	2.808279	El valor D=2.8082 no se encuentra dentro del intervalo [1.746 , 2.254] por lo que <b>Hay Autocorrelación</b>
DU=	1.746	
4-DU=	2.254	

VOLTRIP tiene autocorrelación

Prueba Durbin Watson de Autocorrelación		
DVOLTRIP CON INTERCEPT		# Rezagos= 0
DURBIN STATS	CRITERIO	CONCLUSIÓN
D=	2.107011	El valor D=2.1070 sí se encuentra dentro del intervalo [1.746 , 2.254] por lo que <b>No Hay Autocorrelación</b>
DU=	1.746	
4-DU=	2.254	
DVOLTRIP CON TREND & INTERCEPT		# Rezagos= 0
DURBIN STATS	CRITERIO	CONCLUSIÓN
D=	2.112841	El valor D=2.1128 sí se encuentra dentro del intervalo [1.746 , 2.254] por lo que <b>No Hay Autocorrelación</b>
DU=	1.746	
4-DU=	2.254	
DVOLTRIP CON NONE		# Rezagos= 0
DURBIN STATS	CRITERIO	CONCLUSIÓN
D=	2.106785	El valor D=2.1067 sí se encuentra dentro del intervalo [1.746 , 2.254] por lo que <b>No Hay Autocorrelación</b>
DU=	1.746	
4-DU=	2.254	

DVOLTRIP no tiene autocorrelación

- Si ya desapareció la autocorrelación, se aplica el criterio de decisión para determinar la existencia de raíz unitaria de las tres pruebas ADF a partir del reporte obtenido.

Para hacerlo hay que tomar el valor  $Prob^*$  del *ADF test*, al cual llamaremos  $\alpha^*$ . Según el nivel de confianza que uno determine, 99 %, 95 % o 90 %, determinaremos nuestro  $\alpha$  como  $\alpha = 0.01, \alpha = 0.05$  ó  $\alpha = 0.1$  para realizar el comparativo:

Por lo que realizando las tres pruebas *ADF*, si se cumple para las tres que  $\alpha^* < \alpha$  la serie no tiene raíz unitaria.

En la siguiente tabla, si  $PROBAB^*$  es menor que el nivel de significancia  $\alpha$ , se rechaza la hipótesis de que no hay raíz unitaria. Esta prueba se realiza para tres casos y se deben rechazar los tres para concluir que no hay raíz unitaria. El resumen mostrado toma un  $\alpha = 0.001$  para la prueba realizada a las series *VOLTRIP* y *DVOLTRIP*.

PRUEBA DE RAÍZ UNITARIA		
SERIE	VOLTRIP	
	PROBAB*	CONCLUSIÓN
INTECEPT	0.0010	Coficiente significativo
TREND & INTERCEPT	0.0005	Coficiente significativo
NONE	0.4858	Coficiente NO significativo
		<b>TIENE RAÍZ UNITARIA</b>

VOLTRIP tiene raíz unitaria

PRUEBA DE RAÍZ UNITARIA		
SERIE	DVOLTRIP	
	PROBAB*	CONCLUSIÓN
INTECEPT	0.0000	Coficiente significativo
TREND & INTERCEPT	0.0000	Coficiente significativo
NONE	0.0000	Coficiente significativo
		<b>NO HAY RAÍZ UNITARIA</b>

DVOLTRIP no tiene raíz unitaria

Basándonos en el ejemplo anterior de la tabla de las pruebas *ADF*, podemos decir que si la probabilidad de la tercera prueba de raíz unitaria de la serie *VOLTRIP* es  $\alpha^* = 0.4858$ , al considerar  $\alpha = 0.01$ ,  $\alpha = 0.05$ , y  $\alpha = 0.10$ , entonces, *VOLTRIP* tiene raíz unitaria y por tanto no es estacionaria.

Para poder concluir que una serie es estacionaria se debe cumplir que no haya raíz unitaria y que la varianza sea constante.

Con las pruebas realizadas anteriormente podemos concluir entonces que la serie *VOLTRIP* no es estacionaria en niveles, pero sí lo es en primeras diferencias, quedando demostrado en la variable *DVOLTRIP*.

## 3.2. Modelación

### 3.2.1. Modelo ARIMA univariado

Con el término de la parte anterior, se cubren los puntos 1 a 3 de la sección 2.5.2 del procedimiento del modelado.

Una vez teniendo la serie estacionaria, pasamos a la realización del punto 4 de la sección 2.5.2, donde se procede a analizar el correlograma de la serie *DVOLTRIP* para seleccionar los términos AR y MA que conformarán el modelo econométrico univariado.

Del correlograma se consideran los valores PAC para los términos AR, y los valores AC para los términos MA.

Aquellos valores AC y PAC que sobrepasen el ancho de banda del correlograma, serán utilizados para el modelo econométrico.

Para determinar el ancho de banda del correlograma se utiliza la siguiente fórmula:

$$\text{Ancho de banda} = \frac{2}{\sqrt[2]{\# \text{ de observaciones}}} \quad (3.8)$$

Para hacerlo de manera más sencilla, realicé una fórmula en excel:

Para los valores AR: =  $si(abs(valorPAC) > (\frac{2}{\sqrt[2]{total \text{ de observaciones}}}), \text{pon } \# \text{ obs, " - "})$  (3.9)

Para los valores MA: =  $si(abs(valorAC) > (\frac{2}{\sqrt[2]{total \text{ de observaciones}}}), \text{pon } \# \text{ obs, " - "})$  (3.10)

La siguiente tabla muestra qué términos son los que se considerarán para el modelo.

CORRELOGRAMA DVOLTRIP				TÉRMINOS AR Y MA		
OBS	AC	PAC	Q-Stat	Prob	AR	MA
1	-0.421	-0.421	30.119	0.000	1	1
2	0.045	-0.161	30.466	0.000	2	-
3	0.030	-0.019	30.621	0.000	-	-
4	-0.144	-0.162	34.227	0.000	4	-
5	0.176	0.060	39.623	0.000	-	5
6	-0.146	-0.071	43.384	0.000	-	-
7	0.066	-0.011	44.162	0.000	-	-
8	-0.022	-0.033	44.245	0.000	-	-
9	-0.062	-0.069	44.938	0.000	-	-
10	0.041	-0.064	45.236	0.000	-	-
11	-0.166	-0.207	50.218	0.000	11	11
12	0.390	0.294	77.882	0.000	12	12
13	-0.249	0.005	89.271	0.000	-	13
14	0.079	0.037	90.425	0.000	-	-
15	-0.051	-0.093	90.902	0.000	-	-
16	-0.058	-0.021	91.522	0.000	-	-
17	0.202	0.091	99.202	0.000	-	17
18	-0.187	-0.019	105.820	0.000	-	18
19	0.036	-0.084	106.070	0.000	-	-
20	0.099	0.099	107.960	0.000	-	-
21	-0.116	0.028	110.570	0.000	-	-
22	0.015	-0.091	110.610	0.000	-	-
23	-0.059	-0.035	111.290	0.000	-	-
24	0.192	0.071	118.590	0.000	-	24
25	-0.115	0.035	121.220	0.000	-	-
26	0.104	0.119	123.380	0.000	-	-
27	-0.063	0.060	124.180	0.000	-	-
28	-0.062	-0.056	124.970	0.000	-	-
29	0.207	0.096	133.710	0.000	-	29
30	-0.265	-0.101	148.220	0.000	-	30
31	0.131	0.003	151.780	0.000	-	-
32	-0.020	-0.077	151.860	0.000	-	-
33	-0.090	-0.032	153.580	0.000	-	-
34	0.120	0.046	156.640	0.000	-	-
35	-0.143	0.002	161.030	0.000	-	-
36	0.191	0.055	168.860	0.000	-	36
37	-0.061	0.020	169.680	0.000	-	-
38	0.012	0.028	169.720	0.000	-	-
39	-0.053	-0.066	170.330	0.000	-	-
40	-0.050	-0.077	170.890	0.000	-	-
41	0.170	0.009	177.330	0.000	-	41
42	-0.241	-0.065	190.480	0.000	-	42
43	0.168	-0.008	196.910	0.000	-	43
44	-0.050	0.016	197.490	0.000	-	-
45	-0.048	0.008	198.030	0.000	-	-
46	0.147	0.044	203.080	0.000	-	-
47	-0.219	-0.058	214.390	0.000	-	47
48	0.169	-0.043	221.150	0.000	-	48
49	0.049	0.113	221.720	0.000	-	-
50	-0.053	0.085	222.400	0.000	-	-

Análisis de correlograma  
parte 1

CORRELOGRAMA DVOLTRIP				TÉRMINOS AR Y MA		
OBS	AC	PAC	Q-Stat	Prob	AR	MA
51	-0.065	-0.135	223.440	0.000	-	-
52	0.000	-0.028	223.440	0.000	-	-
53	0.038	-0.093	223.800	0.000	-	-
54	-0.130	-0.103	228.030	0.000	-	-
55	0.115	-0.030	231.370	0.000	-	-
56	0.011	0.141	231.400	0.000	-	-
57	0.004	0.119	231.400	0.000	-	-
58	-0.021	-0.076	231.520	0.000	-	-
59	-0.150	-0.071	237.360	0.000	-	-
60	0.215	0.019	249.580	0.000	-	60
61	-0.018	0.005	249.660	0.000	-	-
62	0.007	0.034	249.670	0.000	-	-
63	-0.084	-0.082	251.600	0.000	-	-
64	0.008	0.058	251.620	0.000	-	-
65	-0.032	-0.112	251.890	0.000	-	-
66	-0.005	-0.019	251.900	0.000	-	-
67	0.059	0.053	252.880	0.000	-	-
68	-0.066	-0.037	254.140	0.000	-	-
69	0.090	0.007	256.490	0.000	-	-
70	-0.027	0.086	256.700	0.000	-	-
71	-0.125	0.062	261.270	0.000	-	-
72	0.248	0.014	279.500	0.000	-	72
73	-0.124	-0.056	284.120	0.000	-	-
74	0.057	0.001	285.100	0.000	-	-
75	-0.056	-0.003	286.070	0.000	-	-
76	-0.061	-0.063	287.240	0.000	-	-
77	0.041	-0.059	287.770	0.000	-	-
78	-0.052	0.014	288.640	0.000	-	-
79	0.003	-0.020	288.640	0.000	-	-
80	0.032	0.010	288.970	0.000	-	-
81	-0.030	-0.080	289.270	0.000	-	-
82	0.015	-0.025	289.350	0.000	-	-
83	-0.074	-0.043	291.200	0.000	-	-
84	0.173	0.017	301.340	0.000	-	84
85	-0.147	-0.058	308.800	0.000	-	-
86	0.092	-0.065	311.720	0.000	-	-
87	-0.063	-0.046	313.140	0.000	-	-
88	-0.041	0.045	313.750	0.000	-	-
89	0.102	0.045	317.520	0.000	-	-
90	-0.068	0.008	319.200	0.000	-	-
91	0.015	-0.008	319.280	0.000	-	-
92	-0.004	-0.049	319.290	0.000	-	-
93	-0.032	-0.046	319.680	0.000	-	-
94	0.037	0.000	320.220	0.000	-	-
95	-0.027	0.004	320.500	0.000	-	-
96	0.103	0.040	324.750	0.000	-	-
97	-0.105	-0.008	329.170	0.000	-	-
98	0.157	0.075	339.290	0.000	-	98
99	-0.126	0.082	345.900	0.000	-	-
100	0.005	0.031	345.910	0.000	-	-

Análisis de correlograma  
parte 2

CORRELOGRAMA DVOLTRIP				TÉRMINOS AR Y MA		
OBS	AC	PAC	Q-Stat	Prob	AR	MA
101	0.076	0.015	348.370	0.000	-	-
102	-0.130	-0.013	355.760	0.000	-	-
103	0.099	0.041	360.080	0.000	-	-
104	-0.072	-0.031	362.410	0.000	-	-
105	0.054	0.015	363.750	0.000	-	-
106	-0.008	-0.055	363.780	0.000	-	-
107	-0.075	-0.035	366.460	0.000	-	-
108	0.073	-0.052	369.000	0.000	-	-
109	-0.079	-0.053	372.040	0.000	-	-
110	0.098	-0.039	376.770	0.000	-	-
111	-0.044	0.102	377.770	0.000	-	-
112	-0.005	0.051	377.790	0.000	-	-
113	0.000	-0.100	377.790	0.000	-	-
114	-0.029	0.098	378.230	0.000	-	-
115	0.091	0.008	382.720	0.000	-	-
116	-0.053	-0.052	384.250	0.000	-	-
117	0.047	-0.033	385.500	0.000	-	-
118	-0.001	-0.019	385.500	0.000	-	-
119	-0.115	-0.060	393.210	0.000	-	-
120	0.122	0.088	402.180	0.000	-	-
121	-0.058	0.035	404.250	0.000	-	-
122	0.047	0.000	405.600	0.000	-	-
123	-0.005	-0.021	405.620	0.000	-	-
124	-0.062	-0.048	408.180	0.000	-	-
125	0.038	-0.003	409.160	0.000	-	-
126	-0.035	0.019	409.980	0.000	-	-
127	0.049	-0.027	411.670	0.000	-	-
128	-0.021	-0.004	412.000	0.000	-	-
129	-0.008	-0.029	412.040	0.000	-	-
130	0.049	0.007	413.870	0.000	-	-
131	-0.083	0.003	419.240	0.000	-	-
132	0.075	-0.031	423.830	0.000	-	-
133	-0.046	0.045	425.560	0.000	-	-
134	0.028	-0.021	426.230	0.000	-	-
135	-0.018	0.010	426.530	0.000	-	-
136	-0.001	0.026	426.530	0.000	-	-
137	-0.005	-0.034	426.560	0.000	-	-
138	-0.009	0.000	426.640	0.000	-	-
139	0.035	-0.062	427.880	0.000	-	-
140	0.007	0.011	427.930	0.000	-	-
141	-0.025	0.000	428.610	0.000	-	-
142	0.019	0.032	429.000	0.000	-	-
143	-0.067	-0.028	434.310	0.000	-	-
144	0.074	-0.040	441.040	0.000	-	-
145	-0.035	0.021	442.640	0.000	-	-
146	0.016	-0.057	442.970	0.000	-	-
147	-0.012	0.007	443.170	0.000	-	-
148	-0.040	-0.035	445.530	0.000	-	-
149	0.020	0.024	446.180	0.000	-	-
150	0.009	-0.010	446.310	0.000	-	-

CORRELOGRAMA DVOLTRIP				TÉRMINOS AR Y MA		
OBS	AC	PAC	Q-Stat	Prob	AR	MA
151	0.002	-0.033	446.320	0.000	-	-
152	0.026	0.035	447.640	0.000	-	-
153	-0.045	0.014	451.720	0.000	-	-
154	0.028	-0.020	453.410	0.000	-	-
155	-0.012	0.041	453.770	0.000	-	-
156	0.032	0.004	456.380	0.000	-	-
157	0.006	0.021	456.470	0.000	-	-
158	-0.021	-0.025	457.910	0.000	-	-
159	-0.001	-0.027	457.920	0.000	-	-
160	-0.003	-0.059	457.950	0.000	-	-
161	0.000	-0.003	457.950	0.000	-	-
162	0.009	0.027	458.420	0.000	-	-
163	-0.016	0.015	460.160	0.000	-	-
164	0.013	-0.015	461.650	0.000	-	-
165	-0.016	0.035	465.18	0.000	-	-
166	0.003	-0.026	465.43	0.000	-	-
TOTAL TÉRMINOS AR'S Y MA'S=					5	20

### Análisis de correlograma parte 4

### Análisis de correlograma parte 3

De lo anterior obtenemos 5 términos AR y 20 términos MA, es decir, un total de 25 términos.

A continuación deben correrse cada una de las combinaciones de los términos AR y MA encontrados.

El total de combinaciones de los modelos ARIMA, se determina por la siguiente fórmula:

$$\text{Combinaciones} = 2^{\text{total de elementos AR y MA}} - 1 = 2^{25} - 1 = 33'554, 431 \quad (3.11)$$

Los términos AR y MA obtenidos del análisis del correlograma de la serie en sus primeras diferencias que se utilizarán para modelar son:

<b>TÉRMINOS AR Y MA</b>	
<b>AR</b>	1, 2, 4, 11, 12
<b>MA</b>	1, 5, 11, 12, 13, 17, 18, 24, 29, 30, 36, 41, 42, 43, 47, 48, 60, 72, 84, 98

Términos con los que se realizarán las combinaciones de los modelos ARIMA

A partir de aquí cubriremos el punto 5 de la sección 2.5.2 del procedimiento, esto es calcular las regresiones de los modelos *ARIMA* necesarias para cubrir todas las combinaciones de los términos *AR* y *MA* mencionados en el punto anterior.

Dado que correrlos de forma manual no es posible, y que la disposición de uso del paquete *E – Views* es restringido, me veo en la necesidad de apoyarme en el paquete de software *R* en el cual desarrollé dos rutinas, una general y una complementaria, para correr de forma automatizada este modelo.

Éstas rutinas de *R* son para correr de forma automatizada todas las combinaciones de modelos *ARIMA*, de los términos obtenidos y resumidos en la tabla anterior. Los scripts de las rutinas se encuentran en el Anexo 2 de este trabajo.

Esta labor se llevó a cabo corriendo de forma paralela en un servidor con 4 CPU's utilizándolos al 100% y tomando un tiempo de 12 días para concluir el proceso.

No se pudo continuar hasta el final ésta metodología en *R*, debido a que algunas pruebas posteriores para seleccionar el mejor modelo pudieron realizarse con ese paquete.

Para el caso del script en *R* se pidió que sólo reportara aquellos modelos que tuvieran los coeficientes completos, significativos y que su estadístico Akaike fuera menor que cualquiera de los previamente seleccionados.

El resumen del script general en *R*, reporta lo siguiente:

RESUMEN SCRIPT "R"			
COMBINACIONES	SE DEBEN ANALIZAR	SE HAN ANALIZADO	SELECCIONADOS
	33'554,431 MODELOS	33'554,431 MODELOS	199 MODELOS

Este cuadro muestra cuantos modelos se debían realizar y analizar, cuantos se lograron y cuantos fueron seleccionados tras cumplir los criterios establecidos

Lo cual implica que se corrieron todos los modelos *ARIMA* con todas las combinaciones de términos *AR* y *MA*, entrando al listado 199 modelos que cumplieron los tres requisitos:

1. Existencia de todos los coeficientes del modelo.
2. Coeficientes significativos, es decir que su estadístico de probabilidad fuera:  $P \leq 0.05$ .
3. Y que tuviera el menor valor del estadístico Akaike.

Una vez ordenando de menor a mayor Akaike, reproduje manualmente los 199 modelos en *E - Views* para buscar el mejor modelo, realizando el análisis de covarianza y ruido blanco de los residuos. De lo anterior obtuve tres tipos de conclusiones en *E - Views*:

1. Failure to achieve convergence: 130 modelos.
2. Coeficientes NO significativos: 31 modelos.
3. Covarianza no constante: 37 modelos.
4. Correctos: 1 modelo.

Dado que la optimización de ambos paquetes se realiza de forma distinta, y un único modelo no me servía para comparar, opté por analizar los términos *MA* que se repetían más en el resumen obtenido de *R* de los modelos que cumplían los requisitos y poder utilizar un script de automatización en *E - Views*, el cual se encuentra en el Anexo 4. Sin embargo, se realizó un script adicional para completar la rutina de *E - Views* en *PHP*, el cual se encarga de crear la parte del código correspondiente a la generación de combinaciones de términos *AR* y *MA*. Este código se encuentra en el Anexo 3 de este trabajo. Sólo tomé los términos *MA* que más se

repetían en los 199 obtenidos con  $R$  e incluí los 5 términos  $AR$ , obteniendo entonces la siguiente selección de términos  $AR$  y  $MA$ :

<b>TÉRMINOS AR Y MA</b>	
<b>AR</b>	1, 2, 4, 11, 12
<b>MA</b>	1, 11, 12, 13, 17, 30, 36

Términos para modelar automáticamente en  
E-Views

Obteniendo ahora un total de combinaciones de:

$$Combinaciones = (2^{12}) - 1 = 4,095 \quad (3.12)$$

El criterio de este script es guardar las 100 mejores regresiones, cumpliendo con el criterio del menor estadístico de Schwarz y coeficientes significativos.

Una vez teniendo el listado de las 100 mejores, las repetí manualmente para realizar las pruebas de significancia de coeficientes, constancia de covarianza y ruido blanco de los residuos.

Terminado lo anterior, obtuve las siguientes conclusiones:

1. Failure to achieve convergence: 46 modelos.
2. Covarianza no constante: 48 modelos.
3. Correctos. 6 Modelos:

La siguiente tabla muestra los valores, tanto de los coeficientes como el estadístico de Schwarz de los mejores modelos:

MODELO:	d(voltrip)= f(combinaciones términos AR y MA)								Estadística de Schwartz
REGRESIÓN	PROBAB $\alpha^*$								
	AR(1)	AR(2)	AR(12)	MA(1)	MA(12)	MA(17)	MA(30)	MA(36)	
1	0.0000	0.0450		0.0237	0.0000	0.0007	0.0000	0.0000	14.77269
2	0.0000		0.0000					0.0039	14.77892
3	0.0000		0.0000						14.78223
4	0.0000	0.0485	0.0000					0.0046	14.78559
5	0.0000	0.0494	0.0000						14.78928
6	0.0000		0.0000				0.0230		14.79303

Modelos que cubren las condiciones mínimas necesarias para ser candidatos al mejor modelo

La teoría indica que el mejor modelo es aquel con menor estadístico de Schwarz, y que cubra el punto 6 de la sección 2.5.2, que es la verificación de ruido blanco en los residuos. Realicé los análisis de los 6 modelos y al mismo tiempo cubrí en la misma sección explicativa de cada uno de los modelos los puntos 6 al 10 de la sección 2.5.2, sólo para comparar.

### Modelo 1

$$DVOLTRIP = f(AR(1), AR(2), MA(1), MA(12), MA(17), MA(30), MA(36)) \quad (3.13)$$

El resultado obtenido de la regresión fué:

Dependent Variable: D(VOLTRIP)				
Method: Least Squares				
Date: 09/06/12 Time: 10:11				
Sample (adjusted): 1994M04 2007M12				
Included observations: 165 after adjustments				
Convergence achieved after 32 iterations				
MA Backcast: 1991M04 1994M03				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.572442	0.103119	-5.551265	0.0000
AR(2)	-0.173291	0.085756	-2.020744	0.0450
MA(1)	0.150950	0.066096	2.283807	0.0237
MA(12)	0.354410	0.058109	6.099090	0.0000
MA(17)	0.188290	0.054582	3.449699	0.0007
MA(30)	-0.254119	0.057681	-4.405565	0.0000
MA(36)	0.323335	0.05942	5.441551	0.0000
R-squared	0.411685	Mean dependent va		1.151515
Adjusted R-squared	0.389344	S.D. dependent var		458.2350
S.E. of regression	358.0854	Akaike info criterior		14.64092
Sum squared resid	20259572	Schwarz criterion		14.77269
Log likelihood	-1200.876	Hannan-Quinn crite		14.69441
Durbin-Watson stat	1.980810			
Inverted AR Roots	-.29+.30i	-.29-.30i		
Inverted MA Roots	.96+.24i	.96-.24i	.93-.09i	.93+.09i
	.88+.40i	.88-.40i	.79+.57i	.79-.57i
	.67-.71i	.67+.71i	.55+.78i	.55-.78i
	.39+.85i	.39-.85i	.27-.96i	.27+.96i
	.10-.96i	.10+.96i	-.11+.98i	-.11-.98i
	-.27+.94i	-.27-.94i	-.41-.86i	-.41+.86i
	-.53+.77i	-.53-.77i	-.70-.71i	-.70+.71i
	-.79+.58i	-.79-.58i	-.90+.39i	-.90-.39i
	-.95+.07i	-.95-.07i	-.96-.25i	-.96+.25i

Salida de E-Views del primer modelo

Analizando la tabla, observamos que la probabilidad de los coeficientes (marcados en verde)  $Prob \leq 0.05$ , implica que éstos son significativos.

Ahora veremos si los residuos son ruido blanco.

Correlograma de Residuos E1			ERRORES EN NIVELES			
OBS	AC	PAC	Q-Stat	Prob	AR	MA
1	-0.090	-0.090	1.277	0.258	-	-
2	-0.119	-0.128	3.528	0.171	-	-
3	0.048	-0.025	3.893	0.273	-	-
4	-0.128	-0.139	6.525	0.163	-	-
5	0.035	-0.019	6.722	0.242	-	-
6	-0.014	-0.046	6.753	0.344	-	-
7	-0.020	-0.011	6.822	0.448	-	-
8	-0.007	-0.039	6.831	0.555	-	-
9	-0.043	-0.045	7.146	0.622	-	-
10	0.024	-0.002	7.242	0.702	-	-
11	-0.020	-0.033	7.311	0.773	-	-
12	-0.033	-0.040	7.502	0.823	-	-
13	0.051	0.026	7.948	0.847	-	-
14	-0.055	-0.056	8.480	0.863	-	-
15	-0.091	-0.105	9.906	0.826	-	-
16	0.031	-0.016	10.069	0.863	-	-
17	0.088	0.077	11.420	0.834	-	-
18	-0.068	-0.072	12.247	0.834	-	-
19	-0.054	-0.074	12.768	0.850	-	-
20	0.144	-0.119	16.490	0.886	-	-
21	-0.133	-0.122	19.715	0.939	-	-
22	-0.066	-0.088	20.524	0.950	-	-
23	0.060	-0.010	21.177	0.970	-	-
24	0.007	-0.032	21.186	0.628	-	-
25	0.076	0.045	22.270	0.620	-	-
26	0.074	0.080	23.312	0.615	-	-
27	-0.040	-0.013	23.618	0.651	-	-
28	0.002	-0.019	23.619	0.701	-	-
29	0.031	-0.038	23.808	0.738	-	-
30	-0.105	-0.126	25.947	0.678	-	-
31	-0.009	-0.011	25.963	0.723	-	-
32	-0.026	-0.027	26.097	0.759	-	-
33	0.002	-0.038	26.097	0.798	-	-
34	-0.027	-0.064	26.242	0.827	-	-
35	-0.069	-0.044	27.221	0.823	-	-
36	0.139	0.087	31.168	0.698	-	-
37	-0.006	-0.040	31.177	0.738	-	-
38	-0.022	0.031	31.283	0.771	-	-
39	-0.038	-0.064	31.590	0.794	-	-
40	-0.040	-0.043	31.923	0.815	-	-
41	0.040	0.030	32.272	0.833	-	-
42	-0.054	-0.091	32.894	0.842	-	-
43	0.061	0.040	33.704	0.844	-	-
44	-0.057	-0.081	34.422	0.849	-	-
45	0.004	-0.001	34.426	0.874	-	-
46	0.109	0.061	37.097	0.823	-	-
47	-0.106	-0.037	39.616	0.769	-	-
48	0.037	0.018	39.926	0.790	-	-
49	0.147	0.099	44.885	0.641	-	-
50	-0.098	-0.026	47.124	0.590	-	-
51	-0.054	-0.105	47.801	0.601	-	-
52	-0.047	-0.089	48.319	0.619	-	-
53	-0.028	-0.070	48.502	0.650	-	-
54	-0.025	-0.100	48.647	0.680	-	-

Correlograma de residuos E1 parte 1

Correlograma de Residuos E1			ERRORES EN NIVELES			
OBS	AC	PAC	Q-Stat	Prob	AR	MA
55	0.053	0.076	49.324	0.690	-	-
56	0.135	0.117	53.896	0.557	-	-
57	-0.076	-0.018	55.279	0.540	-	-
58	-0.123	-0.126	59.102	0.435	-	-
59	-0.032	-0.074	59.357	0.462	-	-
60	0.066	0.050	60.464	0.459	-	-
61	0.070	0.030	61.731	0.450	-	-
62	0.006	-0.028	61.739	0.485	-	-
63	-0.090	-0.143	63.871	0.446	-	-
64	-0.023	0.032	64.008	0.476	-	-
65	-0.048	-0.150	64.629	0.490	-	-
66	0.078	0.003	66.295	0.467	-	-
67	-0.003	0.039	66.297	0.501	-	-
68	-0.048	-0.023	66.939	0.514	-	-
69	0.142	0.074	72.645	0.359	-	-
70	-0.055	-0.003	73.501	0.364	-	-
71	-0.011	0.034	73.537	0.395	-	-
72	0.113	-0.012	77.294	0.313	-	-
73	0.000	-0.009	77.294	0.343	-	-
74	0.014	-0.019	77.350	0.372	-	-
75	-0.012	-0.001	77.392	0.402	-	-
76	-0.084	-0.036	79.590	0.367	-	-
77	-0.027	-0.088	79.822	0.390	-	-
78	-0.009	0.047	79.847	0.421	-	-
79	0.037	0.050	80.288	0.438	-	-
80	0.021	-0.016	80.429	0.465	-	-
81	0.005	-0.017	80.437	0.497	-	-
82	0.004	-0.029	80.442	0.528	-	-
83	-0.040	-0.004	80.997	0.542	-	-
84	0.015	-0.045	81.076	0.570	-	-
85	-0.020	-0.038	81.214	0.596	-	-
86	0.030	-0.020	81.535	0.616	-	-
87	-0.044	-0.052	82.227	0.625	-	-
88	-0.019	0.028	82.355	0.650	-	-
89	0.071	-0.001	84.197	0.624	-	-
90	-0.055	0.039	85.347	0.619	-	-
91	-0.024	-0.060	85.558	0.641	-	-
92	0.028	-0.072	85.871	0.660	-	-
93	-0.045	-0.038	86.671	0.665	-	-
94	0.040	0.004	87.325	0.674	-	-
95	0.038	-0.018	87.901	0.684	-	-
96	0.016	0.019	88.013	0.707	-	-
97	0.011	-0.043	88.063	0.730	-	-
98	0.138	0.116	96.206	0.532	-	-
99	-0.076	-0.006	98.698	0.490	-	-
100	-0.020	0.052	98.872	0.513	-	-
101	0.010	0.014	98.919	0.540	-	-
102	-0.044	-0.026	99.793	0.543	-	-
103	0.018	-0.011	99.951	0.567	-	-
104	-0.031	-0.051	100.400	0.582	-	-
105	0.028	-0.053	100.790	0.598	-	-
106	-0.022	-0.023	100.790	0.625	-	-
107	-0.037	0.009	101.500	0.632	-	-
108	-0.057	-0.078	103.160	0.614	-	-
109	-0.026	-0.072	103.520	0.630	-	-
110	0.020	-0.016	103.730	0.650	-	-

Correlograma de residuos E1 parte 2

Correlograma de Residuos E1			ERRORES EN NIVELES			
OBS	AC	PAC	Q-Stat	Prob	AR	MA
111	0.059	0.082	105.630	0.626	-	-
112	-0.001	-0.004	105.640	0.651	-	-
113	-0.067	-0.017	108.250	0.609	-	-
114	0.050	0.081	109.710	0.596	-	-
115	0.044	-0.078	110.860	0.592	-	-
116	-0.019	-0.077	111.090	0.611	-	-
117	0.025	-0.016	111.500	0.626	-	-
118	-0.037	-0.069	112.410	0.628	-	-
119	-0.079	0.026	116.620	0.545	-	-
120	0.101	0.059	123.680	0.390	-	-
121	-0.028	-0.053	124.260	0.401	-	-
122	0.014	-0.011	124.410	0.422	-	-
123	0.016	-0.032	124.600	0.443	-	-
124	-0.063	-0.072	127.710	0.392	-	-
125	0.040	0.023	129.020	0.384	-	-
126	-0.035	-0.014	130.070	0.384	-	-
127	0.038	-0.042	131.310	0.379	-	-
128	0.014	-0.002	131.500	0.398	-	-
129	-0.022	0.004	131.960	0.411	-	-
130	0.032	0.011	132.990	0.411	-	-
131	-0.032	-0.002	133.520	0.422	-	-
132	-0.011	0.049	133.640	0.444	-	-
133	-0.001	0.036	133.640	0.468	-	-
134	0.023	-0.015	134.260	0.477	-	-
135	0.002	-0.028	134.270	0.502	-	-
136	0.011	-0.004	134.410	0.522	-	-
137	-0.004	-0.022	134.430	0.546	-	-
138	-0.007	-0.002	134.490	0.569	-	-
139	0.042	-0.032	137.220	0.527	-	-
140	0.013	0.023	137.500	0.544	-	-
141	-0.034	-0.035	139.460	0.521	-	-
142	0.008	0.023	139.570	0.542	-	-
143	-0.029	-0.035	141.230	0.526	-	-
144	0.021	0.010	142.220	0.526	-	-
145	-0.004	-0.018	142.260	0.549	-	-
146	-0.028	-0.036	144.410	0.522	-	-
147	0.022	-0.012	145.840	0.512	-	-
148	-0.034	-0.035	149.810	0.443	-	-
149	0.007	0.021	150.000	0.462	-	-
150	0.009	-0.028	150.420	0.475	-	-
151	-0.023	-0.005	153.730	0.423	-	-
152	0.032	0.031	162.060	0.273	-	-
153	-0.016	-0.028	165.130	0.238	-	-
154	0.008	0.022	166.730	0.228	-	-
155	0.015	0.043	173.800	0.143	-	-
156	0.028	-0.024	176.030	0.130	-	-
157	0.008	0.018	176.220	0.140	-	-
158	-0.026	-0.052	178.540	0.126	-	-
159	0.006	-0.017	178.690	0.136	-	-
160	-0.005	-0.060	178.830	0.147	-	-
161	0.003	0.019	178.870	0.159	-	-
162	0.010	-0.012	179.570	0.164	-	-
163	-0.009	0.003	180.390	0.167	-	-
164	0.008	0.021	181.250	0.169	-	-
165	-0.012	-0.004	185.44	0.132	-	-
166	0.004	-0.063	156.49	0.69	-	-
<b>Totales</b>				<b>0</b>	<b>0</b>	

Correlograma de residuos E1 parte 3

Como puede observarse, el total de los términos AR y MA que salen de las bandas es 0, por lo que concluimos que:  $Cov = 0$ .

Para analizar si la media es constante, corremos la regresión auxiliar:

$$E1 = f(c, @trend(0)) \quad (3.14)$$

el resultado de esta regresión es:

Dependent Variable: E1				
Method: Least Squares				
Date: 09/06/12 Time: 11:09				
Sample (adjusted): 1994M04 2007M12				
Included observations: 165 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.818038	28.41665	0.20474	0.838
@TREND(0)	-0.354806	0.575554	-0.61646	0.5385
R-squared	0.002326	Mean dependent var		1.205561
Adjusted R-squared	-0.003795	S.D. dependent var		351.4719
S.E. of regression	352.1382	Akaike info criterion		14.57797
Sum squared resid	20212209	Schwarz criterion		14.61562
Log likelihood	-1200.683	Hannan-Quinn crite		14.59325
F-statistic	0.380023	Durbin-Watson stat		1.985424
Prob(F-statistic)	0.538451			

Regresión auxiliar de residuos con variable de  
tendencia ficticia.

Como la probabilidad ( $\alpha^*$ ) de @trend(0)  $\alpha^* = 0.5385$ , y  $\alpha^* \geq 0.05$ , implica que el coeficiente de @trend(0) no es significativo. por lo tanto la media de  $E1$  es constante.

Aplicando ahora la prueba de heteroscedasticidad de White al resultado de la regresión, obtenemos:

Heteroskedasticity Test: White				
F-statistic	0.870383	Prob. F(2,162)	0.4207	
Obs*R-squared	1.754154	Prob. Chi-Square(2)	0.4160	
Scaled explained SS	1.507364	Prob. Chi-Square(2)	0.4706	
<b>Test Equation:</b>				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/06/12 Time: 11:20				
Sample: 1994M04 2007M12				
Included observations: 165				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	132441.2	18607.30	7.117701	0.0000
@TREND(0)	394.1679	312.4710	1.261454	0.2090
(@TREND(0))^2	-6.180978	6.261035	-0.987213	0.3250
R-squared	0.010631	Mean dependent var		122498.2
Adjusted R-squared	-0.001583	S.D. dependent var		163055.8
S.E. of regression	163184.9	Akaike info criterion		26.86117
Sum squared resid	4.31E+12	Schwarz criterion		26.91764
Log likelihood	-2213.046	Hannan-Quinn crite		26.88409
F-statistic	0.870383	Durbin-Watson stat		2.037126
Prob(F-statistic)	0.42074			

Resultado de la prueba de White

Como  $\alpha^*$  de Obs\*R-squared es:  $\alpha^* = 0.4160 \geq 0.01$ , concluimos que la varianza es constante.

Como conclusión general para este modelo tenemos que:  $Cov = 0$ ,  $\mu = Cte$  y  $Var = Cte$ , los residuos son ruido blanco y por tanto las estimaciones posibles son confiables.

Pasamos a realizar el pronóstico de VOLTRIP. Entonces, la ecuación teórica del modelo  $ARIMA(2, 1, 36)$  queda como:

$$\begin{aligned}
 VOLTRIP_t = & VOLTRIP_{t-1} + \rho_1(VOLTRIP_{t-1} - VOLTRIP_{t-2}) + \\
 & \rho_2(VOLTRIP_{t-2} - VOLTRIP_{t-3}) + \theta_1 E_{t-1} + \theta_{12} E_{t-12} \\
 & + \theta_{17} E_t + \theta_{30} E_{t-30} + \theta_{36} E_t + E_t \quad (3.15)
 \end{aligned}$$

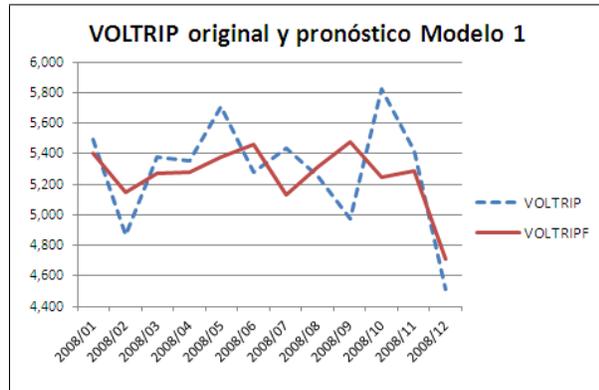
Y la ecuación estimada se expresa de la siguiente manera:

$$\begin{aligned}
 VOLTRIP_t = & VOLTRIP_{t-1} - 0.572442 * VOLTRIP_{t-1} + 0.572442 * VOLTRIP_{t-2} \\
 & - 0.173291 * VOLTRIP_{t-2} + 0.173291 * VOLTRIP_{t-3} \\
 & + 0.150950 * E_{t-1} + 0.354410 * E_{t-12} + 0.188290 * E_{t-17} \\
 & - 0.254119 * E_{t-30} + 0.323335 * E_{t-36} + E_t \quad (3.16)
 \end{aligned}$$

Para los puntos 9 y 10 de la sección 2.5.2, referentes a la realización del pronóstico y graficado, el procedimiento en  $E - Views$  es ampliar el rango de la muestra, seleccionar en el menú *Forecast*, definir la variable a pronosticar, que en este caso es VOLTRIP.

Se indica también el período de pronóstico, el cual en este caso es de enero a diciembre de 2008 y nos da los valores pronosticados. Presento a continuación una tabla y la gráfica del pronóstico:

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO 1			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,401.646	9,092.324
2008/02	4,871	5,149.506	77,565.735
2008/03	5,380	5,267.437	12,670.522
2008/04	5,355	5,283.139	5,163.996
2008/05	5,708	5,374.529	111,203.056
2008/06	5,278	5,462.995	34,223.160
2008/07	5,435	5,126.789	94,994.174
2008/08	5,253	5,315.985	3,967.139
2008/09	4,971	5,480.827	259,923.300
2008/10	5,820	5,247.994	327,191.320
2008/11	5,417	5,287.091	16,876.470
2008/12	4,514	4,712.008	39,207.000
Estadísticos de ajuste de la estimación			
	SSE=	992,078.196	
	MSE=	82,673.183	
	ME=	287.529	
	CORRELACIÓN=	0.561	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo ARIMA 1

Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los  $MSE$  es 287.529, el cual representa un 5.43% del valor de la media de los datos observados, concluyo que en éste modelo tengo un error medio aproximado del 5%.

## Modelo 2

El procedimiento para los modelos 2 a 6 es el mismo que el del modelo 1, por lo que sólo presentaré los resúmenes de cada uno de ellos.

$$DVOLTRIP = f(AR(1), AR(12), MA(36)) \quad (3.17)$$

El resultado de la regresión fue:

Dependent Variable: D(VOLTRIP)				
Method: Least Squares				
Date: 09/06/12 Time: 14:21				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Convergence achieved after 12 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.402418	0.06943	-5.795991	0.0000
AR(12)	0.358709	0.069603	5.153612	0.0000
MA(36)	0.250732	0.085608	2.928824	0.0039
R-squared	0.343555	Mean dependent var	-4.161290	
Adjusted R-squared	0.334917	S.D. dependent var	461.9271	
S.E. of regression	376.7136	Akaike info criterion	14.72001	
Sum squared resid	21570796	Schwarz criterion	14.77892	
Log likelihood	-1137.801	Hannan-Quinn criter.	14.74394	
Durbin-Watson stat	2.087196			
Inverted AR Roots	0.89	.77+.46i	.77-.46i	.43-.79i
	.43+.79i	-.03-.91i	-.03+.91i	-.49-.79i
	-.49+.79i	-.84+.45i	-.84-.45i	-0.96
Inverted MA Roots	.96-.08i	.96+.08i	.93-.25i	.93+.25i
	.87+.41i	.87-.41i	.79-.55i	.79+.55i
	.68-.68i	.68+.68i	.55-.79i	.55+.79i
	.41-.87i	.41+.87i	.25+.93i	.25-.93i
	.08-.96i	.08+.96i	-.08+.96i	-.08-.96i
	-.25+.93i	-.25-.93i	-.41-.87i	-.41+.87i
	-.55+.79i	-.55-.79i	-.68-.68i	-.68+.68i
	-.79-.55i	-.79+.55i	-.87-.41i	-.87+.41i
	-.93+.25i	-.93-.25i	-.96+.08i	-.96-.08i

Salida de E-Views del segundo modelo

Los coeficientes son significativos y el valor de Schwarz es mayor que el modelo anterior; sin embargo, continuaré con el análisis de los residuos. El análisis del correlograma muestra que la covarianza es  $Cov = 0$ , y para la media y varianza constante continuaré con la utilización de la regresión auxiliar.

Dependent Variable: E2				
Method: Least Squares				
Date: 09/06/12 Time: 14:28				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.72172	32.50594	-0.145257	0.8847
@TREND(0)	0.015172	0.673997	0.022511	0.9821
R-squared	0.000003	Mean dependent var		-4.448616
Adjusted R-squared	-0.006533	S.D. dependent var		374.2328
S.E. of regression	375.4532	Akaike info criterion		14.70696
Sum squared resid	21567657	Schwarz criterion		14.74623
Log likelihood	-1137.790	Hannan-Quinn crite		14.72291
F-statistic	0.000507	Durbin-Watson stat		2.087500
Prob(F-statistic)	0.982070			

Regresión auxiliar de residuos con variable de  
tendencia ficticia.

Como la probabilidad ( $\alpha^*$ ) de @trend(0)  $\alpha^* = 0.9821$ , y  $\alpha^* \geq 0.05$ , implica que el coeficiente de @trend(0) no es significativo, por lo tanto la media de  $E2$  es constante.

Aplicando ahora la prueba de heteroscedasticidad de White al resultado de la regresión, obtenemos:

Heteroskedasticity Test: White				
F-statistic	0.119937	Prob. F(2,162)	0.8871	
Obs*R-squared	0.244224	Prob. Chi-Square(2)	0.8850	
Scaled explained SS	0.198378	Prob. Chi-Square(2)	0.9056	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/06/12 Time: 14:40				
Sample: 1995M02 2007M12				
Included observations: 155				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	145842.6	20798.91	7.012034	0.0000
@TREND(0)	142.6797	437.7618	0.325930	0.7449
(@TREND(0))^2	-3.983103	8.13268	-0.489765	0.6250
R-squared	0.001576	Mean dependent var		139146.2
Adjusted R-squared	-0.011562	S.D. dependent var		180254.2
S.E. of regression	181293.2	Akaike info criterion		27.07278
Sum squared resid	5.00E+12	Schwarz criterion		27.13169
Log likelihood	-2095.141	Hannan-Quinn crite.		27.09671
F-statistic	0.119937	Durbin-Watson stat		2.302128
Prob(F-statistic)	0.887060			

Resultado de la prueba de White

Como  $\alpha^*$  de Obs\*R-squared es:  $\alpha^* = 0.8850 \geq 0.01$ , concluimos que la varianza es constante.

Como conclusión general para este modelo tenemos que:  $Cov = 0$ ,  $\mu = Cte$  y  $Var = Cte$ , los residuos son ruido blanco y por tanto las estimaciones posibles son confiables.

Entonces, la ecuación teórica del modelo  $ARIMA(12, 1, 36)$  queda como:

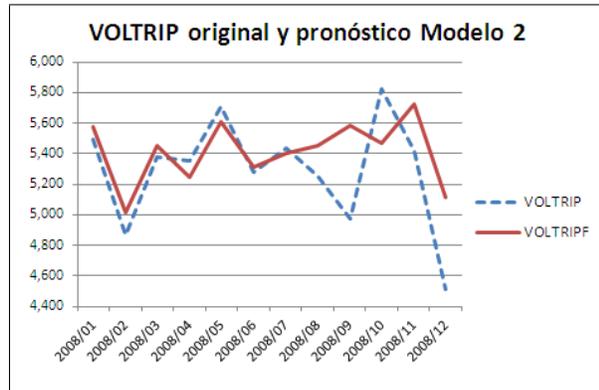
$$VOLTRIP_t = VOLTRIP_{t-1} + \rho_1(VOLTRIP_{t-1} - VOLTRIP_{t-2}) \\ + \rho_{12}(VOLTRIP_{t-12} - VOLTRIP_{t-13}) + \theta_{36}E_{t-36} + E_t \quad (3.18)$$

Y la ecuación estimada se expresa de la siguiente manera:

$$VOLTRIP_t = VOLTRIP_{t-1} - 0.402418 * VOLTRIP_{t-1} \\ + 0.402418 * VOLTRIP_{t-2} + 0.358709 * VOLTRIP_{t-12} \\ - 0.173291 * VOLTRIP_{t-13} + 0.250732 * E_{t-36} + E_t \quad (3.19)$$

Ahora realizaré el pronóstico de VOLTRIP, siguiendo el mismo método del Modelo 1 y su resultado es:

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP 2			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,573.516	5,854.641
2008/02	4,871	5,017.963	21,598.057
2008/03	5,380	5,452.362	5,236.314
2008/04	5,355	5,247.666	11,520.586
2008/05	5,708	5,607.950	10,010.004
2008/06	5,278	5,309.546	995.149
2008/07	5,435	5,402.170	1,077.795
2008/08	5,253	5,454.818	40,730.658
2008/09	4,971	5,587.055	379,524.270
2008/10	5,820	5,469.323	122,974.420
2008/11	5,417	5,724.017	94,259.235
2008/12	4,514	5,110.326	355,604.625
Estadísticos de ajuste de la estimación			
	SSE=	1,049,385.754	
	MSE=	87,448.813	
	ME=	295.717	
	CORRELACIÓN=	0.626	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo ARIMA 2

Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los  $MSE$  es 295.717, el cual representa un 5.58 % del valor de la media de los datos observados, concluyo que en este segundo modelo tengo un error medio aproximado del 6 %.

### Modelo 3

El tercer modelo propuesto es:

$$DVOLTRIP = f(AR(1), AR(12)) \quad (3.20)$$

El resultado de la regresión fue:

Dependent Variable: D(VOLTRIP)				
Method: Least Squares				
Date: 09/06/12 Time: 14:55				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.390820	0.068724	-5.686796	0.0000
AR(12)	0.360406	0.068981	5.224682	0.0000
R-squared	0.319596	Mean dependent var		-4.16129
Adjusted R-squared	0.315148	S.D. dependent var		461.9271
S.E. of regression	382.2712	Akaike info criterion		14.74296
Sum squared resid	22358090	Schwarz criterion		14.78223
Log likelihood	-1140.579	Hannan-Quinn criter.		14.75891
Durbin-Watson stat	2.155681			
Inverted AR Roots	0.89	.77+.46i	.77-.46i	.43-.79i
	.43+.79i	-.03-.91i	-.03+.91i	-.49-.79i
	-.49+.79i	-.84+.45i	-.84-.45i	-0.96

Salida de E-Views del tercer modelo

El análisis del correlograma muestra que la covarianza es  $Cov = 0$ , y para la media y varianza constante continuaré con la utilización de la regresión auxiliar.

Dependent Variable: E3				
Method: Least Squares				
Date: 09/06/12 Time: 15:00				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.202394	33.09356	-0.157203	0.8753
@TREND(0)	0.039084	0.686181	0.056959	0.9547
R-squared	0.000021	Mean dependent var		-4.498885
Adjusted R-squared	-0.006515	S.D. dependent var		381.0014
S.E. of regression	382.2404	Akaike info criterion		14.74280
Sum squared resid	22354479	Schwarz criterion		14.78207
Log likelihood	-1140.567	Hannan-Quinn criter.		14.75875
F-statistic	0.003244	Durbin-Watson stat		2.156031
Prob(F-statistic)	0.954653			

Regresión auxiliar de residuos con variable de tendencia ficticia.

Como la probabilidad ( $\alpha^*$ ) de @trend(0) es  $\alpha^* = 0.9547$ , y  $\alpha^* \geq 0.05$ , implica que el coeficiente de @trend(0) no es significativo, por lo tanto la media de  $E3$  es constante.

Aplicando ahora la prueba de heteroscedasticidad de White al resultado de la regresión, obtenemos:

Heteroskedasticity Test: White				
F-statistic	0.044284	Prob. F(2,162)	0.9567	
Obs*R-squared	0.090263	Prob. Chi-Square(2)	0.9559	
Scaled explained SS	0.070671	Prob. Chi-Square(2)	0.9653	
<b>Test Equation:</b>				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/06/12 Time: 15:15				
Sample: 1995M02 2007M12				
Included observations: 155				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	146823.0	21175.35	6.933674	0.0000
@TREND(0)	-77.87128	445.6850	-0.174723	0.8615
(@TREND(0))^2	-0.515416	8.279875	-0.062249	0.9504
R-squared	0.000582	Mean dependent var	144222.4	
Adjusted R-squared	-0.012568	S.D. dependent var	183425.4	
S.E. of regression	184574.4	Akaike info criterion	27.10866	
Sum squared resid	5.18E+12	Schwarz criterion	27.16756	
Log likelihood	-2097.921	Hannan-Quinn criter.	27.13258	
F-statistic	0.044284	Durbin-Watson stat	2.262756	
Prob(F-statistic)	0.956695			

Resultado de la prueba de White

Como  $\alpha^*$  de Obs\*R-squared es:  $\alpha^* = 0.9559 \geq 0.01$ , concluimos que la varianza es constante.

Como conclusión general para este modelo tenemos que:  $Cov = 0$ ,  $\mu = Cte$  y  $Var = Cte$  los residuos son ruido blanco y por tanto las estimaciones posibles son confiables.

La ecuación teórica del modelo  $ARIMA(12, 1, 0)$  queda como:

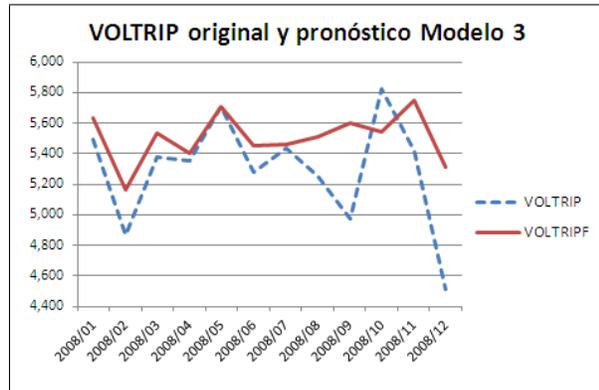
$$VOLTRIP_t = VOLTRIP_{t-1} + \rho_1(VOLTRIP_{t-1} - VOLTRIP_{t-2}) + \rho_{12}(VOLTRIP_{t-12} - VOLTRIP_{t-13}) + E_t \quad (3.21)$$

Y la ecuación estimada se expresa de la siguiente manera:

$$VOLTRIP_t = VOLTRIP_{t-1} - 0.390820 * VOLTRIP_{t-1} + 0.390820 * VOLTRIP_{t-2} + 0.360406 * VOLTRIP_{t-12} - 0.360406 * VOLTRIP_{t-13} + E_t \quad (3.22)$$

El pronóstico de VOLTRIP para el tercer modelo es:

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP 3			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,630.135	17,724.948
2008/02	4,871	5,159.708	83,352.485
2008/03	5,380	5,537.459	24,793.337
2008/04	5,355	5,399.557	1,985.353
2008/05	5,708	5,711.503	12.269
2008/06	5,278	5,454.436	31,129.619
2008/07	5,435	5,459.756	612.843
2008/08	5,253	5,507.773	64,909.280
2008/09	4,971	5,597.129	392,037.020
2008/10	5,820	5,542.024	77,270.721
2008/11	5,417	5,748.809	110,096.896
2008/12	4,514	5,308.308	630,924.975
Estadísticos de ajuste de la estimación			
	SSE=	<b>1,434,849.747</b>	
	MSE=	<b>119,570.812</b>	
	ME=	<b>345.790</b>	
	CORRELACIÓN=	<b>0.640</b>	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo ARIMA 3

Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los  $MSE$  es 345.790, el cual representa un 6.54% del valor de la media de los datos observados, concluyo que en este tercer modelo tengo un error aproximado del 7%.

#### Modelo 4

El cuarto modelo propuesto es:

$$DVOLTRIP = f(AR(1), AR(2), AR(12), MA(36)) \quad (3.23)$$

El resultado de la regresión fue:

Dependent Variable: D(VOLTRIP)				
Method: Least Squares				
Date: 09/06/12 Time: 15:57				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Convergence achieved after 11 iterations				
MA Backcast: 1992M02 1995M01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.466389	0.075927	-6.142569	0.0000
AR(2)	-0.150689	0.075748	-1.989339	0.0485
AR(12)	0.344298	0.069277	4.969913	0.0000
MA(36)	0.247585	0.086125	2.874716	0.0046
R-squared	0.360317	Mean dependent var		-4.16129
Adjusted R-squared	0.347608	S.D. dependent var		461.9271
S.E. of regression	373.102	Akaike info criterion		14.70705
Sum squared resid	21019971	Schwarz criterion		14.78559
Log likelihood	-1135.796	Hannan-Quinn criter.		14.73895
Durbin-Watson stat	1.946939			
Inverted AR Roots	0.87	.75+.46i	.75-.46i	.42-.79i
	.42+.79i	-.04-.92i	-.04+.92i	-.50-.80i
	-.50+.80i	-.83+.46i	-.83-.46i	-0.95
Inverted MA Roots	.96+.08i	.96-.08i	.93+.25i	.93-.25i
	.87-.41i	.87+.41i	.79+.55i	.79-.55i
	.68-.68i	.68+.68i	.55-.79i	.55+.79i
	.41-.87i	.41+.87i	.25+.93i	.25-.93i
	.08+.96i	.08-.96i	-.08+.96i	-.08-.96i
	-.25+.93i	-.25-.93i	-.41-.87i	-.41+.87i
	-.55+.79i	-.55-.79i	-.68+.68i	-.68-.68i
	-.79-.55i	-.79+.55i	-.87-.41i	-.87+.41i
	-.93+.25i	-.93-.25i	-.96+.08i	-.96-.08i

Salida de E-Views del cuarto modelo

El análisis del correlograma muestra que la covarianza es  $Cov = 0$ , y para la media y varianza constante continuaré con la utilización de la regresión auxiliar.

Dependent Variable: E4				
Method: Least Squares				
Date: 09/06/12 Time: 16:03				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.392484	32.08756	-0.168055	0.8668
@TREND(0)	0.030307	0.665322	0.045552	0.9637
R-squared	0.000014	Mean dependent var		-4.846965
Adjusted R-squared	-0.006522	S.D. dependent var		369.4180
S.E. of regression	370.6208	Akaike info criterion		14.68106
Sum squared resid	21016044	Schwarz criterion		14.72033
Log likelihood	-1135.782	Hannan-Quinn criter.		14.69701
F-statistic	0.002075	Durbin-Watson stat		1.947304
Prob(F-statistic)	0.963727			

Regresión auxiliar de residuos con variable de tendencia ficticia.

Como la probabilidad ( $\alpha^*$ ) de @trend(0)  $\alpha^* = 0.9637$ , y  $\alpha^* \geq 0.05$ , implica que el coeficiente de @trend(0) no es significativo, por lo tanto la media de  $E4$  es constante.

Aplicando ahora la prueba de heteroscedasticidad de White al resultado de la regresión, obtenemos:

Heteroskedasticity Test: White				
F-statistic	0.152246	Prob. F(2,162)	0.8589	
Obs*R-squared	0.309880	Prob. Chi-Square(2)	0.8565	
Scaled explained SS	0.260925	Prob. Chi-Square(2)	0.8777	
<b>Test Equation:</b>				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/06/12 Time: 16:14				
Sample: 1995M02 2007M12				
Included observations: 155				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	142320.0	20630.21	6.898622	0.0000
@TREND(0)	190.8966	434.2113	0.439640	0.6608
(@TREND(0))^2	-4.371788	8.066718	-0.541954	0.5886
R-squared	0.001999	Mean dependent var		135587.4
Adjusted R-squared	-0.011132	S.D. dependent var		178830.1
S.E. of regression	179822.8	Akaike info criterion		27.0565
Sum squared resid	4.92E+12	Schwarz criterion		27.1154
Log likelihood	-2093.878	Hannan-Quinn criter.		27.08042
F-statistic	0.152246	Durbin-Watson stat		2.272411
Prob(F-statistic)	0.858908			

Resultado de la prueba de White

Como  $\alpha^*$  de Obs\*R-squared es:  $\alpha^* = 0.8565 \geq 0.01$ , concluimos que la varianza es constante.

Como conclusión general para este modelo tenemos que:  $Cov = 0$ ,  $\mu = Cte$  y  $Var = Cte$  los residuos son ruido blanco y por tanto las estimaciones posibles son confiables.

La ecuación teórica del modelo  $ARIMA(12, 1, 36)$  queda como:

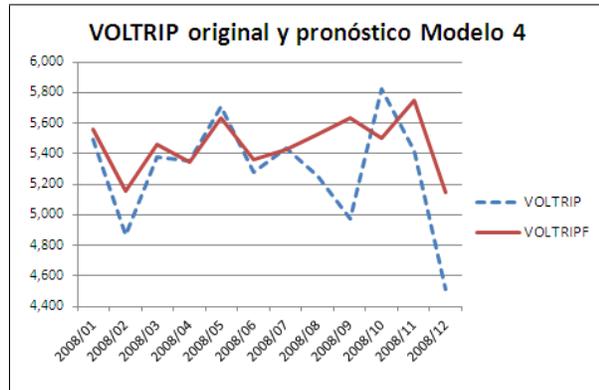
$$VOLTRIP_t = VOLTRIP_{t-1} + \rho_1(VOLTRIP_{t-1} - VOLTRIP_{t-2}) + \rho_2(VOLTRIP_{t-2} - VOLTRIP_{t-3}) + \rho_{12}(VOLTRIP_{t-12} - VOLTRIP_{t-13}) + \theta_{36}E_{t-36} + E_t \quad (3.24)$$

Y la ecuación estimada se expresa de la siguiente manera:

$$VOLTRIP_t = VOLTRIP_{t-1} - 0.466389 * VOLTRIP_{t-1} + 0.466389 * VOLTRIP_{t-2} - 0.150689 * VOLTRIP_{t-2} + 0.150689 * VOLTRIP_{t-3} + 0.344298 * VOLTRIP_{t-12} - 0.344298 * VOLTRIP_{t-13} + 0.247585 * E_{t-36} + E_t \quad (3.25)$$

A continuación se muestra el pronóstico de VOLTRIP.

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP 4			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,558.263	3,753.187
2008/02	4,871	5,156.835	81,701.467
2008/03	5,380	5,457.007	5,930.016
2008/04	5,355	5,343.499	132.268
2008/05	5,708	5,632.636	5,679.748
2008/06	5,278	5,363.275	7,271.752
2008/07	5,435	5,428.565	41.412
2008/08	5,253	5,527.445	75,319.838
2008/09	4,971	5,631.814	436,674.519
2008/10	5,820	5,503.315	100,289.555
2008/11	5,417	5,746.832	108,789.230
2008/12	4,514	5,150.754	405,456.215
Estadísticos de ajuste de la estimación			
	SSE=	1,231,039.208	
	MSE=	102,586.601	
	ME=	320.291	
	CORRELACIÓN=	0.620	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo ARIMA 4

Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los  $MSE$  es 320.291, el cual representa un 6.05 % del valor de la media de los datos observados, concluyo que en este cuarto modelo tengo un error promedio aproximado del 6 %.

### Modelo 5

El quinto modelo propuesto es:

$$DVOLTRIP = f(AR(1), AR(2), AR(12)) \quad (3.26)$$

El resultado de la regresión fué:

Dependent Variable: D(VOLTRIP)				
Method: Least Squares				
Date: 09/06/12 Time: 16:54				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.456668	0.075759	-6.027861	0.0000
AR(2)	-0.147820	0.074625	-1.98083	0.0494
AR(12)	0.351376	0.068483	5.130814	0.0000
R-squared	0.336717	Mean dependent var	-4.161290	
Adjusted R-squared	0.327990	S.D. dependent var	461.9271	
S.E. of regression	378.6703	Akaike info criterion	14.73037	
Sum squared resid	21795467	Schwarz criterion	14.78928	
Log likelihood	-1138.604	Hannan-Quinn criter.	14.75430	
Durbin-Watson stat	2.020488			
Inverted AR Roots	0.88	.75-.46i	.75+.46i	.42-.80i
	.42+.80i	-.04+.92i	-.04-.92i	-.50+.80i
	-.50-.80i	-.83-.46i	-.83+.46i	-0.95

Salida de E-Views del quinto modelo

El análisis del correlograma muestra que la covarianza es  $Cov = 0$ , y para la media y varianza constante continuaré con la utilización de la regresión auxiliar.

Dependent Variable: E5				
Method: Least Squares				
Date: 09/06/12 Time: 17:11				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.023266	32.67352	-0.184347	0.8540
@TREND(0)	0.052968	0.677472	0.078185	0.9378
R-squared	0.000040	Mean dependent var	-5.069843	
Adjusted R-squared	-0.006496	S.D. dependent var	376.1690	
S.E. of regression	377.3888	Akaike info criterion	14.71725	
Sum squared resid	21790613	Schwarz criterion	14.75652	
Log likelihood	-1138.587	Hannan-Quinn criter.	14.73320	
F-statistic	0.006113	Durbin-Watson stat	2.020940	
Prob(F-statistic)	0.937783			

Regresión auxiliar de residuos con variable de tendencia ficticia.

Como la probabilidad ( $\alpha^*$ ) de @trend(0)  $\alpha^* = 0.9378$ , y  $\alpha^* \geq 0.05$ , implica que el coeficiente de @trend(0) no es significativo, por lo tanto la media de  $E5$  es constante.

Aplicando ahora la prueba de heteroscedasticidad de White al resultado de la regresión, obtenemos:

Heteroskedasticity Test: White				
F-statistic	0.152246	Prob. F(2,162)	0.8589	
Obs*R-squared	0.309880	Prob. Chi-Square(2)	0.8565	
Scaled explained SS	0.260925	Prob. Chi-Square(2)	0.8777	
<b>Test Equation:</b>				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/06/12 Time: 16:14				
Sample: 1995M02 2007M12				
Included observations: 155				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	142320.0	20630.21	6.898622	0.0000
@TREND(0)	190.8966	434.2113	0.439640	0.6608
(@TREND(0))^2	-4.371788	8.066718	-0.541954	0.5886
R-squared	0.001999	Mean dependent var	135587.4	
Adjusted R-squared	-0.011132	S.D. dependent var	178830.1	
S.E. of regression	179822.8	Akaike info criterion	27.0565	
Sum squared resid	4.92E+12	Schwarz criterion	27.1154	
Log likelihood	-2093.878	Hannan-Quinn criter.	27.08042	
F-statistic	0.152246	Durbin-Watson stat	2.272411	
Prob(F-statistic)	0.858908			

Resultado de la prueba de White

Como  $\alpha^*$  de Obs\*R-squared es:  $\alpha^* = 0.8565 \geq 0.01$ , concluimos que la varianza es constante.

Como conclusión general para este modelo tenemos que:  $Cov = 0$ ,  $\mu = Cte$  y  $Var = Cte$ , los residuos son ruido blanco y por tanto las estimaciones posibles son confiables.

La ecuación teórica del modelo  $ARIMA(12, 1, 0)$  queda como:

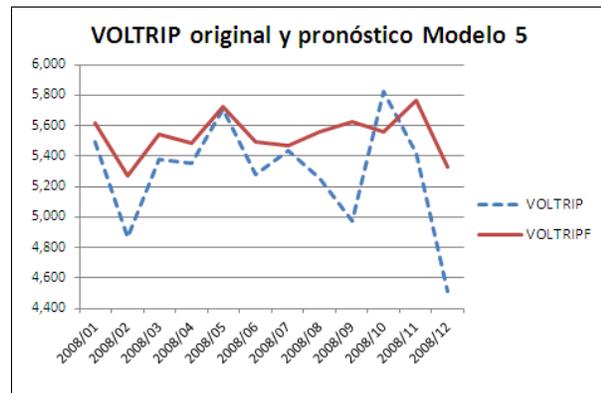
$$VOLTRIP_t = VOLTRIP_{t-1} + \rho_1(VOLTRIP_{t-1} - VOLTRIP_{t-2}) + \rho_2(VOLTRIP_{t-2} - VOLTRIP_{t-3}) + \rho_{12}(VOLTRIP_{t-12} - VOLTRIP_{t-13}) + E_t \quad (3.27)$$

Y la ecuación estimada se expresa de la siguiente manera:

$$VOLTRIP_t = VOLTRIP_{t-1} - 0.456668 * VOLTRIP_{t-1} + 0.456668 * VOLTRIP_{t-2} - 0.147820 * VOLTRIP_{t-2} + 0.147820 * VOLTRIP_{t-3} + 0.352376 * VOLTRIP_{t-12} - 0.331376 * VOLTRIP_{t-13} + E_t \quad (3.28)$$

A continuación se muestra el pronóstico de VOLTRIP.

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP 5			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,616.837	14,360.974
2008/02	4,871	5,273.132	161,710.036
2008/03	5,380	5,545.541	27,403.857
2008/04	5,355	5,481.434	15,985.616
2008/05	5,708	5,722.028	196.772
2008/06	5,278	5,489.867	44,887.460
2008/07	5,435	5,467.559	1,060.106
2008/08	5,253	5,560.906	94,805.831
2008/09	4,971	5,626.988	430,319.732
2008/10	5,820	5,563.335	65,877.138
2008/11	5,417	5,763.242	119,883.454
2008/12	4,514	5,330.687	666,976.991
Estadísticos de ajuste de la estimación			
	SSE=	1,643,467.966	
	MSE=	136,955.664	
	ME=	370.075	
	CORRELACIÓN=	0.646	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo ARIMA 5

Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los  $MSE$  es 370.075, el cual representa un 6.99 % del valor de la media de los datos observados, concluyo que en este quinto modelo tengo un error aproximado del 7 %.

### Modelo 6

El sexto modelo propuesto es:

$$DVOLTRIP = f(AR(1), AR(12), MA(30)) \quad (3.29)$$

El resultado de la regresión fué:

Dependent Variable: D(VOLTRIP)				
Method: Least Squares				
Date: 09/06/12 Time: 17:36				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Convergence achieved after 11 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.383446	0.071095	-5.393457	0.0000
AR(12)	0.327395	0.071678	4.567566	0.0000
MA(30)	-0.205126	0.089284	-2.297459	0.0230
R-squared	0.334223	Mean dependent var	-4.161290	
Adjusted R-squared	0.325462	S.D. dependent var	461.9271	
S.E. of regression	379.3818	Akaike info criterion	14.73413	
Sum squared resid	21877444	Schwarz criterion	14.79303	
Log likelihood	-1138.895	Hannan-Quinn criter.	14.75805	
Durbin-Watson stat	2.140234			
Inverted AR Roots	0.88	.76+.45i	.76-.45i	.43-.78i
		.43+.78i	-.03-.91i	-.03+.91i
		-.49+.78i	-.83+.45i	-.83-.45i
Inverted MA Roots	0.95	.93+.20i	.93-.20i	.87-.39i
		.87+.39i	.77+.56i	.77-.56i
		.63+.70i	.47-.82i	.47+.82i
		.29+.90i	.10-.94i	.10+.94i
		-.10-.94i	-.29+.90i	-.29-.90i
		-.47+.82i	-.63+.70i	-.63-.70i
		-.77+.56i	-.87+.39i	-.87-.39i
		-.93+.20i	-0.95	

Salida de E-Views del sexto modelo

El análisis del correlograma muestra que la covarianza es  $Cov = 0$ , y para la media y varianza constante continuaré con la utilización de la regresión auxiliar.

Dependent Variable: E6				
Method: Least Squares				
Date: 09/06/12 Time: 17:43				
Sample (adjusted): 1995M02 2007M12				
Included observations: 155 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.806039	32.73406	-0.17737	0.8595
@TREND(0)	-0.019961	0.678727	-0.029409	0.9766
R-squared	0.000006	Mean dependent var	-6.165334	
Adjusted R-squared	-0.006530	S.D. dependent var	376.8595	
S.E. of regression	378.0880	Akaike info criterion	14.72095	
Sum squared resid	21871429	Schwarz criterion	14.76022	
Log likelihood	-1138.874	Hannan-Quinn criter.	14.73690	
F-statistic	0.000865	Durbin-Watson stat	2.140822	
Prob(F-statistic)	0.976577			

Regresión auxiliar de residuos con variable de tendencia ficticia.

Como la probabilidad ( $\alpha^*$ ) de @trend(0)  $\alpha^* = 0.9766$ , y  $\alpha^* \geq 0.05$ , implica que el coeficiente

de @trend(0) no es significativo, por lo tanto la media de  $E5$  es constante.

Aplicando ahora la prueba de heteroscedasticidad de White al resultado de la regresión, obtenemos:

Heteroskedasticity Test: White				
F-statistic	0.013408	Prob. F(2,162)	0.9867	
Obs*R-squared	0.027341	Prob. Chi-Square(2)	0.9864	
Scaled explained SS	0.021913	Prob. Chi-Square(2)	0.9891	
<b>Test Equation:</b>				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 09/06/12 Time: 17:46				
Sample: 1995M02 2007M12				
Included observations: 155				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	138660.1	20965.62	6.613691	0.0000
@TREND(0)	-18.54478	441.2707	-0.042026	0.9665
(@TREND(0))^2	1.195038	8.197868	0.145774	0.8843
R-squared	0.000176	Mean dependent var	141106.0	
Adjusted R-squared	-0.012979	S.D. dependent var	181571.8	
S.E. of regression	182746.3	Akaike info criterion	27.08875	
Sum squared resid	5.08E+12	Schwarz criterion	27.14766	
Log likelihood	-2096.378	Hannan-Quinn criter.	27.11268	
F-statistic	0.013408	Durbin-Watson stat	2.226042	
Prob(F-statistic)	0.986682			

Resultado de la prueba de White

Como  $\alpha^*$  de Obs\*R-squared es:  $\alpha^* = 0.9864 \geq 0.01$ , concluimos que la varianza es constante.

Como conclusión general para este modelo tenemos que:  $Cov = 0$ ,  $\mu = Cte$  y  $Var = Cte$ , los residuos son ruido blanco, por tanto las estimaciones posibles son confiables.

La ecuación teórica del modelo  $ARIMA(12, 1, 30)$  queda como:

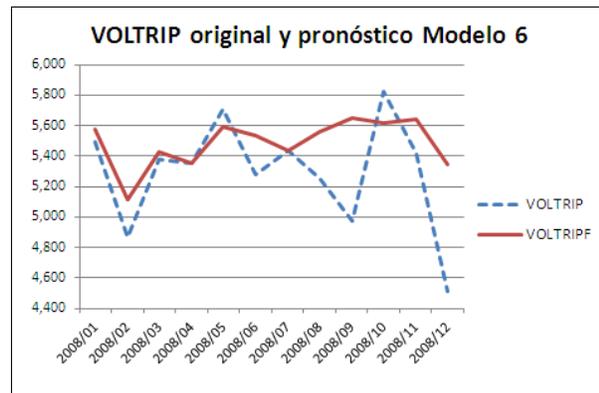
$$VOLTRIP_t = VOLTRIP_{t-1} + \rho_1(VOLTRIP_{t-1} - VOLTRIP_{t-2}) + \rho_{12}(VOLTRIP_{t-12} - VOLTRIP_{t-13}) + \theta_{30}E_{30} + E_t \quad (3.30)$$

Y la ecuación estimada se expresa de la siguiente manera:

$$VOLTRIP_t = VOLTRIP_{t-1} - 0.383446 * VOLTRIP_{t-1} + 0.383446 * VOLTRIP_{t-2} + 0.327295 * VOLTRIP_{t-12} - 0.327395 * VOLTRIP_{t-13} - 0.205126 * E_{30} + E_t \quad (3.31)$$

A continuación se muestra el pronóstico de VOLTRIP.

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP 6			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,579.446	6,797.409
2008/02	4,871	5,118.126	61,071.182
2008/03	5,380	5,430.024	2,502.360
2008/04	5,355	5,349.990	25.099
2008/05	5,708	5,595.053	12,757.115
2008/06	5,278	5,537.590	67,387.122
2008/07	5,435	5,439.263	18.173
2008/08	5,253	5,560.843	94,767.619
2008/09	4,971	5,648.201	458,601.054
2008/10	5,820	5,621.027	39,590.086
2008/11	5,417	5,645.831	52,363.541
2008/12	4,514	5,345.478	691,356.294
Estadísticos de ajuste de la estimación			
	SSE=	1,487,237.054	
	MSE=	123,936.421	
	ME=	352.046	
	CORRELACIÓN=	0.546	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo ARIMA 6

Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los  $MSE$  es 352.046, el cual representa un 6.65 % del valor de la media de los datos observados, concluyo que en este sexto modelo tengo un error aproximado del 7 %.

Siguiendo, las reglas de selección del mejor modelo, para el caso univariado concluyo que el primer modelo es el mejor de los seis presentados y queda como definitivo para posteriormente utilizarlo en el Capítulo 6, durante la comparación de resultados.

### 3.2.2. Modelo *ARMAX* multivariado

1. Corremos tres regresiones: La primera con todas las variables involucradas en la propuesta del modelo.

La segunda con las variables no significativas del primer modelo, para analizar si se hacen significativas en conjunto.

La tercera regresión es sólo con las variables significativas de las dos primeras regresiones y se considera el modelo definitivo.

Sus resultados son:

Dependent Variable: LOG(VOLTRIP)					
Method: Least Squares					
Date: 08/02/12 Time: 16:40					
Sample: 1994M01 2007M12					
Included observations: 168					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	0.376053	0.537808	0.699233	0.4855	No significativa
LOG(VALTRIPDEF)	0.672933	0.03746	17.96415	0.0000	Significativa
LOG(VALCONSTDEF)	-0.008339	0.017306	-0.481823	0.6306	No significativa
LOG(VALMUEBDEF)	0.183299	0.029135	6.291446	0.0000	Significativa
LOG(VALCOLCHDEF)	-0.054624	0.033958	-1.608593	0.1097	No significativa
LOG(VALPAPDEF)	0.135987	0.065281	2.083114	0.0389	Significativa
LOG(VALCARTDEF)	-0.017990	0.031059	-0.579215	0.5633	No significativa
LOG(VALENVCARTDEF)	0.283882	0.056818	4.996316	0.0000	Significativa
LOG(VALEDPERDEF)	-0.271008	0.046098	-5.879023	0.0000	Significativa
LOG(VALEDLIBDEF)	0.011264	0.008607	1.308688	0.1926	No significativa
LOG(VALIMPDEF)	0.054804	0.035331	1.551177	0.1229	No significativa
LOG(IMWDEF)	-0.114490	0.031898	-3.589275	0.0004	Significativa
CETE28DEF	0.533437	0.605653	0.880764	0.3798	No significativa
R-squared	0.907719	Mean dependent var	8.650032		
Adjusted R-squared	0.900575	S.D. dependent var	0.134659		
S.E. of regression	0.042460	Akaike info criterion	-3.406275		
Sum squared resid	0.279445	Schwarz criterion	-3.164540		
Log likelihood	299.1271	Hannan-Quinn criter.	-3.308167		
F-statistic	127.0551	Durbin-Watson stat	1.300949		
Prob(F-statistic)	0.000000				

Modelo 1.1 con todas las series y sacar las significativas para el siguiente modelo.

Dependent Variable: LOG(VOLTRIP)					
Method: Least Squares					
Date: 08/02/12 Time: 16:51					
Sample: 1994M01 2007M12					
Included observations: 168					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	4.336244	0.649073	6.680675	0.0000	Significativa
LOG(VALCONSTDEF)	0.057499	0.035100	1.638150	0.1033	No significativa
LOG(VALCOLCHDEF)	0.026869	0.042555	0.631381	0.5287	No significativa
LOG(VALCARTDEF)	0.143147	0.053358	2.682789	0.0081	Significativa
LOG(VALEDLIBDEF)	-0.030548	0.017132	-1.783108	0.0765	No significativa
LOG(VALIMPDEF)	0.338485	0.055125	6.140350	0.0000	Significativa
CETE28DEF	1.900632	1.311765	1.448912	0.1493	No significativa
R-squared	0.472968	Mean dependent var	8.650032		
Adjusted R-squared	0.453327	S.D. dependent var	0.134659		
S.E. of regression	0.099563	Akaike info criterion	-1.735276		
Sum squared resid	1.595963	Schwarz criterion	-1.605111		
Log likelihood	152.7632	Hannan-Quinn criter.	-1.682449		
F-statistic	24.08075	Durbin-Watson stat	0.573998		
Prob(F-statistic)	0.000000				

Modelo 1.2 seleccionamos sólo las variables NO significativas del modelo 1.1

Dependent Variable: LOG(VOLTRIP)					
Method: Least Squares					
Date: 08/02/12 Time: 16:58					
Sample: 1994M01 2007M12					
Included observations: 168					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	0.372259	0.524769	0.709376	0.4791	No significativa
LOG(VALTRIPDEF)	0.658275	0.034722	18.95836	0.0000	Significativa
LOG(VALMUEBDEF)	0.149274	0.025816	5.782279	0.0000	Significativa
LOG(VALPAPDEF)	0.156875	0.059634	2.630626	0.0094	Significativa
LOG(VALCARTDEF)	0.011721	0.02977	0.393721	0.6943	No significativa
LOG(VALENCARTDEF)	0.237945	0.043189	5.509340	0.0000	Significativa
LOG(VALEDPERDEF)	-0.293561	0.041072	-7.147398	0.0000	Significativa
LOG(VALIMPDEF)	0.076486	0.034311	2.229184	0.0272	Significativa
LOG(IMWDEF)	-0.099504	0.030839	-3.226603	0.0015	Significativa
R-squared	0.902313	Mean dependent var	8.650032		
Adjusted R-squared	0.897398	S.D. dependent var	0.134659		
S.E. of regression	0.043133	Akaike info criterion	-3.396963		
Sum squared resid	0.295815	Schwarz criterion	-3.229608		
Log likelihood	294.3449	Hannan-Quinn criter.	-3.329043		
F-statistic	183.5819	Durbin-Watson stat	1.224356		
Prob(F-statistic)	0.000000				

Modelo 1.3 lo consideramos el definitivo por conjuntar las variables significativas de los modelos 1.1 y 1.2

2. Genero la serie de residuos  $E1 = resid$  para realizar el análisis del correlograma y obtener los términos  $AR$  y  $MA$ .

Correlograma de E1		ERRORES EN NIVELES				
OBS	AC	PAC	Q-Stat	Prob	AR	MA
1	0.383	0.383	25.089	0.000	1	1
2	0.255	0.127	36.280	0.000	-	2
3	0.028	-0.126	36.417	0.000	-	-
4	-0.102	-0.129	38.236	0.000	-	-
5	-0.042	0.070	38.550	0.000	-	-
6	-0.159	-0.136	43.014	0.000	-	6
7	-0.084	-0.007	44.267	0.000	-	-
8	-0.074	-0.001	45.245	0.000	-	-
9	-0.059	-0.029	45.875	0.000	-	-
10	0.087	0.116	47.255	0.000	-	-
11	0.111	0.075	49.488	0.000	-	-
12	0.212	0.108	57.739	0.000	-	12
13	0.085	-0.083	59.070	0.000	-	-
14	-0.047	-0.122	59.473	0.000	-	-
15	-0.081	-0.023	60.686	0.000	-	-
16	-0.146	-0.031	64.712	0.000	-	-
17	-0.032	0.072	64.901	0.000	-	-
18	-0.089	-0.057	66.393	0.000	-	-
19	-0.071	-0.040	67.355	0.000	-	-
20	-0.031	0.008	67.537	0.000	-	-
21	-0.072	-0.056	68.549	0.000	-	-
22	0.065	0.054	69.385	0.000	-	-
23	0.095	0.064	71.176	0.000	-	-
24	0.024	-0.117	71.295	0.000	-	-
25	0.032	0.011	71.494	0.000	-	-
26	-0.110	-0.050	73.911	0.000	-	-
27	0.033	0.133	74.136	0.000	-	-
28	-0.039	-0.047	74.451	0.000	-	-
29	-0.018	-0.038	74.515	0.000	-	-
30	0.036	0.050	74.780	0.000	-	-
31	-0.012	0.022	74.811	0.000	-	-
32	0.033	-0.015	75.035	0.000	-	-
33	0.024	0.040	75.154	0.000	-	-
34	0.089	0.048	76.846	0.000	-	-
35	0.067	-0.053	77.807	0.000	-	-
36	-0.005	-0.012	77.812	0.000	-	-
37	-0.019	-0.001	77.892	0.000	-	-
38	-0.061	0.017	78.708	0.000	-	-
39	-0.035	-0.041	78.972	0.000	-	-
40	-0.024	-0.019	79.101	0.000	-	-
41	0.010	0.078	79.126	0.000	-	-
42	0.022	-0.030	79.232	0.000	-	-
43	0.008	0.013	79.247	0.001	-	-
44	-0.014	-0.063	79.293	0.001	-	-
45	-0.113	-0.164	82.240	0.001	45	-
46	-0.125	-0.069	85.920	0.000	-	-
47	-0.100	0.015	88.297	0.000	-	-
48	-0.161	-0.074	94.504	0.000	-	48
49	-0.093	0.013	96.565	0.000	-	-
50	-0.117	-0.104	99.869	0.000	-	-
51	-0.011	0.074	99.898	0.000	-	-
52	-0.043	-0.109	100.350	0.000	-	-
53	0.008	0.021	100.370	0.000	-	-
54	0.011	-0.053	100.400	0.000	-	-

Correlograma de  
residuos E1 parte 1

Correlograma de E1		ERRORES EN NIVELES				
OBS	AC	PAC	Q-Stat	Prob	AR	MA
55	-0.023	-0.040	100.530	0.000	-	-
56	0.004	0.032	100.540	0.000	-	-
57	-0.005	0.051	100.540	0.000	-	-
58	0.041	0.079	100.980	0.000	-	-
59	0.046	0.023	101.520	0.000	-	-
60	0.043	0.016	102.010	0.001	-	-
61	0.070	0.001	103.310	0.001	-	-
62	0.042	0.047	103.800	0.001	-	-
63	0.024	-0.045	103.960	0.001	-	-
64	0.020	-0.002	104.070	0.001	-	-
65	-0.041	-0.063	104.540	0.001	-	-
66	-0.025	-0.034	104.710	0.002	-	-
67	-0.112	-0.066	108.280	0.001	-	-
68	-0.044	0.024	108.830	0.001	-	-
69	-0.031	0.005	109.120	0.001	-	-
70	0.108	0.129	112.500	0.001	-	-
71	0.161	0.047	120.180	0.000	-	71
72	0.139	0.049	125.950	0.000	-	-
73	0.044	-0.107	126.530	0.000	-	-
74	0.041	0.015	127.050	0.000	-	-
75	-0.054	-0.024	127.930	0.000	-	-
76	-0.112	-0.151	131.820	0.000	-	-
77	-0.163	-0.006	140.160	0.000	-	77
78	-0.156	-0.033	147.890	0.000	-	78
79	-0.151	0.015	155.200	0.000	-	-
80	-0.124	-0.045	160.190	0.000	-	-
81	-0.098	-0.122	163.330	0.000	-	-
82	-0.034	-0.029	163.710	0.000	-	-
83	0.018	-0.052	163.820	0.000	-	-
84	0.050	0.014	164.690	0.000	-	-
85	0.023	-0.034	164.880	0.000	-	-
86	-0.001	0.001	164.880	0.000	-	-
87	0.013	0.041	164.940	0.000	-	-
88	-0.015	0.026	165.030	0.000	-	-
89	0.010	0.063	165.060	0.000	-	-
90	-0.023	0.006	165.250	0.000	-	-
91	-0.028	-0.088	165.530	0.000	-	-
92	-0.009	0.004	165.560	0.000	-	-
93	-0.002	-0.046	165.560	0.000	-	-
94	0.010	-0.059	165.600	0.000	-	-
95	0.072	0.043	167.600	0.000	-	-
96	-0.004	-0.088	167.610	0.000	-	-
97	0.032	-0.041	168.010	0.000	-	-
98	-0.001	0.005	168.010	0.000	-	-
99	0.011	0.020	168.060	0.000	-	-
100	-0.048	-0.120	169.020	0.000	-	-
101	-0.048	-0.018	170.020	0.000	-	-
102	-0.017	-0.034	170.140	0.000	-	-
103	-0.042	0.046	170.930	0.000	-	-
104	0.038	0.044	171.570	0.000	-	-
105	-0.004	-0.057	171.580	0.000	-	-
106	-0.011	-0.033	171.640	0.000	-	-
107	0.028	0.028	172.000	0.000	-	-
108	-0.079	-0.033	174.940	0.000	-	-
109	-0.025	-0.008	175.240	0.000	-	-
110	-0.034	0.077	175.790	0.000	-	-

Correlograma de  
residuos E1 parte 2

Correlograma de E1		ERRORES EN NIVELES				
OBS	AC	PAC	Q-Stat	Prob	AR	MA
111	-0.001	0.000	175.790	0.000	-	-
112	0.063	0.003	177.820	0.000	-	-
113	0.007	-0.020	177.850	0.000	-	-
114	0.066	0.018	180.130	0.000	-	-
115	0.089	0.070	184.400	0.000	-	-
116	0.000	-0.054	184.400	0.000	-	-
117	0.034	0.012	185.070	0.000	-	-
118	-0.019	0.077	185.280	0.000	-	-
119	-0.012	0.010	185.360	0.000	-	-
120	-0.077	-0.063	188.900	0.000	-	-
121	-0.047	-0.034	190.250	0.000	-	-
122	0.007	-0.052	190.280	0.000	-	-
123	0.027	0.023	190.740	0.000	-	-
124	0.066	-0.032	193.520	0.000	-	-
125	0.009	-0.065	193.580	0.000	-	-
126	0.067	0.004	196.590	0.000	-	-
127	0.066	-0.019	199.620	0.000	-	-
128	0.037	-0.016	200.570	0.000	-	-
129	0.068	-0.024	202.950	0.000	-	-
130	0.017	-0.014	204.170	0.000	-	-
131	0.045	0.017	205.740	0.000	-	-
132	0.018	-0.029	206.000	0.000	-	-
133	-0.030	-0.027	206.730	0.000	-	-
134	0.003	0.009	206.740	0.000	-	-
135	-0.011	0.061	206.850	0.000	-	-
136	0.000	0.008	206.850	0.000	-	-
137	0.016	0.056	207.090	0.000	-	-
138	-0.018	-0.036	207.400	0.000	-	-
139	-0.004	0.040	207.420	0.000	-	-
140	-0.013	-0.036	207.590	0.000	-	-
141	0.003	-0.006	207.590	0.000	-	-
142	0.045	-0.027	209.880	0.000	-	-
143	0.017	-0.018	210.180	0.000	-	-
144	0.017	0.002	210.530	0.000	-	-
145	0.017	0.000	210.870	0.000	-	-
146	0.004	-0.085	210.900	0.000	-	-
147	-0.045	-0.016	213.620	0.000	-	-
148	-0.017	-0.007	214.030	0.000	-	-
149	-0.001	0.009	214.040	0.000	-	-
150	-0.040	-0.060	216.590	0.000	-	-
151	-0.024	0.008	217.550	0.000	-	-
152	-0.031	0.034	219.210	0.000	-	-
153	0.038	0.002	221.960	0.000	-	-
154	0.042	-0.028	225.580	0.000	-	-
155	0.031	0.018	227.660	0.000	-	-
156	0.021	-0.021	228.670	0.000	-	-
157	0.026	-0.026	230.370	0.000	-	-
158	0.056	-0.007	234.160	0.000	-	-
159	-0.012	-0.044	234.590	0.000	-	-
160	-0.021	-0.012	236.100	0.000	-	-
161	-0.009	-0.005	236.400	0.000	-	-
162	-0.019	0.007	238.160	0.000	-	-
163	-0.035	-0.011	245.260	0.000	-	-
164	-0.020	0.019	248.020	0.000	-	-
165	-0.006	-0.021	248.310	0.000	-	-
166	-0.01	-0.027	249.690	0.000	-	-
167	-0.005	0.018	250.340	0.000	-	-
					Totales	2 8

Correlograma de  
residuos E1 parte 3

### 3. Análisis de los residuos:

#### a) Prueba de Raíz Unitaria:

Null Hypothesis: E1 has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic - based on SIC, maxlag=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.563811	0.0000
Test critical values:		
1% level	-3.469691	
5% level	-2.878723	
10% level	-2.57601	
*MacKinnon (1996) one-sided p-values.		

Prueba de Raíz Unitaria con  
constante

Null Hypothesis: E1 has a unit root		
Exogenous: Constant, Linear Trend		
Lag Length: 0 (Automatic - based on SIC, maxlag=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.538737	0.0000
Test critical values:		
1% level	-4.013946	
5% level	-3.436957	
10% level	-3.142642	
*MacKinnon (1996) one-sided p-values.		

Prueba de Raíz Unitaria con  
constante y tendencia

Null Hypothesis: E1 has a unit root		
Exogenous: None		
Lag Length: 0 (Automatic - based on SIC, maxlag=0)		
	<b>t-Statistic</b>	<b>Prob.*</b>
Augmented Dickey-Fuller test statistic	-8.590106	0.0000
Test critical values:	1% level	-2.578883
	5% level	-1.942745
	10% level	-1.615438
*MacKinnon (1996) one-sided p-values.		

Prueba de Raíz Unitaria sin constante ni tendencia

Como puede observarse, los valores absolutos de los “Tests critical values” son menores a los valores del “Augmented Dickey-Fuller test statistic”, por lo que puedo concluir con un 90, 95 y 99 % de nivel de confianza que las series están cointegradas.

b) Analizando el correlograma podemos obtener los siguientes términos *AR* y *MA*, los cuales son los que se salen de las bandas:

<b>Términos AR y MA para variable E1A</b>	
<b>AR</b>	<b>MA</b>
1, 45	1, 2, 6, 12,48, 71,77, 78

Términos seleccionados del correlograma, para correr los modelos de E1A

4. Las regresiones se correrán teniendo como variable dependiente *E1* y como variables independientes todas las combinaciones de los términos *AR* y *MA* obtenidas, en este caso:  $2^{10} - 1 = 1023$ .

Dado que son 1023 regresiones me apoyaré en el script de *E - Views* del Anexo 4 de este trabajo. Para agilizar el procedimiento solicité guardar los 50 mejores modelos que cumplan la condición del menor estadístico de Schwarz, con la finalidad de analizarlas manualmente.

Resultados del script de E-Views			
Combinaciones	Van	Son	Gustaron
	1023	1023	26

Resumen de resultados de script de automatización de  
E-Views

5. Resumen de resultados en *EViews*:

Términos AR y MA	MODELO: REGRESIÓN	E1A= f (combinaciones AR y MA)									Estadístico de Schwarz	Conclusión
		PROBAB $\alpha^*$										
		AR(1)	AR(45)	AR(76)	MA(1)	MA(2)	MA(6)	MA(12)	MA(48)	MA(71)		
ar(45) ma(48)	1		0.0227						0.0000		-4.270102	Cov $\neq$ 0
ma(71)	2								0.0000		-4.323465	Cov $\neq$ 0
ar(1) ma(12) ma(48) ma(71)	3	0.0000						0.0020	0.0000	0.0000	-3.832721	Cov $\neq$ 0
ar(1) ar(45) ma(48)	4	0.0048	0.0359						0.0000		-4.294995	Cov $\neq$ 0
ma(2) ma(71)	5	0.0001				0.0000			0.0000		-3.999284	Cov $\neq$ 0
ma(2) ma(12) ma(71)	6	0.0002				0.1021		0.0642	0.0666		-3.676214	COEF NO SIG
ma(1) ma(2) ma(71)	7	0.0000			0.0058	0.0007			0.0000		-4.066948	FAILURE
ma(1) ma(2) ma(12) ma(71)	8				0.0005	0.0002		0.0011	0.0000		-3.925118	FAILURE
ar(1) ma(48)	9				0.0000				0.0000		-3.825108	FAILURE
ma(48)	10								0.0000		-4.019877	Cov $\neq$ 0
ar(1)	11	0.0000									-3.63285	Cov $\neq$ 0
ma(1)	12				0.0002						-3.587375	Cov $\neq$ 0
ma(2)	13					0.0000					-3.562395	Cov $\neq$ 0
ar(1) ma(2)	14	0.0000				0.0067					-3.639774	Cov $\neq$ 0
ma(1) ma(2)	15				0.0000	0.0000					-3.646951	Cov $\neq$ 0
ma(12)	16							0.0001			-3.539786	Cov $\neq$ 0
ar(1) ma(12)	17	0.0000						0.0000			-3.66166	Cov $\neq$ 0
ar(45) ma(12)	18		0.0378					0.0129			-3.573295	Cov $\neq$ 0
ma(1) ma(12)	19				0.0000			0.0000			-3.620234	Cov $\neq$ 0
ma(2) ma(12)	20					0.0000		0.0000			-3.590927	Cov $\neq$ 0
ar(1) ma(2) ma(12)	21	0.0000				0.0103		0.0004			-3.662466	Cov $\neq$ 0
ar(45) ma(2) ma(12)	22		0.0240			0.0001		0.0059			-3.641689	Cov $\neq$ 0
ar(1) ar(45) ma(2) ma(12)	23	0.0001	0.0298			0.0201		0.0394			-3.729655	OK,RB
ma(1) ma(2) ma(12)	24				0.0001	0.0003		0.0172			-3.641848	OK,RB
ma(6) ma(48)	25							0.0120	0.0000		-3.93986	Cov $\neq$ 0
ar(1) ma(48) ma(71)	26	0.0000							0.0000	0.0000	-3.895877	Cov $\neq$ 0

Los renglones en verde muestran los modelos que cumplieron las condiciones de menor estadístico de Schwarz, coeficientes significativos y residuos ruido blanco.

De la tabla anterior tenemos que el mejor modelo es el 24, ya que los coeficientes son significativos pues la probabilidad es:  $PROBAB < 0.05$ .

Sus residuos (Variable  $E2$ ) son ruido blanco, dado que ningún elemento del correlograma se sale de las bandas por lo que  $Cov = 0$ .

La regresión  $E2 = f(c, @trend(0))$  indica que el  $t - statistic$  de  $@trend(0) \geq 0.05$ , por lo que la media de  $E2_t$  es constante.

Finalmente, aplicando la prueba de heteroscedasticidad de residuales de White indica que el  $\alpha^*$  de  $Obs^*R^2 \geq 0.01$  por lo que la varianza es constante.

Al cumplir los tres criterios anteriores concluimos que la serie de residuos  $E2$  es ruido blanco. Podemos establecer entonces que los errores son bien comportados y que las estimaciones posibles son confiables.

6. Se calcula el pronóstico de  $E1$  para todo el período *Ene 2008 - Dic 2008*.

Donde ahora,  $E1_t \text{ pronosticado} = EF_t$

7. Regresando a la ecuación del modelo definitivo y añadiendo la variable  $EF_t$  se corre entonces el modelo a continuación:

$$LVOLTRIP = f(c, lvaltripdef, lvalmuebdef, lvalpapdef, lvalcartdef, lvalenvcartdef, lvaldedperdef, lvalimpdef, limwdef, EF) \quad (3.32)$$

El resultado de la regresión estimada es:

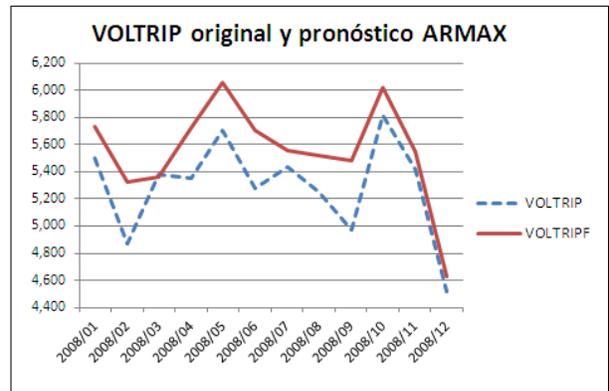
$$\begin{aligned} LVOLTRIP_{t+i} = & \hat{\beta}_1 + \hat{\beta}_2 lvaltripdef_{t+i} + \hat{\beta}_3 lvalmuebdef_{t+i} + \hat{\beta}_4 lvalpapdef_{t+i} \\ & + \hat{\beta}_5 lvalcartdef_{t+i} + \hat{\beta}_6 lvalenvcartdef_{t+i} - \hat{\beta}_7 lvaldedperdef_{t+i} \\ & + \hat{\beta}_8 lvalimpdef_{t+i} - \hat{\beta}_9 limwdef_{t+1} + \hat{\beta}_{10} EF_t \quad (3.33) \end{aligned}$$

Los valores pronosticados se muestran en el siguiente punto.

8. Ampliando el rango de la muestra pedimos a *E-Views* calcular el pronóstico de VOLTRIP donde obtenemos los siguientes resultados y su gráfica.

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,736.024	57,132.473
2008/02	4,871	5,325.582	206,644.795
2008/03	5,380	5,357.446	508.683
2008/04	5,355	5,718.799	132,349.712
2008/05	5,708	6,055.471	120,736.096
2008/06	5,278	5,705.338	182,617.766
2008/07	5,435	5,557.841	15,089.911
2008/08	5,253	5,523.204	73,010.202
2008/09	4,971	5,481.047	260,147.942
2008/10	5,820	6,020.620	40,248.384
2008/11	5,417	5,546.244	16,704.012
2008/12	4,514	4,627.266	12,829.187
Estadísticos de ajuste de la estimación			
	SSE=	1,118,019.16	
	MSE=	93,168.26	
	ME=	305.235	
	CORRELACIÓN=	0.902	

Donde VOLTRIPF son los valores pronosticados



Gráfica de VOLTRIP y VOLTRIPF Modelo ARMAX

9. Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los errores cuadrados es 305.235, el cual representa un 5.77% del valor de la media de los datos observados, concluyo que en este modelo tengo un error aproximado del 6%.

En este capítulo se mostró la elaboración completa de los modelos ARIMA y ARMAX, desde la proposición del modelo de oferta maderera, pasando por la recabación de la información y la realización de los pronósticos de los modelos univariado y multivariado, cuyas respectivas predicciones eventualmente se compararon en el capítulo 5 de este trabajo.

## Capítulo 4

# Introducción a los Modelos de Redes Neuronales Artificiales (*RNA*)

Este capítulo se conforma de tres partes: La primera que muestra los fundamentos y conceptos básicos de las *RNA*, la segunda algunos tipos de *RNA* y la tercera que explica la estructura y operación de los modelos *NAR* y *NARX* que se utilizaron para hacer los pronósticos univariado y multivariado, los cuales se compararon con sus similares econométricos *ARIMA* y *ARMAX*.

### 4.1. Fundamentos y conceptos básicos

En 1943, Warren McCulloch y Walter Pitts realizaron una propuesta de un modelo de neurona cerebral. Por ello se les puede considerar como los padres de las Redes Neuronales Artificiales (*RNA*). Por esa época se creía que las neuronas biológicas eran de carácter binario, lo cual a pesar de ser inexacto, sirvió como fundamento para realizar estudios posteriores del sistema nervioso.

El modelo de neurona McCulloch y Pitts<sup>1</sup> es un dispositivo binario con umbral fijo, el cual hay que superar para cambiar de estado. Esta neurona recibe señales (entradas) de otros elementos, llamadas “sinapsis” si son de valor positivo o “inhibidoras” si son negativas. Si la

---

<sup>1</sup>W. S. McCulloch y W. Pitts, *A logical calculus of the ideas immanent in neurons activity*", Bull. Math. Biophys., 5, 115-133, (1943).

suma total de dichas señales iguala o supera al umbral de la neurona, entonces ésta se activa y envía una señal (salida), de lo contrario la neurona permanece en su estado original de reposo.<sup>2</sup>

La neurona artificial se encuentra inspirada en la neurona biológica. Es decir, que los investigadores piensan en la organización cerebral al considerar configuraciones y algoritmos de Redes Neuronales.

La neurona se considera un elemento formal, un módulo o unidad básica de la red que recibe la información de otros módulos o del entorno. Es entonces que esa información es integrada, computada y emite una única señal de salida que va a transmitir idéntica a múltiples neuronas posteriores.

Consideremos los pesos sinápticos como referencia de la noción biológica de la fuerza de unión entre los elementos; es decir, a la fuerza de la sinapsis. Se considera que una sinapsis es fuerte; es decir, tiene un alto grado de conexión, cuando la información que transmite contribuye, en gran medida, a un nuevo estado o a la alteración que se produzca en la neurona receptora y, por tanto, en la respuesta que ésta elabora.

En una red de neuronas existe un peso o fuerza sináptica que se traduce en un valor numérico, el cual pondera las señales que se reciben por sus entradas. Este peso será un valor que determinará la fuerza de conexión entre 2 neuronas. Cuando se evalúa una neurona se debe calcular el conjunto de todas las fuerzas o valores (denominado *NET*) que se reciben por sus entradas. Una vez calculado el valor conjunto de todas las entradas se aplica una función de activación (*FA*), que determinará el valor del estado interno de la neurona y será lo que se transmita a su salida.

La combinación de las señales que recibe una neurona se puede calcular como muestra en la siguiente ecuación:

$$NET_i(t) = \sum_{j=1}^{N-1} [W_{ij} \cdot O_j \cdot (t - 1)]$$

Donde  $W_{ij}$  representa el peso de la conexión entre una neurona emisora  $j$  y neurona receptora  $i$ .

---

<sup>2</sup>Stephen I. Gallant. Neural Network Learning and Expert Systems. Third edition. The MIT Press 1995.

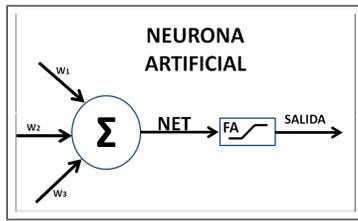


Gráfico de operación de  
una Neurona Artificial

Las funciones de activación y de transferencia son las encargadas de definir el nuevo estado de activación  $A_i$  y la respuesta  $O_i$  de la neurona. Tanto el estado de activación como la salida de la neurona van a variar en función de las entradas que recibe en un determinado momento y del estado de activación previo que tenga esa neurona (ver figura previa de Neurona Artificial).

## 4.2. Tipos de redes neuronales

Existen diversos tipos de modelos de redes neuronales, algunos de los más reconocidos son:

1. Perceptrón
2. Adaline
3. Perceptrón multicapa
4. Memorias asociativas
5. Máquina de Boltzmann
6. Máquina de Cauchy
7. Propagación hacia atrás (backpropagation)
8. Redes de Elman
9. Redes de Hopfield
10. Red de contrapropagación
11. Redes de neuronas de base radial

12. Redes de neuronas de aprendizaje competitivo
13. Mapas Autoorganizados (RNA) (Redes de Kohonen)
14. Crecimiento dinámico de células
15. Gas neuronal creciente
16. Redes ART (Adaptative Resonance Theory)

Para este trabajo analizaremos principalmente el perceptrón, el perceptrón multicapa, y el de retropropagación.

#### 4.2.1. El perceptrón

El Perceptrón es un tipo de red neuronal artificial desarrollado por Frank Rosenblatt<sup>3</sup>. También puede entenderse como perceptrón la neurona artificial y unidad básica de inferencia en forma de discriminador lineal que constituye este modelo de red neuronal artificial, esto debido a que el perceptrón puede usarse como neurona dentro de un perceptrón más grande u otro tipo de red neuronal artificial.<sup>4</sup>

#### Funcionamiento

En la siguiente figura se representa una neurona “artificial” que intenta modelar el comportamiento de la neurona biológica. Aquí el cuerpo de la neurona se representa como un sumador lineal de los estímulos externos  $z_j$ , seguida de una función no lineal  $y_j = f(z_j)$ . La función  $f(z_j)$  es llamada función de activación, y es la que utiliza la suma de estímulos para determinar la actividad de salida de la neurona.

---

<sup>3</sup>Rosenblatt, Frank (1957), "The Perceptron a perceiving and recognizing automaton". Report 85-460-1, Cornell Aeronautical Laboratory.

<sup>4</sup>Martin Anthony, Peter L. Bartlett Neural Network Learning: Theoretical Foundations. Cambridge University Press 2009.

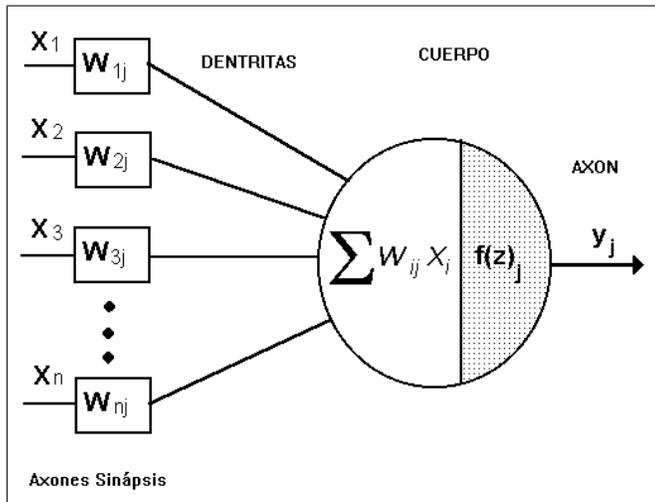


Gráfico de funcionamiento de un Perceptrón

Este modelo se conoce como perceptrón de McCulloch-Pitts, y es la base de la mayor parte de las arquitectura de las RNA que se interconectan entre sí. Las neuronas emplean funciones de activación diferentes según la aplicación: Algunas veces son funciones lineales; otras, funciones sigmoidales (p.ej. la tanh), y otras más funciones de umbral de disparo. La eficiencia sináptica se representa por factores de peso de interconexión  $w_{ij}$ , desde la neurona  $i$ , hasta la neurona  $j$ .

Los pesos pueden ser positivos (excitación) o negativos (inhibición). Los pesos junto con las funciones  $f(z)$  dictan la operación de la red neuronal. Normalmente las funciones no se modifican, de tal forma que el estado de la red neuronal depende del valor de los factores de peso (sinapsis) que se aplica a los estímulos de la neurona.

En un perceptrón, cada entrada es multiplicada por el peso  $W$  correspondiente y los resultados son sumados para evaluarlos contra el valor de umbral; si el resultado es mayor al mismo, el perceptrón se activa.

### Limitantes

El perceptrón sólo es capaz de resolver funciones definidas por un hiperplano (objeto de dimensión  $N - 1$  contenido en un espacio de dimensión  $N$ ) que corte un espacio de dimensión  $N$ . Un ejemplo de una función que no puede ser resuelta es el operador lógico *XOR*.

Una explicación más sencilla de un hiperplano sería, hablando en un plano de dos dimensiones, una línea que separa los elementos existentes en dos grupos. El perceptrón sólo puede resolver una función si todos los posibles resultados del problema pueden separarse en dos secciones es decir, que no se combinen entre sí.

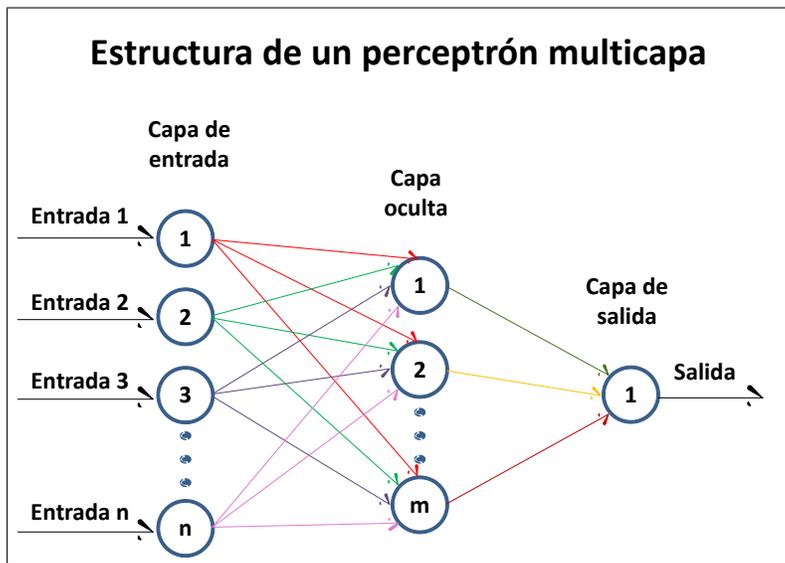
#### **4.2.2. El perceptrón multicapa**

El perceptrón multicapa es una red neuronal artificial formada por múltiples capas, esto le permite resolver problemas que no son linealmente separables. La salida de una neurona está determinada por la suma ponderada con las salidas de las neuronas de la capa anterior y una función de activación.

A diferencia del perceptrón simple, la arquitectura del multicapa permite que las entradas se encuentren total o localmente conectadas a las capas subsecuentes, esto es que la información de entrada de una neurona puede conectarse a todas las neuronas de la siguiente capa oculta, o sólo a una o algunas de ellas.

Las capas que conforman el perceptrón multicapa pueden ser de tres tipos:

1. Capa de entrada. Es la conformada por la información que alimentará la red. En las neuronas que componen esta capa no se realiza procesamiento alguno (datos de entrada).
2. Capas ocultas. Están conformadas por neuronas que reciben información de capas anteriores y envían información a capas posteriores. En éstas se realizan diversos procesamientos de la información.
3. Capa de salida. Está conformada por neuronas cuyos valores de salida corresponden a las salidas de toda la red (resultado o datos de salida).



Arquitectura de un perceptrón multicapa

Una característica importante de un perceptrón multicapa es que las funciones de transferencia deben ser derivables.

### Limitantes

La extrapolación no es perfecta. Esto significa que en caso de no ser suficiente o bueno el entrenamiento, la información de salida puede resultar imprecisa.

La existencia de mínimos locales en la función de error complica su entrenamiento, pues una vez alcanzado un valor mínimo en el entrenamiento, éste se detiene pudiendo no ser un mínimo global y no satisfacer el valor mínimo del error permitido.

### 4.2.3. Retropropagación

El perceptrón sólo es el ejemplo más elemental de una red neuronal. De hecho, no puede siquiera ser considerado una “red”, puesto que no intervienen otros elementos. Si se combinan varios perceptrones en una “capa”, y los estímulos de entrada se suman después, tendremos ya una red neuronal.

Una red neuronal muy eficaz para resolver fundamentalmente problemas de reconocimiento de patrones es la red neuronal de propagación hacia atrás, en inglés back propagation network.

Este tipo de algoritmo de aprendizaje es “supervisado” y utilizado para entrenar las redes neuronales, mediante la minimización de los errores mediante un descenso de gradiente, calculando las derivadas parciales del error respecto de los parámetros de la red neuronal.

*“Al hablar de redes de retropropagación o redes de propagación hacia atrás, hacemos referencia a un algoritmo de aprendizaje más que a una arquitectura determinada. La retropropagación consiste en propagar el error hacia atrás, es decir, de la capa de salida hacia la capa de entrada, pasando por las capas ocultas intermedias y ajustando los pesos de las conexiones con el fin de reducir dicho error. Hay distintas versiones o reglas del algoritmo de retropropagación y distintas arquitecturas conexionistas a las que pueden ser aplicados.*

*Durante mucho tiempo no se dispuso de algoritmos para entrenar las redes multicapa, y como las redes de una capa estaban muy limitadas en cuanto a lo que eran capaces de representar, el campo de las redes neuronales artificiales estaba estancado. La invención y perfeccionamiento del algoritmo de retropropagación dio un gran impulso al desarrollo de este campo. Tiene un buen fundamento matemático y a pesar de sus limitaciones ha expandido enormemente el rango de problemas donde se aplican las redes neuronales artificiales”.*<sup>5</sup>

#### 4.2.4. Algoritmo de aprendizaje

El algoritmo de aprendizaje proporciona una forma de entrenar una red multicapa con alimentación hacia adelante. Comienza alimentando los valores de la entrada de acuerdo con las siguientes ecuaciones:

$$net_i = \sum_{j \in A} O_j W_{ji} \quad \forall i : i \in B,$$

Donde  $A$  es el grupo de neuronas en una capa y  $B$  en la otra capa.  $O_j$  es la activación para la neurona  $J$ , y  $W_{ji}$  son los pesos asignados a la conexión entre las neuronas  $j$  e  $i$ .

En la ecuación anterior se toman los valores de salida y se alimenta a la siguiente capa a través de los pesos. Esta operación se realiza para cada neurona en la siguiente capa, produciendo un valor de red.

Este valor es la suma de todos los valores de activación en las neuronas de la capa anterior, y

---

<sup>5</sup>[http://rna.50webs.com/tutorial/RNA\\_backprop.html](http://rna.50webs.com/tutorial/RNA_backprop.html)

cada valor de red es aplicado ahora a la siguiente ecuación, conocida como función de activación, para producir la activación de esa neurona.

$$O_i = f(\text{net}_i) = \frac{1}{1 + e^{-\text{net}_i}}$$

Después de que todas las neuronas tienen un valor de activación asociado a un patrón de valores de entrada, el algoritmo sigue buscando errores en cada neurona que no es entrada.

Los errores encontrados para las neuronas de salida son propagados hacia atrás, a la capa anterior, para que puedan ser asignados a neuronas de las capas escondidas. Esto se calcula por:

$$\delta_i = f'(\text{net}_i) \sum_{j \in E} \delta_j W_{ij} \quad \forall i : i \in D,$$

Donde  $D$  es el grupo de neuronas en una capa que no es de entrada y  $E$  es el grupo de neuronas de la siguiente capa. Este cálculo se repite para cada capa escondida en la red.

Después de que se ha encontrado la activación y el error asociado a cada grupo de neuronas, los pesos se actualizan, primero encontrando el valor que cada peso debe modificarse. Esto se logra calculando:

$$\Delta W_{ij} = CO_i \delta_j \quad \forall i, j : i \in A, j \in B,$$

Donde  $C$ , conocida como la razón de aprendizaje, es una constante que controla el valor del cambio de los pesos y  $W_{ij}$  es el cambio de los pesos entre la neurona  $i$  y  $j$ . El peso es cambiado evaluando:

$$W_{ij_{t+1}} = W_{ij_t} + \Delta W_{ij}$$

### 4.3. Aproximación por redes neuronales

Las redes neuronales pueden catalogarse en estáticas o dinámicas. Para el caso de las estáticas no tienen elementos de realimentación ni rezagos, sus salidas son calculadas directamente de los datos de entrada mediante conexiones de realimentación (feed-forward).

El teorema de aproximación universal de George Cybenko<sup>6</sup> de 1989, afirma que la red multicapa feed-forward con una capa oculta puede aproximar una función continua arbitraria en subconjuntos compactos de  $R^n$ . Con más precisión:

Teorema.: Sea  $\phi(x)$  una función no constante, continua, acotada y monótonamente creciente. Sea  $I_m$  el hipercubo unitario en  $R^m$ .  $C(I_m)$  el espacio de funciones continuas en  $I_m$ . Entonces dada una función  $f$  en  $C(I_m)$  y  $\epsilon > 0$ , existe un entero  $N$  y constantes  $a_j, b_j$  reales,  $w_j$  en  $R^m$ ,  $j = 1, 2, \dots, N$  tales que si definimos

$$F(x) = \sum_{i=1}^N a_i \phi(w_i^T x + b_i)$$

Satisface que

$$|F(x) - f(x)| < \epsilon \forall x \text{ en } I^m$$

Para el caso de redes neuronales dinámicas, las salidas dependen no solo de la entrada actual a la red, sino también de entradas anteriores (rezagos), salidas y estado de la red.<sup>7</sup>

En este trabajo se utilizarán dos modelos dinámicos, un Modelo Neuronal Autorregresivo no Lineal (*NAR*) para el modelo univariado y un Modelo Autorregresivo no Lineal con variables exógenas (*NARX*) para el multivariado. Ambos modelos dinámicos no lineales son considerados por *MATLAB* como los más adecuados para la predicción de series temporales.

#### 4.3.1. Modelo NAR caso univariado

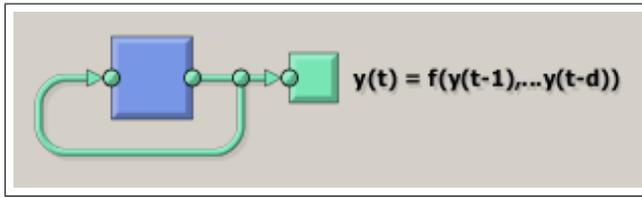
Este es un modelo neuronal autorregresivo no lineal, basado en una única serie de datos que funciona como input y output. Éste modelo predice los valores de la serie utilizando únicamente las observaciones pasadas de dicha serie a pronosticar. Esta propuesta de modelo se considera una de las mejores para analizar y predecir información de series de tiempo y ya se encuentra incluida dentro del módulo de redes neuronales de *MATLAB*, expresamente para este tipo de análisis de series temporales.

La arquitectura general se presenta a continuación:

---

<sup>6</sup>Cybenko., G. (1989) "Approximations by superpositions of sigmoidal functions", Mathematics of Control, Signals, and Systems, 2 (4), pp 303-314

<sup>7</sup>Matlab Student version R1012a.



Arquitectura general de un modelo NAR de  
Redes neuronales

Una red *NAR*, es un tipo de perceptrón multicapa, en donde una vez teniendo la serie a utilizar se definen los rezagos, el número de capas ocultas y la función de entrenamiento. La sintaxis en *MATLAB* para este efecto es: *narnet(inputDelays, hiddenSizes, trainFcn)*. Este tipo de red es de dos capas, una oculta y una de salida.

La función de entrenamiento que utiliza es “*trainlm*”, que es la función de retropropagación de Levenberg-Marquardt<sup>8</sup> y su sintaxis en *MATLAB* es: *Trainingfunction(default = trainlm)*. Esto nos regresa una red neuronal *NAR*.

El término retropropagación se refiere a la forma en que el gradiente de la función de transferencia es calculado para redes neuronales feedforward. El entrenamiento de la red involucra un ajuste de los pesos comparando la salida deseada con la respuesta de la red, de manera que se minimice el error.

La función que usualmente se utiliza para medir el error es la suma de los errores al cuadrado (*SSE*) o su promedio (*MSE*).

La función de transferencia utilizada en la capa oculta de este tipo de red es la sigmoideal, pues al ser un modelo no lineal, ésta debe ser estrictamente monótona, acotada y diferenciable, requisitos que cubre la función sigmoideal.

Para el caso de la capa de salida, la función de transferencia es lineal, la cual produce tanto su entrada como su salida.

Como el valor de la salida durante el entrenamiento se encuentra disponible para este tipo de red, se puede utilizar una arquitectura de bucle abierto, en donde el verdadero valor de la salida

<sup>8</sup>Kenneth Levenberg (1944). "A Method for the Solution of Certain Non-Linear Problems in Least Squares". Quarterly of Applied Mathematics 2: pp 164–168.

Donald Marquardt (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". SIAM Journal on Applied Mathematics 11 (2): pp 431–441. doi:10.1137/0111030.

sea utilizado en lugar de retroalimentar la salida estimada. Las dos ventajas que ofrece esto son que en primer lugar, la entrada a la red de alimentación directa (feedforward) es más precisa, y en segundo lugar que la red resultante tiene una arquitectura completamente feedforward, y por lo tanto un algoritmo más eficaz puede ser utilizado para su entrenamiento.<sup>9</sup>

En Redes Neuronales Artificiales los modelos se obtienen por prueba y error. Esto implica que no hay un procedimiento específico para conseguir la mejor red, se parte de una red inicial, y pobando diversas arquitecturas, es que puede obtenerse una red final que cumpla las expectativas del modelador.

Los elementos manipulables para la modelación son principalmente la defnición de rezagos y de neuronas ocultas. Se podrían definir los pesos de conexión entre las neuronas; pero al depender de la interconexión entre ellas, esto se puede volver muy complejo y dada la capacidad de aprendizaje de las redes, es mejor dejar que los pesos de conexión se calculen por sí mismos.

Dentro del proceso de la construcción de una red debe dividirse la información de entrada y salida en base al número de observaciones, para determinar qué porción de la misma se utiliza para entrenar, validar y probar la red. Por deaful, el módulo deja fijo el 70 % de la información para entrenamiento y el 30 % restante puede uno definir cómo se va a repartir entre validación y comprobación.

Una vez entrenada, validada y comprobada la red, se analizan los resultados. Si el resultado no fue adecuado, siempre se puede reentrenar o modificar su arquitectura (cambiar rezagos y/o número de neuronas, así como el porcentaje de división de datos para entrenamiento) para obtener un mejor resultado.

### 4.3.2. Modelo NARX caso multivariado

El modelo *NARX* también es un modelo dinámico autorregresivo no lineal pero con variables exógenas, lo cual implica que tiene una variable dependiente de una o varias variables que la explican. Al igual que en el modelo *NAR*, se considera la mejor propuesta para el análisis con el módulo de series de tiempo en redes neuronales de *MATLAB*.

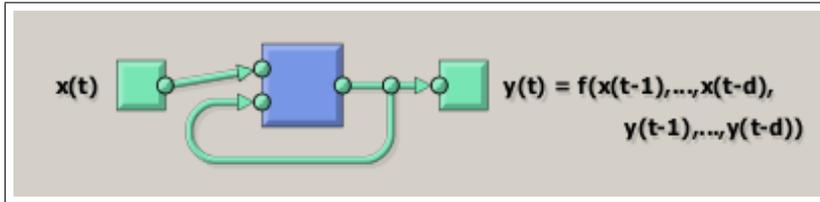
Éste modelo predice los valores de  $y_t$  en función de los valores de:

---

<sup>9</sup>MATLAB Help (F1), Neural Network Toolbox/Getting Started/Getting Started/Time Series Prediction.

$$y_{t-1} \dots y_{t-n}, x_{1_{t-1}} \dots x_{1_{t-n}}, \dots, x_{m_{t-1}} \dots x_{m_{t-n}}$$

Su arquitectura general se muestra en el siguiente gráfico:



Arquitectura general de un modelo NARX de Redes Neuronales

Similar a una red *NAR*, la red *NARX* es un tipo de perceptrón multicapa en donde una vez teniendo las series a utilizar (introducidas de forma independiente la variable dependiente de un grupo de independientes) se definen los rezagos, el número de capas ocultas y la función de entrenamiento. La sintaxis en *MATLAB* para este efecto es: `narxnet(inputDelays, hiddenSizes, trainFcn)`. Este tipo de red también se conforma de dos capas, una oculta y una de salida.

La función de entrenamiento que utiliza es “*trainlm*” que es la función de retropropagación de Levenberg-Marquardt y su sintaxis en *MATLAB* es: `Trainingfunction(default = ttrainlm)`. Esto nos regresa una red neuronal *NARX*.

Las funciones de transferencia son las mismas que en el caso de la red *NAR*: Sigmoidal en capa oculta y lineal en capa de salida.<sup>10</sup>

Una vez explicado en este capítulo cómo funcionan los modelos *NAR* y *NARX*, en el siguiente capítulo se detalla la forma en que se construyeron ambos modelos en *MATLAB*, para obtener las predicciones.

<sup>10</sup>MATLAB Help (F1), Neural Network Toolbox/Getting Started/Getting Started/Time Series Prediction.

## Capítulo 5

# Construcción, especificación, modelación y resultados de los modelos *RNA*

En este capítulo se muestra la forma en que se construyeron las *RNA* utilizando el módulo de redes neuronales de *MATLAB* para los modelos *NAR* y *NARX* de predicción de oferta maderera, para eventualmente realizar la comparación entre modelos en el capítulo 6 de éste trabajo.

Aunque el módulo de redes neuronales de *MATLAB* cuenta con una interfaz sencilla de operar, al final opté por utilizar las líneas de comandos con unas rutinas que se muestran en los anexos 5 y 6 para poder realizar la predicción de varios periodos.

### 5.1. Construcción de un modelo *RNA*

Dado que la información que se utilizó en la modelización con Redes Neuronales Artificiales es la misma que en los modelos *ARIMA* y *ARMAX*, sólo restará especificar lo siguiente en cuanto a la utilización de la información en *MATLAB*:

1. En el modelo *NAR*, la variable dependiente *VOLTRIP* se insertará al *MATLAB* como una *matriz*( $n, m$ ), con  $n = 1$  y  $m = 180$ , donde  $n$  es el número de variables involu-

cradas y  $m$  el número de observaciones de dichas variables, es decir, la serie *VOLTRIP* transpuesta. En este caso la nombré *VOTRIPT*.

2. Para el modelo *NARX* se tiene también la variable dependiente *VOLTRIP* transpuesta como una *matriz*( $n, m$ ), con  $n = 1$  y  $m = 180$ , y aparte una *matriz*( $n, m$ ) de variables explicativas transpuestas con  $n = 8$  y  $m = 180$ , con las mismas variables del modelo *ARMAX* con  $n$  y  $m$  definidas de la misma forma que en el punto anterior.

## 5.2. Especificación de un modelo *RNA*

La forma de diseñar una red neuronal se define en los siguientes pasos básicos si se utiliza el módulo wizard:

1. Colectar la información, es decir tener las bases de datos con las que se va a trabajar.
2. Ingresar al módulo de Redes Neuronales con la instrucción *nnstart*, seleccionando la opción de series de tiempo.
3. Crear la red, en este caso *NAR* o *NARX*.
4. Configurar la red, esto es definir el número de neuronas y rezagos, así como la división de datos para entrenamiento, validación y comprobación.
5. Entrenar la red.
6. Validar la red analizando su desempeño. Si los resultados no son los requeridos, reentrenar o reconfigurar la red.
7. Usar la red

Al utilizar este módulo, el problema está en que a lo sumo se puede realizar una predicción de sólo un período adelante. Por ello conviene generar el código en la opción correspondiente del mismo módulo y modificarlo especialmente, cambiando la red una vez entrenada a una arquitectura de bucle cerrado y cambiar la red a una configuración paralela para poder realizar las predicciones con varios períodos por delante.

Otra forma de utilizar el módulo de redes neuronales es utilizando las líneas de comando. Buscando la forma de cerrar el bucle de la red encontré un script de comandos para una red de predicción de varios pasos adelante, que finalmente fue el que utilicé para modelar. Los scripts se encuentran en el Anexo 5 para el caso univariado y en el Anexo 6 para el multivariado.<sup>1</sup>

## 5.3. Modelación

### 5.3.1. Modelo NAR univariado

En esta sección se realizaron varios modelos NAR probando diferentes arquitecturas con probando de 1 a 4 capas ocultas, y de 4 a 15 neuronas. Consideré 12 rezagos debido a que la frecuencia de la información es mensual. Existe una propuesta informal para definir una aproximación a un adecuado número de neuronas que es la suma de la cantidad de variables dependientes e independientes más una. Sin embargo, en este caso probé varias.

Cada una de las redes creadas se reentrenaron varias veces, para ver si se lograba obtener una con mejores resultados.

Para poder descartar más rápidamente los modelos menos eficientes, tanto en *NAR* como en *NARX*, puse en *Excel* dos tablas como las que a continuación se muestran:

---

<sup>1</sup>[www.mathworks.com/matlabcentral/answers/14970](http://www.mathworks.com/matlabcentral/answers/14970)

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO REDES UNIVARIADO 1				VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO REDES UNIVARIADO 2			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR	PERÍODO	VOLTRIP	VOLTRIPF	ERROR
2008/01	5497	5166.897	108967.925	2008/01	5497	4713.923	613210.096
2008/02	4871	5351.642	231016.859	2008/02	4871	5321.213	202692.119
2008/03	5380	5608.021	51993.648	2008/03	5380	5359.755	409.841
2008/04	5355	5312.947	1768.417	2008/04	5355	5199.129	24295.857
2008/05	5708	5641.804	4381.860	2008/05	5708	5467.858	57668.104
2008/06	5278	5175.091	10590.338	2008/06	5278	5224.565	2855.319
2008/07	5435	5360.331	5575.466	2008/07	5435	5548.418	12863.653
2008/08	5253	5450.855	39146.787	2008/08	5253	6040.532	620207.200
2008/09	4971	5270.306	89584.292	2008/09	4971	5747.959	603665.837
2008/10	5820	5397.921	178150.408	2008/10	5820	5105.010	511210.473
2008/11	5417	5157.212	67489.818	2008/11	5417	5371.654	2056.221
2008/12	4514	4833.152	101858.035	2008/12	4514	5332.549	670022.098
Estadísticos de ajuste de la estimación				Estadísticos de ajuste de la estimación			
SSE=			890523.853	SSE=			3,321,156.82
MSE=			74,210.32	MSE=			276,763.07
ME=			272.416	ME=			526.083
CORRELACIÓN=			0.616	CORRELACIÓN=			-0.242
% ME de la media de voltrip=			5.14809456	% ME de la media de voltrip=			9.94187797

Cuadros comparativos de modelos de redes para descarte

Es posible definir de forma visual si el modelo previo es mejor o peor que el nuevo mediante la siguiente fórmula condicional en Excel.

$$SI (MSE2 < MSE1, "2 mejor que 1", "1 sigue mejor")$$

Con un formato condicional para la celda que contiene la fórmula, de tal forma que si “2 es mejor que 1” la celda se muestra en verde.

Entonces los valores de *VOLTRIPF* de la tabla 2 se sustituyen en la tabla 1.

Dado que los valores pronosticados en redes salen en un arreglo de (1, 12) y en la tabla los utilizo como un arreglo (12, 1), también en *Excel* utilicé la fórmula de trasponer los datos que pego de *MATLAB*, y a la vez cada celda de datos *VOLTRIPF* de la tabla dos toman los valores del arreglo traspuesto.

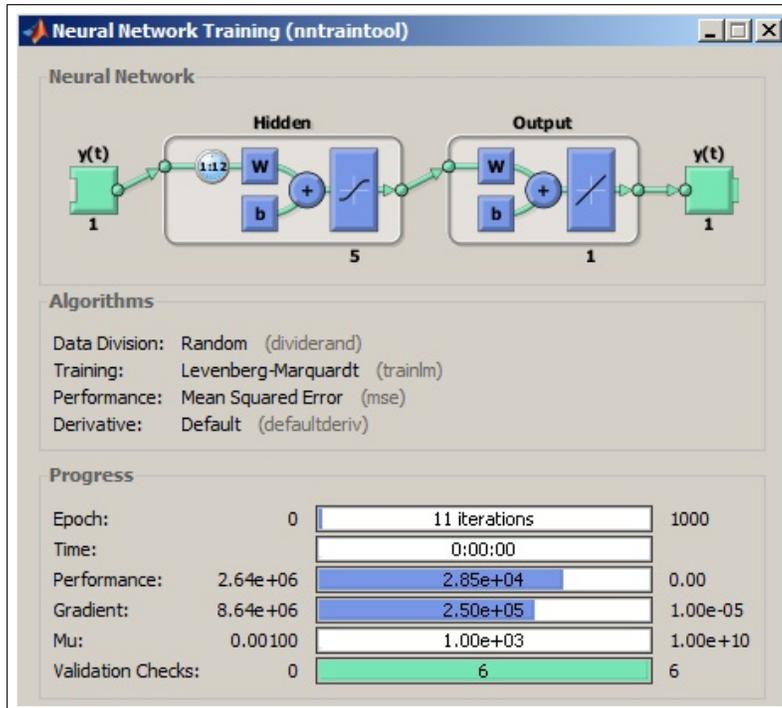
De esta forma se facilita la introducción de los nuevos valores pronosticados de cada modelo calculado.

A continuación, únicamente presento el resumen del mejor modelo obtenido en redes neuronales, el criterio de descarte entre modelos fue el que tuviese el menor promedio de errores cuadrados (*MSE*) de las observaciones correspondientes al año 2008, tanto observadas como

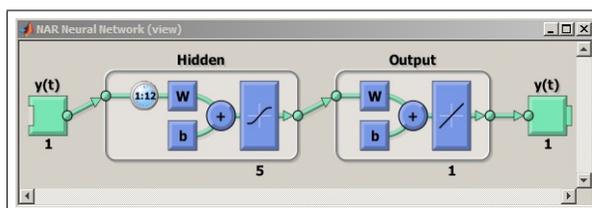
pronosticadas.

Se presenta el seleccionado debido a que se corrieron alrededor de 100 modelos obteniendo a veces mejores, a veces peores, y descartando los peores.

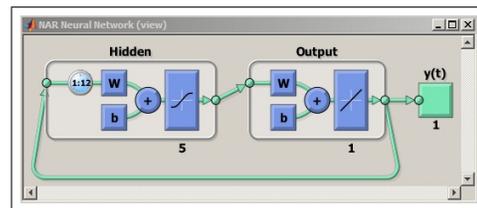
El mejor modelo obtenido fue uno con 12 rezagos y 5 neuronas. Presento a continuación las gráficas pertinentes a este modelo:



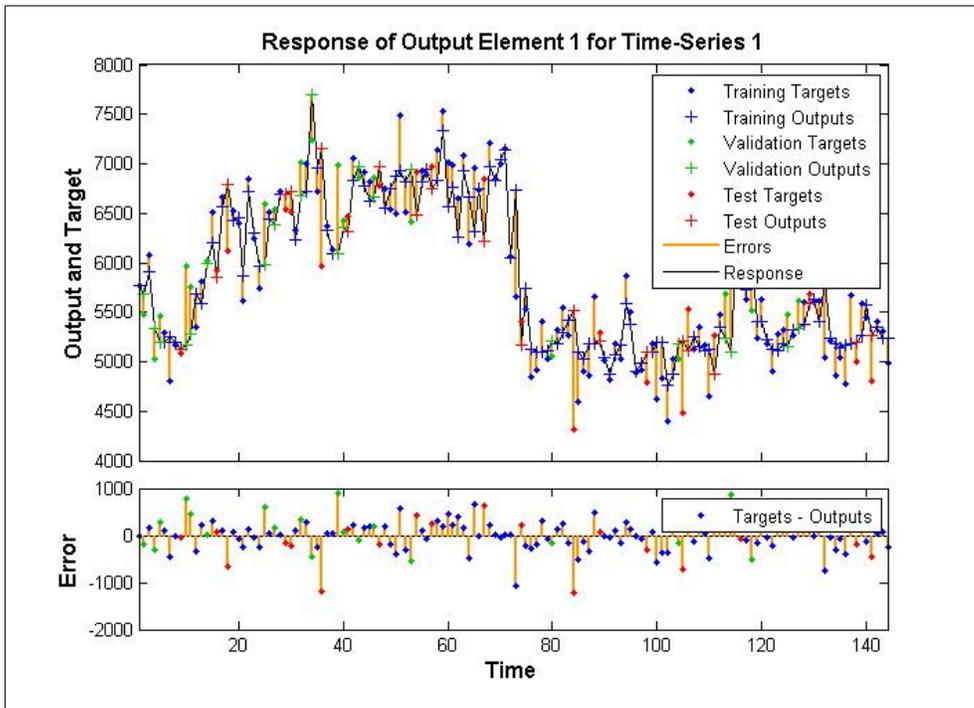
Arquitectura de la red y progreso de entrenamiento



Arquitectura con bucle abierto para  
entrenamiento



Arquitectura con bucle cerrado  
para predicción de varios periodos



Resumen gráfico del desempeño de la red

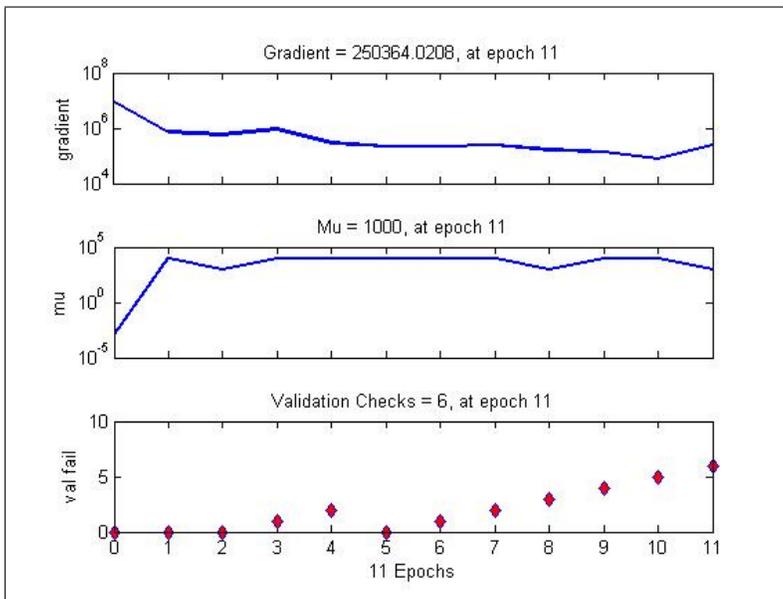


Gráfico del estado del entrenamiento

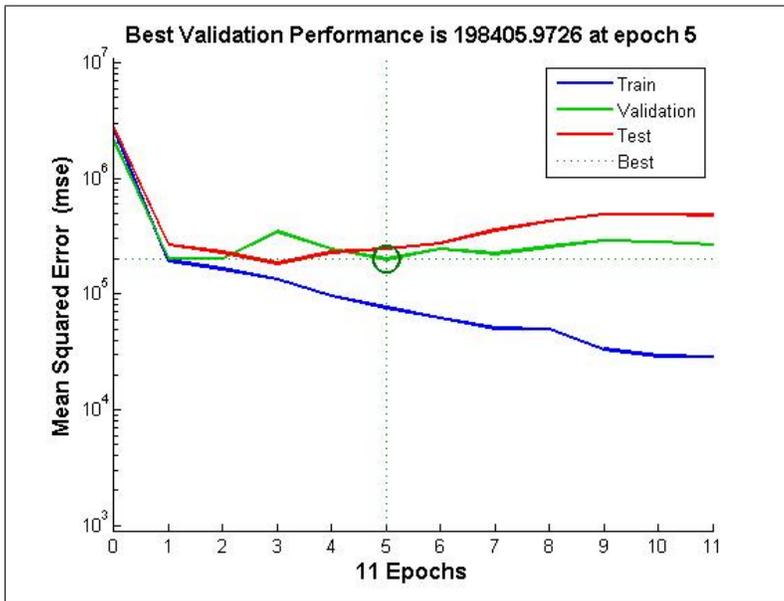
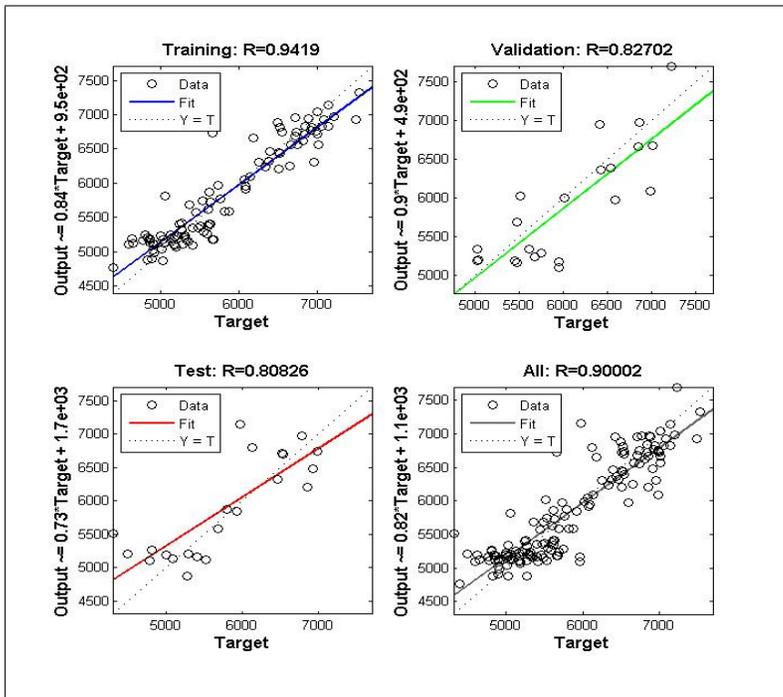
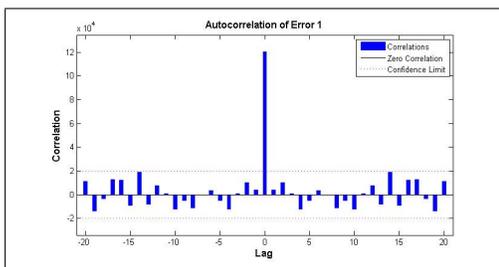


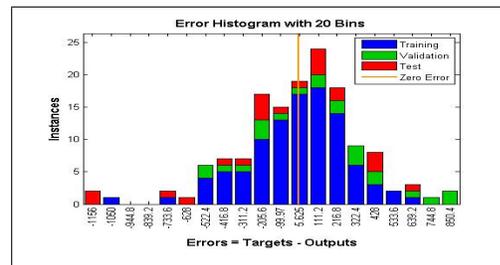
Gráfico del desempeño de entrenamiento, validación y comprobación



Gráfica de la regresión durante entrenamiento, validación comprobación y general



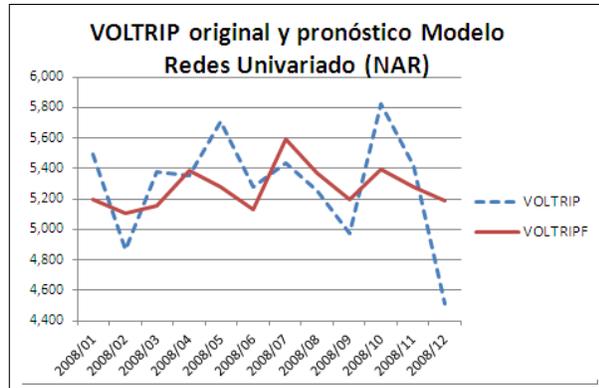
Autocorrelación del error



Histograma del error

A continuación, presento los valores de la predicción y la gráfica entre los valores esperados y calculados de la serie VOLTRIP.

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO REDES UNIVARIADO (NAR)			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,199.210	88,678.884
2008/02	4,871	5,102.532	53,607.067
2008/03	5,380	5,156.930	49,760.225
2008/04	5,355	5,386.400	985.960
2008/05	5,708	5,280.837	182,468.229
2008/06	5,278	5,134.458	20,604.306
2008/07	5,435	5,589.951	24,009.812
2008/08	5,253	5,369.473	13,565.960
2008/09	4,971	5,198.654	51,826.344
2008/10	5,820	5,393.879	181,579.107
2008/11	5,417	5,277.875	19,355.766
2008/12	4,514	5,192.450	460,294.402
Estadísticos de ajuste de la estimación			
	SSE=	1,146,736.06	
	MSE=	95,561.34	
	ME=	309.130	
	CORRELACIÓN=	0.452	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo RNA Univariado (NAR)

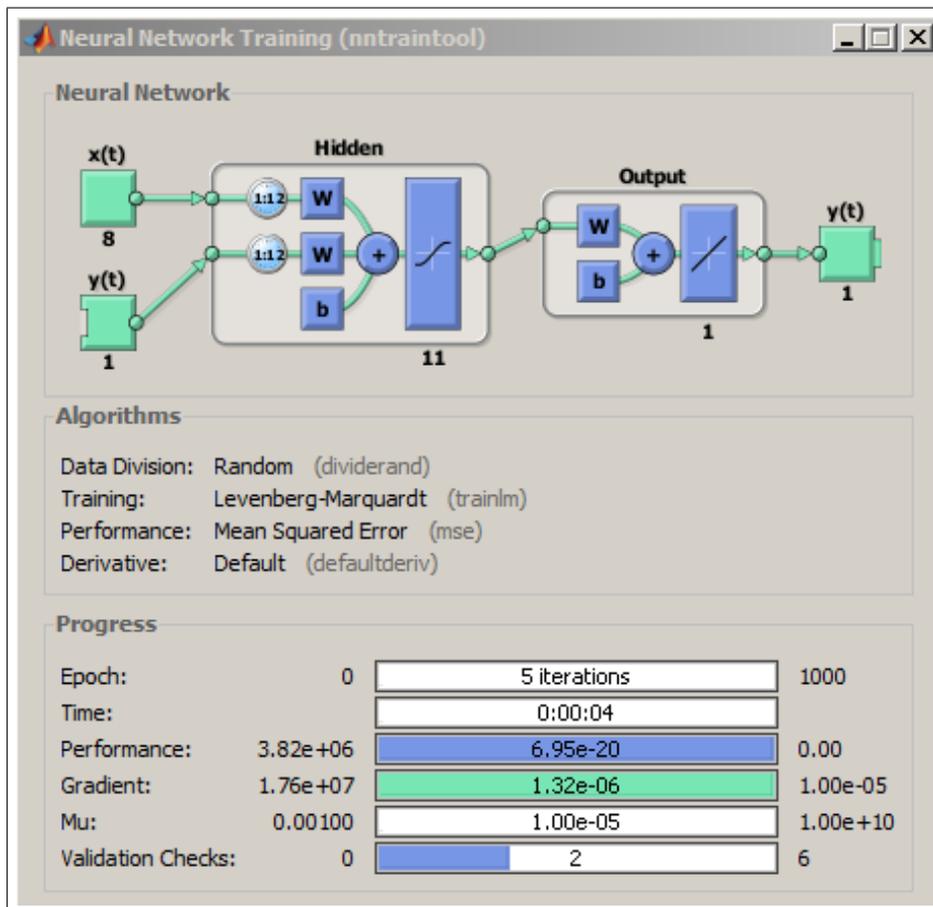
Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los errores cuadrados es 309.130, el cual representa un 5.84 % del valor de la media de los datos observados, concluyo que en este modelo tengo un error aproximado del 6 %.

### 5.3.2. Modelo NARX multivariado

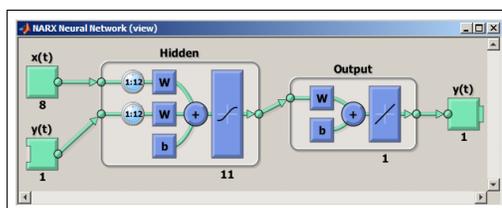
Al igual que con el proceso realizado en el modelo *NAR*, probé con diferentes arquitecturas en cuanto al número de neuronas seleccionadas, dejando fijo el número de rezagos para el modelo *NARX*.

Por las razones mencionadas en el modelo *NAR*, sólo presento el resumen del mejor modelo seleccionado después de haber probado una cantidad considerable de cada una de las arquitecturas seleccionadas, que van desde 5 a 15 neuronas en la capa oculta.

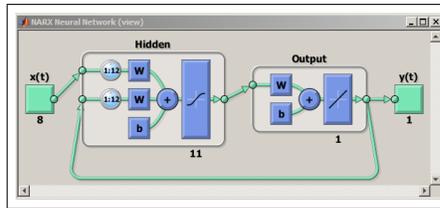
El mejor modelo obtenido fue uno con 12 rezagos y 11 neuronas. A continuación muestro el resumen gráfico de dicho modelo:



Arquitectura de la red y progreso de entrenamiento del modelo NARX



Arquitectura con bucle abierto para entrenamiento



Arquitectura con bucle cerrado para predicción de varios períodos

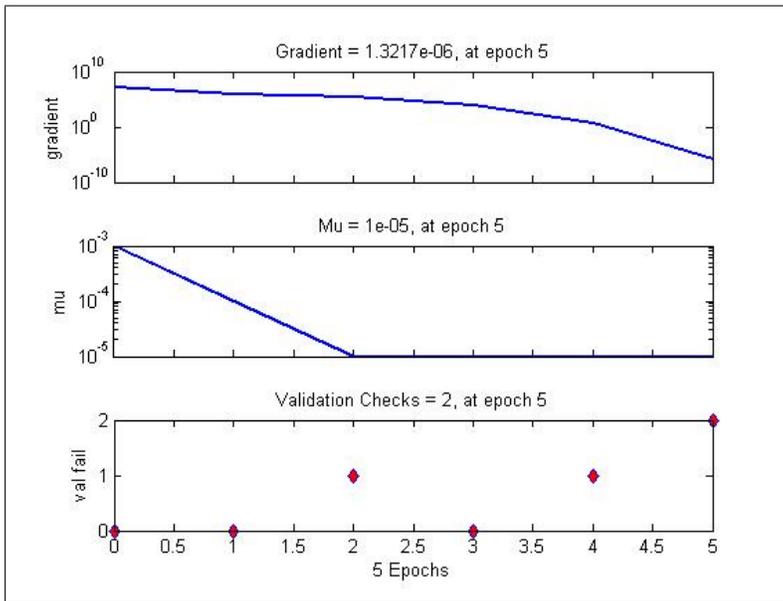


Gráfico del estado del entrenamiento

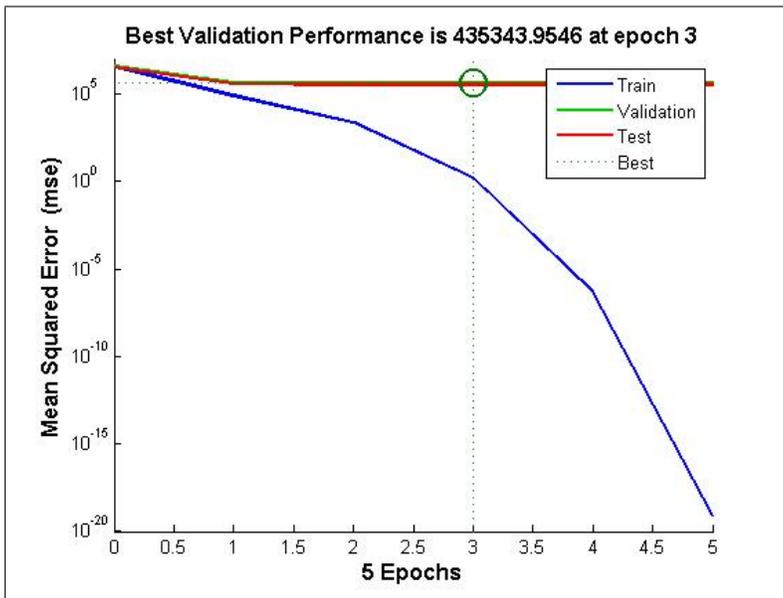
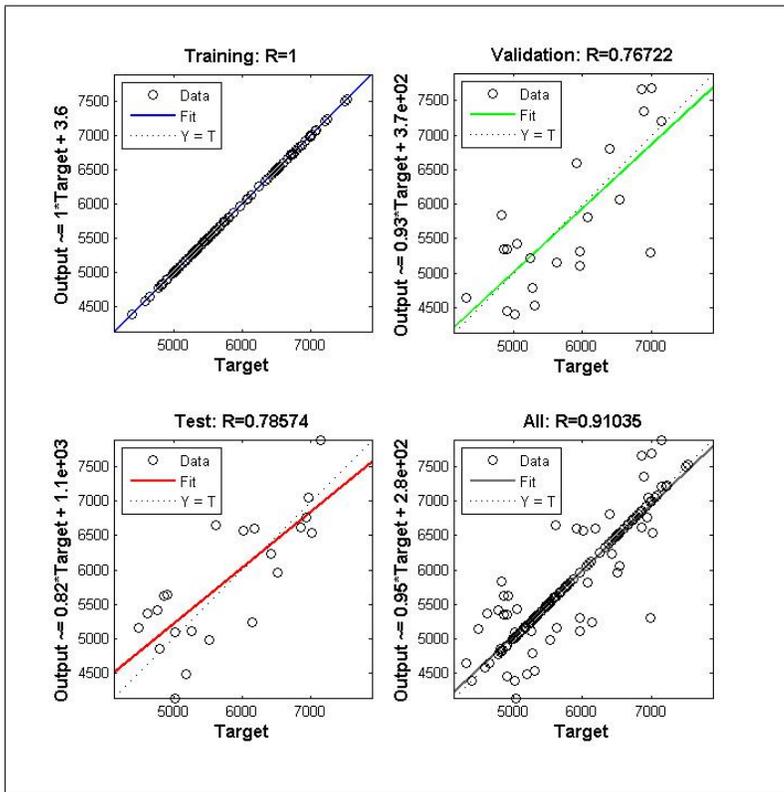
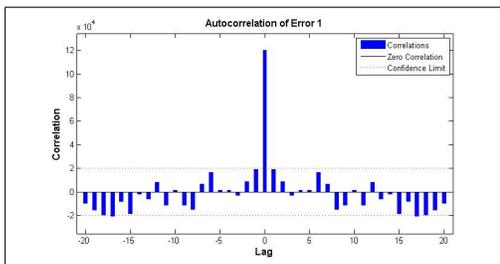


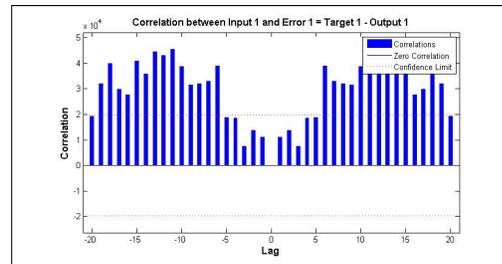
Gráfico del desempeño de entrenamiento, validación y comprobación



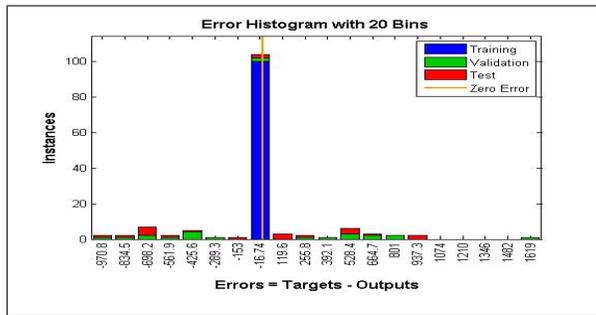
Gráfica de la regresión durante entrenamiento, validación comprobación y general



Autocorrelación del error



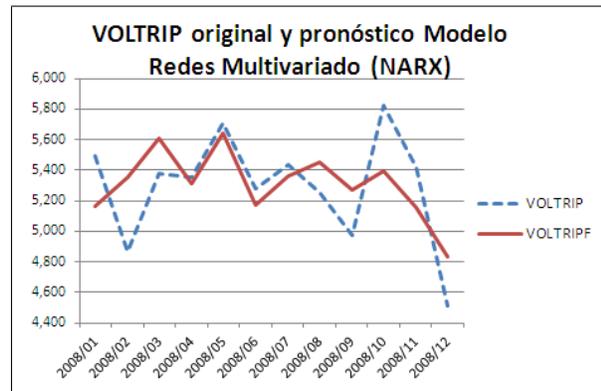
Correlación= Target1 - Output1



Histograma del error

A continuación presento los valores de la predicción y la gráfica entre los valores esperados y calculados de la serie VOLTRIP.

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO REDES MULTIVARIADO (NARX)			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,166.897	108,967.991
2008/02	4,871	5,351.642	231,016.732
2008/03	5,380	5,608.021	51,993.576
2008/04	5,355	5,312.947	1,768.455
2008/05	5,708	5,641.804	4,381.910
2008/06	5,278	5,175.091	10,590.262
2008/07	5,435	5,360.331	5,575.460
2008/08	5,253	5,450.855	39,146.601
2008/09	4,971	5,270.306	89,584.082
2008/10	5,820	5,397.921	178,150.682
2008/11	5,417	5,157.212	67,489.805
2008/12	4,514	4,833.152	101,857.999
Estadísticos de ajuste de la estimación			
	SSE=	890,523.56	
	MSE=	74,210.30	
	ME=	272.416	
	CORRELACIÓN=	0.616	



Donde VOLTRIPF son los valores pronosticados

Gráfica de VOLTRIP y VOLTRIPF Modelo RNA Multivariado (NARX)

Dado que la media de las observaciones es 5,291.58 y la media de la raíz de los errores cuadrados es 272.416, el cual representa un 5.15 % del valor de la media de los datos observados, concluyo que en este modelo tengo un error medio aproximado del 5 %.

En este capítulo se explicó cómo se realizaron los modelos de redes neuronales artificiales *NAR* y *NARX*, y se presentaron sus respectivas predicciones.

## Capítulo 6

# Comparación de resultados y conclusiones

Este capítulo se compone de dos secciones principales. La primera es la comparación de los resultados entre los modelos univariados econométrico *ARIMA* y de redes neuronales *NAR*, como los multivariados econométrico *ARMAX* y de redes neuronales *NARX*. La segunda parte se refiere a las conclusiones de la comparación de modelos.

### 6.1. Comparación de resultados

#### 6.1.1. Modelos univariados

A continuación muestro la comparación de los modelos *ARIMA* y *NAR* univariados. El criterio de selección del mejor modelo se llevó a cabo por comparación de menor esperanza del error cuadrático, esto es, menor varianza de los residuales.

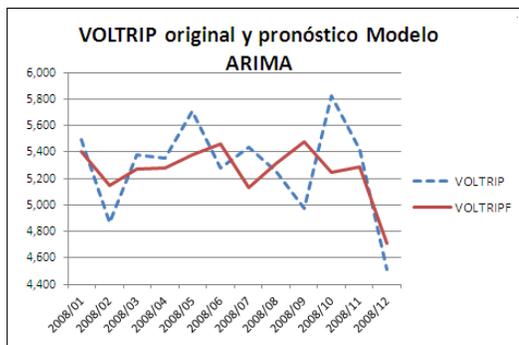
VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO ARIMA				VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO NAR			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO	PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,401.646	9,092.324	2008/01	5,497	5,199.210	88,678.884
2008/02	4,871	5,149.506	77,565.735	2008/02	4,871	5,102.532	53,607.067
2008/03	5,380	5,267.437	12,670.522	2008/03	5,380	5,156.930	49,760.225
2008/04	5,355	5,283.139	5,163.996	2008/04	5,355	5,386.400	985.960
2008/05	5,708	5,374.529	111,203.056	2008/05	5,708	5,280.837	182,468.229
2008/06	5,278	5,462.995	34,223.160	2008/06	5,278	5,134.458	20,604.306
2008/07	5,435	5,126.789	94,994.174	2008/07	5,435	5,589.951	24,009.812
2008/08	5,253	5,315.985	3,967.139	2008/08	5,253	5,369.473	13,565.960
2008/09	4,971	5,480.827	259,923.300	2008/09	4,971	5,198.654	51,826.344
2008/10	5,820	5,247.994	327,191.320	2008/10	5,820	5,393.879	181,579.107
2008/11	5,417	5,287.091	16,876.470	2008/11	5,417	5,277.875	19,355.766
2008/12	4,514	4,712.008	39,207.000	2008/12	4,514	5,192.450	460,294.402
Estadísticos de ajuste de la estimación				Estadísticos de ajuste de la estimación			
SUMA		992,078.196		SUMA		1,146,736.06	
MSE=		82,673.183		MSE=		95,561.34	

$MSE_{ARIMA} < MSE_{NAR}$   
**MEJOR MODELO = ARIMA**

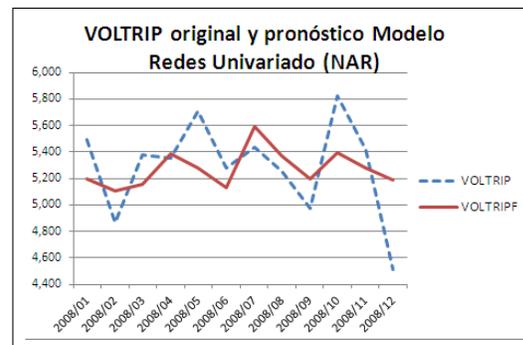
Comparativo de resultados entre modelos

*ARIMA* y *NAR* univariados

Las gráficas de los modelos respectivos se muestran a continuación:



Gráfica de valores de la serie original VOLTRIP y pronosticada VOLTRIPF del modelo ARIMA univariado



Gráfica de valores de la serie original VOLTRIP y pronosticada VOLTRIPF del modelo NAR univariado

Observando lo anterior, pareciera ser que en este caso univariado, el modelo econométrico *ARIMA* es mejor que el *NAR* de redes neuronales artificiales.

Sin embargo, para determinar si son significativamente diferentes ambos resultados, se realizó una prueba de hipótesis de comparación de varianzas de los residuales de las doce fechas del periodo de pronóstico.

Se utilizó en *Excel* una prueba *F* bilateral, con  $\alpha = 0.01$ , al ser bilateral,  $\alpha = \frac{\alpha}{2} = 0.005$ , con  $gl = 11$  tanto en el numerador como en el denominador.

Las hipótesis a probar con  $F_{0.995,11,11}$  son:

$$H_0 : \sigma_{ARIMA}^2 = \sigma_{NAR}^2$$

$$H_A : \sigma_{ARIMA}^2 \neq \sigma_{NAR}^2$$

El criterio de decisión está dado por: si  $F_c > F_{0.995,11,11}$ , se rechaza  $H_0$ .

La tabla resumen de la prueba  $F$  de *Excel* se muestra a continuación:

Prueba F para varianzas de dos muestras		
$\alpha=0.01$		
	Error Cuadrático NAR	Error Cuadrático ARIMA
MSE	95,569.63475	82,673.18298
Observaciones	12	12
Grados de libertad	11	11
$F_c = \text{MSE}_{NAR} / \text{MSE}_{ARIMA}$	1.155993169	
Valor crítico para $F_{0.995,11,11}$	5.319667079	

Resumen prueba  $F_{0.995,11,11}$  bilateral para comparación de varianzas del error cuadrático *ARIMA* y *NAR*

Por lo que se concluye que:

Como  $F_c = 1.1559 < F_{0.995,11,11} = 5.3196$ .

Dada la evidencia muestral, no se rechaza la hipótesis nula de que la esperanza del error cuadrático de los modelos *ARIMA* y *NAR* son iguales, con 1 % de significancia.

Por lo que sus varianzas son estadísticamente iguales, y por tanto ambos modelos son igualmente efectivos en sus predicciones.

### 6.1.2. Modelos multivariados

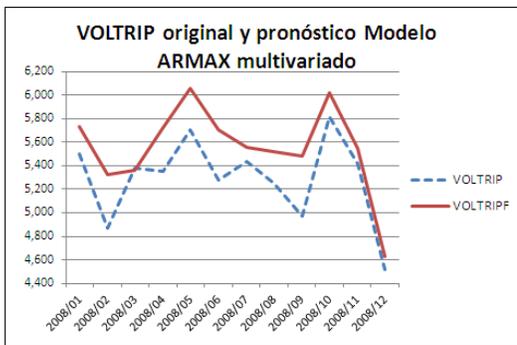
A continuación muestro la comparación de los modelos *ARMAX* y *NARX* multivariados, el criterio de selección del mejor modelo se llevó a cabo de la misma manera que en la comparación de los modelos anteriores.

VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO ARMAX				VALORES ORIGINALES Y PRONOSTICADOS DE VOLTRIP MODELO NARX			
PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO	PERÍODO	VOLTRIP	VOLTRIPF	ERROR CUADRÁTICO
2008/01	5,497	5,736.024	57,132.473	2008/01	5,497	5,166.897	108,967.991
2008/02	4,871	5,325.582	206,644.795	2008/02	4,871	5,351.642	231,016.732
2008/03	5,380	5,357.446	508.683	2008/03	5,380	5,608.021	51,993.576
2008/04	5,355	5,718.799	132,349.712	2008/04	5,355	5,312.947	1,768.455
2008/05	5,708	6,055.471	120,736.096	2008/05	5,708	5,641.804	4,381.910
2008/06	5,278	5,705.338	182,617.766	2008/06	5,278	5,175.091	10,590.262
2008/07	5,435	5,557.841	15,089.911	2008/07	5,435	5,360.331	5,575.460
2008/08	5,253	5,523.204	73,010.202	2008/08	5,253	5,450.855	39,146.601
2008/09	4,971	5,481.047	260,147.942	2008/09	4,971	5,270.306	89,584.082
2008/10	5,820	6,020.620	40,248.384	2008/10	5,820	5,397.921	178,150.682
2008/11	5,417	5,546.244	16,704.012	2008/11	5,417	5,157.212	67,489.805
2008/12	4,514	4,627.266	12,829.187	2008/12	4,514	4,833.152	101,857.999
Estadísticos de ajuste de la estimación				Estadísticos de ajuste de la estimación			
SUMA		1,118,019.16		SUMA		890,523.56	
MSE=		93,168.26		MSE=		74,210.30	

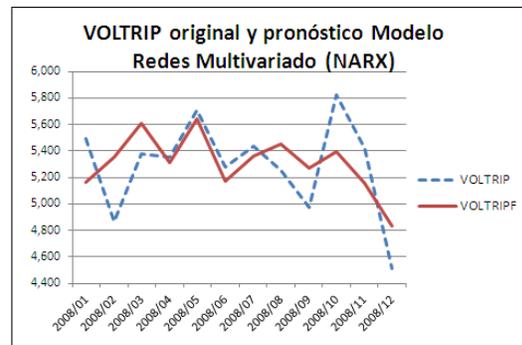
$MSE_{NARX} < MSE_{ARMAX}$   
**MEJOR MODELO = NARX**

Comparativo de resultados entre modelos  
*ARMAX* y *NARX* multivariados

Las gráficas de los modelos respectivos se muestran a continuación:



Gráfica de valores de la serie original  
 VOLTRIP y pronosticada VOLTRIPF  
 del modelo ARMAX multivariado



Gráfica de valores de la serie original  
 VOLTRIP y pronosticada VOLTRIPF  
 del modelo NARX multivariado

Comparando los *MSE* de *ARMAX* y *NARX*, pareciera que en el caso multivariado, el modelo de redes neuronales artificiales *NARX* es mejor que el econométrico *ARMAX*.

Al igual que en el caso univariado, para determinar si son significativamente diferentes ambos resultados, se realizó también una prueba de hipótesis de comparación de varianzas de los residuales de las doce fechas del periodo de pronóstico. Se utilizó también en *Excel* una prueba *F* bilateral, con  $\alpha = 0.01$ , al ser bilateral,  $\alpha = \frac{\alpha}{2} = 0.005$ , con  $gl = 11$  tanto en el numerador como en el denominador.

Las hipótesis a probar con  $F_{0.995,11,11}$  son:

$$H_0 : \sigma_{ARMAX}^2 = \sigma_{NARX}^2$$

$$H_A : \sigma_{ARMAX}^2 \neq \sigma_{NARX}^2$$

El criterio de decisión está dado por: si  $F_c > F_{0.995,11,11}$ , se rechaza  $H_0$ .

La tabla resumen de la prueba  $F$  de *Excel* se muestra a continuación:

Prueba F para varianzas de dos muestras		
$\alpha=0.01$		
	Error Cuadrático ARMAX	Error Cuadrático NARX
MSE	93,168.26354	74,210.29627
Observaciones	12	12
Grados de libertad	11	11
$F_c = \text{MSE}_{ARMAX} / \text{MSE}_{NARX}$	1.255462762	
Valor crítico para $F_{0.995,11,11}$	5.319667079	

Resumen prueba  $F_{0.995,11,11}$  bilateral para comparación de varianzas del error cuadrático *ARMAX* y *NARX*

Por lo que se concluye que:

Como  $F_c = 1.2554 < F_{0.995,11,11} = 5.3196$ .

Dada la evidencia muestral, no se rechaza la hipótesis nula de que la esperanza del error cuadrático de los modelos *ARMAX* y *NARX* son iguales, con 1 % de significancia.

Por lo que sus varianzas son estadísticamente iguales, y por tanto ambos modelos son igualmente efectivos en sus predicciones.

## 6.2. Conclusiones y perspectivas

Proponiendo un modelo de oferta de madera en el mercado mexicano, se realizaron las predicciones correspondientes a los doce meses del año 2008, utilizando como criterio de selección de mejor modelo el que tuviese el menor varianza en su error cuadrático. Realicé la prueba de hipótesis de comparación de esperanzas del error cuadrático, para constatar que efectivamente, la varianza entre ambos modelos sea significativamente diferente, y por tanto la selección del mejor modelo sea correcta.

Las hipótesis centrales a probar en esta tesis de maestría en matemáticas aplicadas son:

$H_0$  : Las RNA ofrecen mejores predicciones que la econometría

$H_A$  : Las RNA no ofrecen mejores predicciones que la econometría

### 6.2.1. Caso univariado

Este modelo predice los valores futuros de la serie *VOLTRIP* que determiné como oferta de madera, la cual depende de sí misma a lo largo del período 1994-2007.

El pronóstico realizado fue de las 12 observaciones correspondientes a Enero - Diciembre de 2008.

En la sección previa se analizó la efectividad de una predicción univariada entre un modelo econométrico *ARIMA* y uno de redes neuronales artificiales *NAR*.

Dado el criterio de selección de mejor modelo por menor *Mean Squared Error (MSE)*, se concluye que el modelo *ARIMA* parece predecir mejor que el *NAR* debido a que su *MSE* es menor.

Sin embargo, para determinar que la diferencia entre ambos *MSE* se considere significativa, realicé una prueba de hipótesis de comparación de la esperanza del error cuadrático entre ambos modelos.

El resultado de la prueba arroja que, tanto el modelo *ARIMA* como el *NAR* son igualmente efectivos en su predicción, dado que no se rechaza que sus varianzas son iguales, con un 1 % de nivel de significancia.

Ello me hace rechazar la hipótesis nula de este trabajo, para el caso del modelo univariado.

### 6.2.2. Caso multivariado

Para el modelo multivariado, la serie *VOLTRIP* depende de sí misma y depende también de otras variables que se encuentran relacionadas con ella.

Estas variables representan la demanda de madera utilizada como insumo en la producción de otros bienes, luego entonces se esperaría que las variaciones en la demanda de madera modifiquen la oferta de la misma.

Para el caso multivariado se aplicaron los mismos criterios de evaluación (menor *MSE*) a

los modelos *ARMAX* (econométrico) y *NARX* (de redes neuronales artificiales).

La conclusión para el caso multivariado, por inspección de *MSE*, es que el *NARX* parece predecir mejor que el *ARMAX*.

De la misma manera que en el caso univariado, se realizó su respectiva prueba de hipótesis para asegurar que la diferencia entre la esperanza del error cuadrático de las predicciones fuese significativo.

Con la prueba de análisis de varianza, se concluye que no existe diferencia estadísticamente significativa entre las predicciones de ambos modelos, pues no se rechaza que ambas varianzas sean iguales con un 1 % de nivel de significancia.

Con lo anterior, en éste caso no hay una evidencia estadísticamente significativa para concluir que se acepta o rechaza la hipótesis nula principal de este trabajo, que establece que las *RNA* ofrecen una mejor predicción que el modelo econométrico, pues sus resultados fueron prácticamente iguales para el caso multivariado.

### **6.2.3. Conclusión general**

Dado que en ambos casos se realizó modelación y ésta a su vez se realiza mediante eurísticos, porque hasta este momento no es seguro que se determine que el proceso va a encontrar “*El MODELO*” perfecto, y depende mucho de la habilidad del modelador en muchos aspectos.

Para este trabajo, se concluye que ambas técnicas, son igualmente efectivas al momento de realizar predicciones, pues la diferencia entre ambas no resulta estadísticamente significativa como para dar una preferencia entre una técnica y otra.

Siempre existe la posibilidad de encontrar un modelo que prediga mejor especialmente en el caso de las *RNA*, ya que al iniciar el entrenamiento con valores aleatorios en los pesos de la función de activación, se puede llegar a obtener un mejor modelo que el que se haya obtenido previamente. Para el caso de los modelos econométricos, al utilizar un eurístico para seleccionar los términos *AR* y *MA*, podría quedar alguno fuera de la selección que tenga un impacto bien sea marginal o considerable en el resultado.

Por otra parte, una ventaja de los modelos econométricos, es que los coeficientes que se obtienen son de interpretación inmediata y proveen información adicional desde la perspectiva económica. En el caso de las *RNA*, su ventaja es que pueden reducir el tiempo de modelación

para el caso de los pronósticos aunque no aporten más información adicional.

Considero que las Redes Neuronales Artificiales (*RNA*) pueden ser una muy buena herramienta para los econométricos, ya que pronostican bien, rápidamente, y son muy fáciles de usar, dando para el caso de los pronósticos tan buenos resultados como los obtenidos mediante métodos econométricos tradicionales.

#### **6.2.4. Perspectivas**

Los modelos econométricos tradicionales de predicción, a pesar de llevar mayor trabajo por parte del analista, tienen tan buen desempeño como las *RNA*, pero además permiten estudiar y comprender los coeficientes de los modelos y sus implicaciones económicas cuando requieren ese tipo de análisis.

Sin embargo, en el caso de sólo requerir la predicción sin el análisis de coeficientes de los modelos, las *RNA* facilitan de forma considerable la labor de predecir.

Habría también que probar con otros sectores económicos para ver si se obtienen resultados similares a los de este trabajo.

Otro tema interesante sería ver qué sucede en predicciones de largo plazo si se cuenta con suficiente información, tanto para modelar como para comprobar los resultados.

Otro posible análisis sería hasta qué nivel impacta mayor información histórica sobre el cálculo de un pronóstico, y cuál es su magnitud. En otras palabras, qué tanto más preciso es un pronóstico cuando la o las series que nos dan la información cubren un mayor periodo.

# Anexos

# Anexo 1: Base de datos final

BASE DE DATOS DEFINITIVA													
obs	VOLTRIP	VALTRIPDEF	VALCONSTDEF	VALMUEBDEF	VALCOLCHDEF	VALPAPEDEF	VALCARTDEF	VALENCARTDEF	VALEDPERDEF	VALEDLIBDEF	VALIMPDEF	IMWDEF	CETE28DEF
1994M01	4529	1709.28594	1351.99941	1955.17079	2202.61998	17708.93600	2522.36485	12228.53461	4394.52769	2132.70159	4828.86918	0.98848	0.00197
1994M02	4430	1756.26450	1444.24569	1859.10911	2260.65397	17026.66878	2620.69126	12056.92866	4184.97701	3368.55209	4569.60850	1.06376	0.00506
1994M03	4929	1919.83486	1561.40708	2031.41061	1961.76618	19438.73851	2457.41742	12579.95981	4802.55480	3723.00811	4862.31992	1.06205	0.00433
1994M04	4755	1901.45860	1563.76039	1861.05375	2084.48250	18985.27200	2692.18449	13377.33456	4716.27656	4880.74844	4975.90874	1.13287	0.00774
1994M05	5153	2071.46695	1433.70244	1997.82292	1993.98184	19099.81860	2923.69002	13558.37643	4699.96521	7246.18652	5014.60816	1.12396	0.00826
1994M06	5187	1999.05732	1292.75465	2051.66395	1952.02209	18016.12537	2846.95033	13454.20916	4642.86779	8700.11523	4851.18180	1.02794	0.00703
1994M07	5208	1827.11198	1297.26961	2224.59665	1864.64083	17934.74855	3021.05218	13134.29385	4803.05816	8096.55346	4976.54789	1.02115	0.00785
1994M08	5323	2029.66502	1161.05603	2154.25794	2086.00416	17651.69146	3040.62741	13753.70132	4767.26794	6042.32897	5964.21155	1.01632	0.00573
1994M09	5277	1852.93336	856.35232	2108.95870	2133.97446	17141.74592	2923.46514	12702.40465	4392.33823	5653.95396	5577.63356	0.83101	0.00509
1994M10	6313	2170.17065	594.76129	2420.19440	2398.79488	16972.61823	3095.05464	13511.25474	4526.95066	3187.87821	5389.80105	0.88694	0.00697
1994M11	6361	2170.83710	612.52645	2657.70946	2424.15272	16273.42553	2756.48561	14066.80602	4405.40786	1971.51776	5561.55005	0.99766	0.00641
1994M12	5823	2030.00694	560.26619	2916.05752	2518.69586	15872.38275	2638.22854	12737.14101	4363.68299	2212.96523	5059.16511	1.49998	0.00617
1995M01	5764	2282.80934	627.85620	1984.46238	2270.41942	23333.34378	2934.80821	15823.47638	4286.69230	2335.71486	4253.31808	1.02889	-0.02621
1995M02	5480	2245.39337	641.46514	1784.44265	2134.07440	19760.82150	2723.40861	13397.35625	4017.61017	3871.33428	4179.22052	1.01596	-0.00953
1995M03	6081	2521.88197	693.31750	1811.08647	1977.58340	21075.37376	3298.49246	13162.30227	4158.49376	5341.03856	5735.47125	0.97174	0.00353
1995M04	5028	1993.55749	650.16136	1186.58047	1119.68959	21316.50734	2935.35294	10980.25981	3509.02503	3252.16131	3662.44577	0.98944	-0.02472
1995M05	5461	2131.92105	680.00538	1205.86323	1296.37975	19360.65445	2932.71999	12344.03924	3473.58707	5866.09378	4673.80041	1.05850	0.00698
1995M06	5300	2177.90228	615.41331	1148.04846	1019.27627	18466.00019	3023.19243	12430.96513	3273.52031	5347.46515	4574.05897	0.88903	0.00964
1995M07	4804	1772.91435	526.87643	1156.98451	1049.32847	17686.22523	3327.14980	12229.53424	3477.18041	7306.59159	4780.55544	0.99666	0.01503
1995M08	5161	1924.55650	698.17902	1344.34265	1255.76608	16951.05143	3124.06656	12640.08996	3436.89165	4607.84561	4802.90363	0.90837	0.01261
1995M09	5085	1850.64600	646.79354	1497.01161	1238.26149	1743.74860	2743.28183	11669.20271	3324.69519	3382.13501	4344.09282	0.74273	0.00867
1995M10	5959	2130.93814	775.56286	1520.21368	1586.62943	16243.76240	2645.45637	12397.08118	3205.78159	2067.94041	4428.17026	0.85258	0.01216
1995M11	5750	2137.75299	607.06955	1750.22032	1732.89019	17070.03561	2743.18303	12089.27802	3161.92060	1887.03756	5026.00483	0.80423	0.01506
1995M12	5355	2041.19252	984.54885	1947.83221	1592.44219	17945.24215	2460.35676	11313.55882	3108.14243	2783.62112	4289.68054	1.07778	0.00218
1996M01	5815	2305.86374	838.14162	1389.02069	1522.65378	18267.21408	2305.71058	13669.38362	3114.87411	1247.08561	4721.16358	0.97390	0.00140
1996M02	6014	2340.33082	806.85880	1406.11282	1438.48575	17249.48461	2428.43106	13594.62200	2956.48167	1567.09729	4306.53503	0.88228	0.00842
1996M03	6505	2486.63984	784.62710	1405.03283	1454.11257	16808.15583	2737.61987	13775.12493	3269.63172	3021.35045	4861.68233	0.89763	0.01117
1996M04	5920	2106.22966	741.34996	1325.09303	1118.98200	17366.15173	2535.74517	12542.67792	2942.15687	2631.48318	4620.32356	0.84822	0.00600
1996M05	6660	2275.88045	865.61461	1550.81383	1302.36288	16030.22491	2679.66908	13466.99964	2894.98606	2176.80552	4821.02467	0.88281	0.00902
1996M06	6124	2148.84784	863.58543	1325.57287	1216.81737	16752.38268	2525.80176	12641.99533	2981.72160	3563.03935	4942.32083	0.86804	0.01207
1996M07	6518	2529.86992	993.11157	1466.82341	1532.24377	17434.22371	2551.15155	12656.15172	3015.49831	4624.92431	5034.54904	0.89241	0.01355
1996M08	6393	2346.76806	1024.96786	1605.61132	1608.68269	16817.72942	2655.32149	11801.87522	2882.55531	4299.70763	4839.89772	0.85383	0.00949
1996M09	5615	1993.19568	934.22952	1561.82059	1512.95339	16151.75255	2790.98834	11599.35662	2794.31064	3785.91976	4718.99777	0.71320	0.00929
1996M10	6843	2295.76347	1087.18997	1962.68386	1946.11988	17230.72271	2845.36565	13405.19055	2989.35362	1776.01599	5167.96783	0.75411	0.00903
1996M11	6248	2223.59288	1027.33822	1990.61433	1846.49355	17558.23810	2832.64187	13059.34394	2791.45997	1920.69072	5538.19344	0.77400	0.00793
1996M12	5735	1913.86711	971.38002	2315.18798	1641.76480	18053.43371	2961.83112	12070.56586	2956.66144	2608.35799	5011.50394	1.11667	-0.00307
1997M01	6596	2281.08460	1071.06868	1704.19812	1670.93550	19515.27208	2313.16535	13652.41363	2855.75174	1495.13222	4858.31483	0.85210	-0.00299
1997M02	6504	2259.48411	1020.12989	1626.94448	1592.98668	18396.46651	2202.27156	13278.61404	2655.15570	1971.43066	4959.98012	0.86717	0.00174
1997M03	6539	2135.25296	1153.54910	1536.98800	1349.71924	19310.68872	2558.19821	12366.78021	2952.69084	2000.02843	5414.60991	0.90403	0.00837
1997M04	6715	2199.58350	1321.46362	1727.34539	1561.56635	20241.96158	1535.60251	14192.23258	2822.91637	2775.70522	5658.15803	0.86634	0.01080
1997M05	6537	2185.61738	1257.81714	1798.44693	1534.21663	19973.53071	1537.89643	13272.47348	2996.06306	3048.25175	5544.02679	0.92567	0.00839
1997M06	6513	2024.85171	977.95626	1602.47043	1470.46144	18968.87561	1926.36470	13307.40941	2864.07454	4749.10667	5634.77132	0.83893	0.00893
1997M07	6332	2146.46170	1210.06300	1753.51581	1606.25697	20029.72688	2586.86419	13933.92703	2987.06202	5409.42126	5817.98389	0.87924	0.00816
1997M08	7019	2107.30191	930.49311	1805.50352	1648.42363	20447.90974	2464.13275	13391.31858	3023.84830	9906.80746	5864.70184	0.82070	0.00770
1997M09	6996	2184.79291	1014.14291	1985.86809	1806.53527	19459.39343	2382.84078	13361.48141	2823.43266	3610.79747	5954.97143	0.69019	0.00555
1997M10	7233	2297.96694	972.16342	2179.21479	2118.11436	19397.76280	2508.21062	15058.79891	2975.68832	1978.11401	6594.80579	0.74901	0.00708
1997M11	6714	2140.27392	903.00279	2220.94360	2006.98099	19645.26372	2625.33738	14216.50068	2830.70815	2031.21070	6342.13769	0.71638	0.00014
1997M12	5968	1855.07572	747.78642	2762.84231	1830.60288	19269.95807	2399.57514	13286.87848	2819.34305	2415.12522	6498.91642	1.09202	0.00484

Base de datos Parte 1.

BASE DE DATOS DEFINITIVA													
obs	VOLTRIP	VALTRIPDEF	VALCONSTDEF	VALMUEBDEF	VALCOLCHDEF	VALPAPDEF	VALCARTDEF	VALENCARTDEF	VALEDPERDEF	VALEDLIDDEF	VALIMPDEF	IMWDEF	CETE28DEF
1998M01	6373	1951.76767	870.01352	1704.72178	1879.19181	21108.54201	2827.01364	14978.49808	3417.77767	1946.72699	5830.06009	0.86269	-0.01059
1998M02	6139	1860.78363	825.30827	1809.95442	1785.60164	19607.79863	2557.72867	12390.89762	3240.90966	1950.26596	5262.88762	0.86668	-0.00258
1998M03	6989	2114.59043	928.22786	2023.10807	1814.35929	20662.55855	2543.90227	14391.74778	3370.94537	3571.59377	6491.99855	0.93053	0.00631
1998M04	6433	2084.50593	819.47171	1755.47872	1540.98664	19169.56511	2258.67610	14367.34464	3317.13878	2381.73432	5588.37464	0.87638	0.00944
1998M05	6467	2086.24369	937.75590	2021.82559	1706.51286	19861.37791	2514.13870	14315.18443	3329.12014	3595.87192	5169.65679	0.96969	0.01034
1998M06	7056	2172.36509	792.48307	1920.35193	1675.92966	20393.58101	2578.24405	15203.70968	3374.33834	3898.91949	5655.54865	0.94096	0.00455
1998M07	6862	2263.91282	890.47303	1932.78886	1749.36410	22209.98577	2851.40349	16047.22370	3389.80819	8296.66367	5725.06042	0.98477	0.00716
1998M08	6920	2255.30580	876.91976	2044.07116	1842.82129	20895.71736	2711.92626	14409.69271	3424.48586	6466.02529	5702.12479	0.86618	0.00620
1998M09	6813	2368.87143	791.01618	1868.02124	1850.07746	18823.55490	2633.20527	14522.55924	3313.81168	3313.33656	6510.54043	0.74370	0.00453
1998M10	6859	2340.23052	832.32126	2469.12937	2156.71363	19985.30179	2983.30802	16539.07768	3341.91999	2273.97171	6101.75061	0.80567	0.00853
1998M11	6775	2274.91335	772.92981	2693.81796	2211.78009	20113.73828	2719.85144	14234.36040	3235.18652	2582.19985	6493.62184	0.76566	0.01258
1998M12	6745	2371.72599	773.82932	2586.64955	2140.98199	19481.84327	2731.05292	14538.96059	3164.54570	2114.90904	6914.91591	1.09658	0.01180
1999M01	6544	2221.53932	776.72214	2034.11295	1920.06506	21181.34122	2269.21153	14697.60011	3391.73441	2000.11946	5083.38525	0.99671	0.00321
1999M02	6490	2130.82602	827.78496	1716.91061	1818.02067	19312.79355	2714.78451	14234.87659	3310.02080	2266.75392	4864.81243	0.88392	0.01658
1999M03	7493	2333.18915	904.01893	1971.67582	1987.05620	21691.38581	2979.38784	16481.48582	3528.04601	2834.31433	5212.55195	0.98584	0.01534
1999M04	6516	2061.01546	845.57703	2000.72863	1822.14904	20381.41173	2810.99834	14885.92976	3366.95910	2557.05677	5447.72051	0.91827	0.01548
1999M05	6419	2027.07401	927.93118	1930.16798	1735.85498	20738.16371	2813.53112	16129.92342	3480.68163	3788.05136	4831.92294	0.99555	0.01357
1999M06	6918	2222.71871	878.07037	2057.15437	1705.53091	21320.69982	2703.62330	16654.02479	3485.56088	6346.77489	6464.53237	0.93075	0.01049
1999M07	6935	2193.52728	990.26470	2042.48122	1849.66921	21387.17079	2857.60235	16122.21980	3559.92607	5668.55180	5951.81510	0.99965	0.01205
1999M08	6883	2256.63636	994.56413	2035.73084	1921.74601	21532.96217	2639.43311	16110.77222	3657.89332	4750.94650	5925.68575	0.90002	0.01301
1999M09	6979	2288.18766	880.66015	2088.61785	1965.37534	19820.87609	2576.63068	15354.35105	3627.54158	3111.67804	5211.19049	0.79058	0.00998
1999M10	7143	2330.82690	757.05195	2406.59882	2010.74554	21350.64142	2779.00548	16652.33468	3762.04614	1913.24133	6325.36982	0.86510	0.00683
1999M11	7535	2474.61632	790.91136	2457.48045	2271.71912	20986.36857	2732.25673	17098.86519	3687.31041	2112.83973	6311.16495	0.83096	0.00731
1999M12	7007	2280.90662	624.23259	3082.74139	2198.44124	20971.55678	2667.91774	15569.38462	3697.41222	1550.58329	6869.06225	1.17525	0.00591
2000M01	6990	2315.80531	680.76011	2350.25274	1772.03897	20793.98715	2812.63751	16544.01503	3700.31118	2351.38301	5528.22267	1.06603	0.00207
2000M02	6653	2207.38937	728.41981	2047.11963	1841.81146	20797.19494	2583.71268	15521.42557	3573.15991	2380.17202	5366.16896	0.95668	0.00585
2000M03	7083	2347.59907	733.91181	1873.97445	1921.81810	23532.25247	2761.21601	17057.17742	3688.32156	3443.31124	5337.28883	1.00802	0.00589
2000M04	6183	2012.50971	737.04868	1751.50068	1473.23641	21682.81458	2607.66385	15260.67987	3530.04292	2586.26202	5148.72878	1.04153	0.00255
2000M05	6961	2171.40836	818.93607	2132.94760	2002.66477	22751.57685	2647.45126	17354.32551	3501.43922	5586.53939	5776.37365	1.06314	0.00788
2000M06	6733	2127.42231	763.84038	2233.59958	1921.73017	22315.75511	2725.66519	16328.06295	3608.05705	5132.56298	5791.45608	1.00885	0.00388
2000M07	6850	2177.05631	572.53441	2170.66723	2080.58155	22476.65302	2812.90587	16126.09995	3824.07987	5152.53869	5799.89030	1.02313	0.01108
2000M08	7207	2311.30323	679.87591	2340.95972	2286.67815	22033.36446	2895.25028	16430.36958	3852.16539	4080.93605	5155.98405	0.94772	0.00909
2000M09	6841	2120.99767	684.16735	2245.12810	2087.18695	21210.32981	2344.53764	15707.91635	3779.56829	3186.13536	5131.94381	0.82404	0.01013
2000M10	7006	2209.64255	756.74176	2569.83759	2368.24531	21787.91929	2824.67362	16979.18535	3894.05061	2040.71605	5419.42547	0.86355	0.00534
2000M11	7147	2250.37493	723.65996	2876.52113	2561.21529	19537.33699	2737.38919	16781.98509	3879.54922	2585.75807	5853.28162	0.88052	0.00840
2000M12	6064	2084.91097	729.51210	2728.62614	2055.04386	20501.17812	2090.20760	14701.20667	3763.89326	2155.86321	5623.63148	1.26403	0.00846
2001M01	5664	1894.45886	624.12593	2168.39026	2015.11824	21053.70449	2051.42998	16436.74624	3532.97763	1736.98039	5161.77373	1.15341	0.00297
2001M02	5400	1734.97831	581.78664	1922.48226	1946.49966	19658.93946	2094.80560	14586.88610	3572.85290	2169.81353	4073.35771	0.94005	0.01441
2001M03	5525	1781.74407	676.21314	2285.96967	2046.26328	21511.98676	2411.86236	16818.75231	3786.23126	2953.42342	5446.14094	0.97975	0.00501
2001M04	4849	1592.12568	562.63042	1971.70494	1566.80675	19040.80167	2125.03552	15404.67076	3621.73876	2994.91730	4604.46768	0.94330	0.01162
2001M05	4915	1605.20214	551.99198	2165.87875	2020.99183	20629.93892	2606.69358	16759.90116	3598.92944	4613.52258	4724.79477	0.96701	0.01226
2001M06	5406	1647.48189	611.66674	1856.68953	1994.70087	20392.01212	2495.13144	16307.37867	3597.81412	8797.49125	5153.74951	0.97626	0.00870
2001M07	5031	1559.09422	695.53520	1941.28476	1900.50221	19852.08842	2499.82530	15716.27233	3643.54946	5744.50036	5435.17286	1.00484	0.01005
2001M08	5049	1559.31010	631.50956	1962.46984	2100.30412	19584.33079	2301.53144	16039.16721	3636.71561	3907.95530	4925.28668	0.91075	0.00234
2001M09	5322	1646.80892	648.24710	1828.58894	2010.63304	18887.82950	2147.69342	14985.13453	3450.47213	2427.73584	4821.32476	0.74492	0.00115
2001M10	5549	1750.07963	669.78399	2122.38838	2338.55654	19423.92495	2383.67636	16994.57243	3711.18677	2219.92277	5512.37769	0.77834	0.00413
2001M11	5260	1739.52971	602.36065	2042.53724	2375.93472	19244.45654	2318.26616	15810.91162	3616.08582	1117.69454	5534.70544	0.86680	0.00604
2001M12	4310	1432.07622	526.42992	2038.75347	2252.09248	18525.99706	1777.49038	14330.72160	3479.73565	1383.07099	5563.25131	1.20275	0.00846

Base de datos Parte 2.

BASE DE DATOS DEFINITIVA													
obs	VOLTRIP	VALTRIPDEF	VALCONSTDEF	VALMUEBDEF	VALCOLCHDEF	VALPAPEDEF	VALCARTDEF	VALENCARTDEF	VALEDPERDEF	VALEDLIBDEF	VALIMPDEF	IMWDEF	CETE28DEF
2002M01	4591	1395.44473	592.77601	1848.78910	2411.26723	18292.15553	1905.91019	16805.23581	3373.88600	1756.90734	4665.21453	0.97671	0.00075
2002M02	4900	1479.29099	539.65925	1748.36102	2265.39094	17412.17222	1909.72316	15002.51987	3201.10023	2114.35980	4089.83459	0.85783	0.00756
2002M03	4854	1455.50214	515.85010	1688.84698	1866.52366	20097.27436	1747.18268	15812.11131	3476.69177	1797.43713	3985.55578	0.92043	-0.00059
2002M04	5658	1620.56731	640.94843	1943.94849	2304.20787	19661.56652	2098.71867	17480.06947	3437.26142	2894.03140	4632.85429	0.90259	0.00067
2002M05	5293	1498.54483	609.97854	2313.66122	2188.49835	20157.61515	2150.50145	17472.69156	3498.79386	7053.95082	5228.69149	0.94048	-0.00181
2002M06	5014	1418.97920	565.78371	2018.65075	1773.68322	20105.02662	2082.64071	16150.17244	3486.39216	6821.80913	4399.72918	0.92452	-0.00159
2002M07	4822	1396.23739	600.19333	1894.00715	2087.47712	20501.15976	2198.65470	16836.52318	3620.76977	3998.49047	4486.34653	0.95222	0.00094
2002M08	5186	1506.59575	664.18458	2310.40358	2279.53081	20743.74447	1993.02626	16377.19725	3793.48053	2364.07849	4426.85252	0.89167	0.00391
2002M09	5023	1396.54420	602.55372	2070.98849	2055.10092	18580.32233	1740.62075	15082.25548	3480.61432	2242.96852	4421.54261	0.76754	0.00145
2002M10	5870	1688.41171	728.77845	2318.21817	2541.45537	19966.57074	1834.80540	16976.90968	3625.93764	2327.68705	5223.17293	0.75221	0.00013
2002M11	5499	1610.66380	655.30897	2393.01616	2692.52215	19687.62881	1591.95909	15601.70624	3683.63079	2360.71921	4913.91426	0.80778	-0.00432
2002M12	4889	1407.32967	540.71876	2755.49394	2510.55516	20306.44596	1525.27743	14383.08935	3516.63214	2719.04828	5587.72466	1.08652	0.00258
2003M01	4910	1481.86732	511.29657	1818.75730	2435.71263	20531.08190	1525.25259	16241.25273	3558.39336	1085.84599	3963.70990	0.89398	0.00230
2003M02	4785	1447.32041	468.71928	1846.55068	2236.21282	19613.18357	1666.10952	15214.92209	3383.22685	1956.78426	3760.82332	0.97609	-0.00457
2003M03	5181	1567.87527	571.68993	2060.41310	2226.42737	21987.95452	1826.39985	16058.61793	3603.08903	2871.53829	3590.43000	0.98835	-0.00548
2003M04	4624	1394.93924	532.48406	1909.30116	1951.64364	20892.95232	1480.67001	15756.54315	3636.37677	4326.51415	3750.29000	1.01129	0.00931
2003M05	4832	1397.18772	512.06245	1919.48401	2046.52732	21283.61521	1873.97713	16259.36753	3567.30678	8716.74375	3907.28000	0.92440	0.01058
2003M06	4395	1255.43625	555.74376	1843.74372	2048.79082	20037.92849	1857.89823	14905.08778	3603.25420	4106.19943	4018.43000	0.94938	0.00089
2003M07	5028	1509.67784	532.36199	2018.60031	2046.42166	20250.17776	1907.67159	16082.49820	3746.34742	3051.82741	4084.25000	0.93753	0.00251
2003M08	5025	1490.30210	588.07513	1780.98098	2170.08463	20609.95452	1920.32423	15093.74674	3604.46988	2057.88203	3920.34000	0.92012	-0.00075
2003M09	4485	1385.74311	570.28151	1604.29070	2289.37952	19649.26349	1859.25994	14984.98141	3430.49785	1919.96011	3750.37000	0.75104	-0.00092
2003M10	5526	1612.02254	564.90212	2184.19623	2634.74317	20989.04095	1801.68570	17368.61425	3547.23677	1578.93047	4142.27000	0.95850	-0.00323
2003M11	5128	1607.91350	598.63618	2102.36647	2603.72864	20512.00916	1723.83120	15742.61655	3546.63862	1756.27280	4507.80000	0.79872	-0.00554
2003M12	5344	1606.82000	504.60000	2331.05000	2551.85000	20867.31000	1476.47000	15537.48000	3445.66000	2017.91000	5141.01000	1.00000	-0.00338
2004M01	5168	1607.24630	578.74016	1699.35613	2362.15811	20382.54083	1873.95177	16825.21180	3477.82795	1383.30869	3642.49646	0.82762	0.00009
2004M02	4649	1421.83585	581.88521	1564.11946	2180.44722	20290.72142	1800.49626	15393.21655	3419.88517	1592.79910	3293.93376	0.74689	-0.01100
2004M03	5268	1657.73686	607.38984	1710.09540	2400.12601	22332.11211	2076.00256	17797.88553	3710.54712	2671.62181	3646.31134	0.87429	-0.00909
2004M04	5472	1742.16943	668.07041	1427.62927	2034.09491	21142.65277	1958.45875	16300.76300	3448.87702	3336.35800	3591.88154	0.79110	-0.00893
2004M05	5686	1795.42681	739.42640	1794.13329	2172.99864	22333.93124	2140.95674	16208.43181	3601.45977	6130.22134	3440.35249	0.79862	0.00161
2004M06	5954	1897.12548	837.93995	1671.21579	2055.20200	22250.95413	2209.14504	16905.70402	3617.22670	6914.05427	3787.84790	0.79077	0.00329
2004M07	6085	1981.39549	972.88648	1658.32892	2065.88377	22669.95843	2314.77599	17259.19656	3603.01890	3962.55840	3777.39083	0.78109	0.00294
2004M08	5788	1935.20706	882.51844	1811.22946	2374.32276	22764.51059	2090.57825	16702.37615	3628.58391	3284.39708	3827.17518	0.78073	-0.00066
2004M09	5631	1840.51737	889.00594	1868.08662	2334.76026	21274.56790	2107.79732	16976.30132	3651.29561	1692.02933	3981.26599	0.79400	0.00113
2004M10	5512	1869.68292	805.58345	1784.67939	2653.30961	22805.15989	2358.68645	18998.70157	3679.73358	1075.55534	4391.51731	0.78961	0.00255
2004M11	5240	1692.36200	689.03912	1643.81292	2736.90651	22066.39687	2234.98655	18401.74672	3674.04081	1442.43161	4432.73215	0.77916	-0.00047
2004M12	5625	1893.81370	563.30926	2423.92211	2730.57336	22579.31908	1977.58231	17244.70530	3487.39583	1398.19551	5042.54556	0.92483	0.01000
2005M01	5175	1590.64057	593.90113	1489.34834	2413.10692	22416.13731	2393.40407	17740.26824	3544.81326	1454.87406	3721.76449	0.97324	0.00635
2005M02	4898	1383.83280	560.43606	1378.84921	2088.29731	20931.27785	2154.35584	16779.06378	3407.70553	1973.13189	3037.89654	0.76615	0.00152
2005M03	5274	1488.01568	629.24844	1655.44712	2228.13277	23126.45385	1904.64717	17849.32103	3489.24075	1814.88897	3523.32973	0.73095	0.00452
2005M04	5319	1625.47784	659.84072	1483.37738	2311.37510	22728.05888	2197.33804	19625.47861	3626.75101	4188.94669	3575.79092	0.73874	0.00389
2005M05	5482	1582.80843	590.69720	1623.50124	2256.93553	23636.98121	2095.92700	18664.13089	3571.00573	3906.56827	3746.98773	0.77595	0.01384
2005M06	5268	1535.59091	581.20422	1635.09563	2218.96601	23827.67032	2108.22388	18085.42749	3500.85573	6621.13168	3908.44526	0.75736	0.01418
2005M07	5614	1661.94320	561.36128	1571.56949	2271.26576	23080.95276	2356.45630	18777.63787	3560.07580	4124.05213	3599.37525	0.75228	0.00257
2005M08	5601	1574.34678	575.32791	1598.03585	2457.77550	22947.64028	2040.67312	19228.64467	3466.13308	3050.25057	4017.14953	0.71711	0.00843
2005M09	5685	1621.10282	575.32077	1808.27793	2369.99761	20305.99596	1970.87770	18057.95786	3583.05100	1527.14550	3958.11727	0.68706	0.00325
2005M10	5604	1670.03912	539.65819	1767.51231	2720.31021	21629.33755	1961.83597	19503.01232	3528.73429	1616.37915	3789.31207	0.70064	0.00296
2005M11	5618	1579.35407	526.17562	1888.84877	2924.10092	21046.81775	2063.12708	18930.47835	3517.52925	1433.84292	3704.05111	0.70703	0.00384
2005M12	5048	1529.62709	517.45934	1911.31469	2775.27300	21429.28156	1835.99544	18105.43968	3413.10836	1332.33943	4644.60294	0.88482	0.00232

Base de datos Parte 3.

BASE DE DATOS DEFINITIVA													
obs	VOLTRIP	VALTRIPDEF	VALCONSTDEF	VALMUEBDEF	VALCOLCHDEF	VALPAPDEF	VALCARTDEF	VALENCARTDEF	VALEDPERDEF	VALELIBDEF	VALIMPDEF	IMWDEF	CETE28DEF
2006M01	5209	1547.15188	615.87640	1624.19662	2506.84474	21910.05617	1767.64203	19690.31372	3446.00964	2102.05931	3433.75076	0.73047	-0.00102
2006M02	4859	1484.59945	613.11911	1419.31722	2192.56876	20668.67308	1842.84923	18168.20286	3259.44611	1688.90745	3639.81228	0.73298	0.00230
2006M03	5047	1479.25645	677.25866	1522.21823	2551.28848	23317.98720	1746.15785	19918.87462	3347.63179	2234.71413	3576.50121	0.68603	-0.00056
2006M04	4770	1319.38312	646.19973	1401.52285	2087.71452	21954.60770	1600.74198	18355.92167	3241.09288	2938.96309	3494.27065	0.72395	-0.00622
2006M05	5672	1584.51270	710.74335	1474.51476	2617.96123	22473.86182	1764.87891	20042.91711	3262.14041	3271.22919	4002.89848	0.76909	-0.00696
2006M06	5001	1532.20440	651.80902	1480.10323	2519.91209	22541.29960	1823.98143	19291.38226	3274.21853	5073.81767	4406.68255	0.76601	-0.00501
2006M07	5585	1551.33979	692.91840	1544.17565	2668.12315	23099.12898	1867.69410	19588.68397	3324.03342	4704.18283	3714.79153	0.76355	0.00439
2006M08	5443	1568.04963	651.41689	1676.57890	2682.76060	22453.05304	1795.57490	19884.83799	3479.11998	2812.00570	3731.04159	0.77157	-0.00047
2006M09	4809	1406.11921	581.92090	1442.17329	2530.08921	21662.14581	1726.61617	19453.06229	3437.26048	3011.41435	3923.16923	0.73374	0.00126
2006M10	5401	1522.91992	510.82103	1406.28553	2859.45123	22382.55324	1862.66457	21169.53179	3488.33118	1667.19506	4094.77149	0.75347	0.00601
2006M11	5308	1633.27857	435.20033	1535.32154	2933.44551	21171.21570	1734.46953	20455.51971	3201.15944	1968.11818	4181.96665	0.79067	0.00649
2006M12	4993	1446.03932	376.25325	1508.46884	2764.63675	21378.80329	1493.85995	18775.86895	3266.27091	1579.64610	3678.11535	1.00541	0.00283
2007M01	5329	1588.61552	502.85442	1377.41810	2931.63658	23419.25312	1856.29499	21318.27083	3359.54277	2259.72489	3475.49057	0.77493	0.00355
2007M02	4578	1266.17289	458.79842	1305.04574	2315.79659	21736.56027	1647.41712	19207.11877	3111.94617	2022.00721	3351.09410	0.71387	0.00179
2007M03	5116	1503.59770	516.12323	1377.77541	2505.36071	23576.93675	1822.37563	22468.87521	3298.88990	2871.69775	3741.74704	0.78158	-0.00082
2007M04	5143	1501.47464	424.29124	1303.02011	2335.55713	22672.28439	1516.11149	19676.33470	3220.12936	2292.18738	3563.11535	0.73670	0.00465
2007M05	5859	1645.37860	527.84050	1458.10281	2486.94013	23408.09448	1941.12548	21733.91197	3249.58579	4160.68955	3771.40147	0.82190	0.01303
2007M06	5484	1626.01019	530.92530	1410.26588	2455.63701	23186.52466	1830.64990	20500.78711	3205.35114	3887.08391	3789.46441	0.78877	0.00408
2007M07	5220	1556.13118	584.21245	1276.75247	2581.73731	24725.51634	1844.32053	21151.93749	3468.44183	4566.61119	4028.95777	0.84816	0.00139
2007M08	5359	1434.43391	612.35321	1379.91937	2782.43959	24134.24542	1891.01311	21094.38111	3460.68181	3042.95792	4083.72573	0.81074	-0.00395
2007M09	5659	1510.42103	614.32467	1356.50541	2651.20448	21947.38799	1680.18442	20146.63431	3390.27270	2010.30383	3745.10446	0.82121	-0.00184
2007M10	5603	1528.06626	626.86555	1636.14930	3064.16154	24656.17745	1822.41051	23710.76310	3567.22190	1694.78653	4251.33667	0.79931	0.00482
2007M11	6117	1705.59868	557.42936	1744.88427	3458.06172	23557.93112	1799.71999	21009.75159	3250.79955	1552.66212	4216.46144	0.81522	0.00272
2007M12	5119	1422.27802	400.77458	1631.47754	3344.36192	23055.35091	1756.61606	19130.93621	3503.01389	1046.97283	4489.52904	1.08743	0.00618
2008M1	5497	1632.81382	552.8505443	1460.634242	3191.493973	23742.07397	1784.55138	22014.72213	3258.145643	2333.851772	3516.7997	0.82434	-0.00224
2008M2	4871	1480.71083	484.6604367	1407.556773	3080.735245	23424.70929	1669.556329	20849.11002	3130.259455	2534.845946	3173.06767	0.758336	-0.00487
2008M3	5380	1648.97107	466.8424683	1347.316637	2411.87078	24594.98875	1818.521835	19596.10834	3484.710319	3071.512579	3344.61995	0.83592	-0.00407
2008M4	5355	1657.75968	512.0551277	1490.703084	2896.120069	24284.29591	1919.244982	22076.87827	3350.476664	3401.04513	3309.41309	0.810069	-0.00119
2008M5	5708	1755.44835	510.8066478	1332.891268	2713.822466	24031.70135	1839.131932	22587.28256	3287.185763	4684.445202	4039.94548	0.838569	0.00186
2008M6	5278	1687.26324	502.5131739	1427.601141	2609.191405	23381.18597	1871.112955	21760.48633	3392.461833	6757.047838	3294.78833	0.827643	0.00032
2008M7	5435	1703.63916	477.1110495	1247.322827	2754.079156	24071.43236	1907.681709	22403.07578	3896.417564	10840.25788	3850.74639	0.863767	-0.00247
2008M8	5253	1582.14755	457.7636117	1167.876276	2610.278635	23823.83188	1894.897128	22935.77507	3457.833352	3339.87502	3676.56254	0.814181	0.00617
2008M9	4971	1528.90215	373.4140182	1195.675396	2567.005337	22862.09395	1819.403125	20938.66986	3295.630358	1819.294045	3703.90744	0.801163	0.00140
2008M10	5820	1827.75498	353.431687	1241.312399	2935.677215	23336.44531	2044.744127	21695.23007	3475.064983	1751.116594	4315.14749	0.746622	-0.01874
2008M11	5417	1687.96511	306.0962851	1372.124111	2916.516135	22801.42598	1706.96888	19155.81256	3507.027606	1200.589446	4432.76885	0.709434	-0.00238
2008M12	4514	1322.00480	261.1337976	1307.07659	2988.825362	22661.73919	1392.771493	17760.20135	3150.39027	1123.6972	3932.94834	0.880478	0.00333

Base de datos definitiva Parte 4.

# Anexo 2: Scripts en R de automatización

## Script general: generalPAR.R

```
#Script General, eventualmente llamara al segundo script klzPAR.R
# Nombre de archivo para este caso: generalPAR.R
# INSTRUCCIONES
# 1 instalar R
# 2 instalar los paquetes parallel,doParallel,foreach,tools,utils,forecast,gregmisc
# 3 cambiar define el paralelo, a usar los cores adecuados
# 4 correr este archivo (generalPAR.R) usando R
# 5 esperar a que termine
# define (captura los datos de la serie a pronosticar) como y1 Para este caso especifico
fue la serie DVOLTRIP cuya primera observacion era nula por lo que inserte como primera
observacion el valor de la segunda
y1 <-c(-99,-99,499,-174,398,34,21,115,-46,1036,48,-538,-59,-284,601,-1053,433,-161,-496,357
,-76,874,-209,-395,460,199,491,-585,740,-536,394,-125,-778,1228,-595,-513,861,-92,35,176,-178,-24,-
181,687,-23,237,-519,-746,405,-234,850,-556,34,589,-194,58,-107,46,-84,-30,-201,-54,1003,-977,-97,499,17,-
52,96,164,392,-528,-17,-337,430,-900,778,-228,117,357,-366,165,141,-1083,-400,-264,125,-676,66,491,-
375,18,273,227,-289,-950,281,309,-46,804,-365,-279,-192,364,-163,847,-371,-610,21,-125,396,-557,208,-
437,633,-3,-540,1041,-398,216,-176,-519,619,204,214,268,131,-297,-157,-119,-272,385,-450,-277,376,45,163,-
214,346,-13,84,-81,14,-570,161,-350,188,-277,902,-671,584,-142,-634,592,-93,-315,336,-751,538,27,716,-
375,-264,139,300,-56,514,-998);
# convierte la serie y1 en una serie de tiempo con periodo cada 12 meses
yt=ts(y1,freq=12);
# AR a considerar erase c(1,4)
lasAR <- c(1,2,4,11,12)
# MA a considerar
lasMA <- c(1,5,11,12,13,17,18,24,29,30,36,41,42,43,47,48,60,72,84,98)
# numero de modelos con minimo AIC a registrar
numModelos<-10
```

```

# archivo en que se registra cada modelo que logra entrar en la lista de minimos
archivoReporte <- "/home/claudia/tesisR/paralel_v_batch_01.csv"
# incluye el intercepto ? poner TRUE si lo incluye, FALSE para no incluirlo
conC<-FALSE
# probabilidad del criterio descarte a utilizar
probCoef<-0.05
# indica si ir reportando el avance y cuales han gustado. TRUE si reporta, FALSE no
reporta
  reportaAvance<-TRUE
  reportaGustaron<-TRUE
# carga el programa
source("klzPAR.R")
# define el paralelo, usa 4 cores
cl <- makeCluster(4)
registerDoParallel(cl,cores=4)
# ejecuta todas las combinaciones en paralelo (definida en klzPAR.R)
timea();
# termina el paralelo
stopCluster(cl)

```

## Script complementario: klzPAR.R

```
# Nombre del erchivo: klzPAR.R
# carga los paquetes de R requeridos
library(parallel)
library(doParallel)
library(foreach)
library(tools)
library(utis)
library(forecast)
library(gregmisc)

# funcion que concatena los terminos a usar en el modelo a calcular
klz_juntaEQ <- function(letra,term,fin=""),coef) {
  klzJuntaEQ_s<-" "
  klzJuntaEQ_Build<-0
  for(klzJuntaEQ_x in term) {
    klzJuntaEQ_Build<- klzJuntaEQ_Build + 1
    klzJuntaEQ_s<-paste(" ",klzJuntaEQ_s,"+ ",coef[klzJuntaEQ_Build],letra, klzJuntaEQ_x,fin,sep=" ",
collapse=" ")
  }
  klzJuntaEQ_s
}

# funcion que ordena por minimo AIC. usa el algoritmo bubblesort
klz_bubblesortAIC <- function(v) {
  itemCount <- length(v)
  repeat {
    hasChanged <- FALSE
    itemCount <- itemCount - 1
    for(i in 1:itemCount) {
      if ( v[[i]]$aic > v[[i+1]]$aic ) {
        t <- v[i]
```

```

v[i] <- v[i+1]
v[i+1] <- t
hasChanged <- TRUE
}
}
if ( !hasChanged ) break;
}
v
}

# calcula cada modelo combinando las ar recibida y las ma regresando los 10 mejores
modelos

klzPAR_arma <- function( klz_yt,klz_archivoReporte,klz_conC,klz_probCoef,klz_lasAR,
klzNumAR,klz_lasMA,klzNumMA,klzPorMA,klziPorAR,klz_ReportaAvance=FALSE,
klz_saveFiles=FALSE) {
  # inicializa las variables
  klzCombs<-array(list(NULL)); # array para guardar los modelos seleccionados
  klzVan <- 0 # van calculadas
  klzGustaron <- 0 # han entrado a la lista de los 10 minimos AIC
  klzMinAIC <- c(9e12,9e12,9e12,9e12,9e12,9e12,9e12,9e12,9e12,9e12) # los 10 minimos AIC
  klzMaxAIC <- 9e12 # el maximo AIC
  klzFileOut <- paste( klz_archivoReporte,klziPorAR,".csv",sep="_ ") # archivo en que
guardar los resultados
  if(klziPorAR==0) { # si no hay AR ponlo en vacio
    klzCombAR <- NULL # combinaciones no hay
    klzNar <- c(0) # hay 0
    klziNar <-0 # numero de AR 0
  } else {
    klzCombAR <- combinations(klzNumAR,klziPorAR,klz_lasAR) # toma todas la combi-
naciones de AR con klziPorAR elementos
    klzNar <- 1:dim(klzCombAR)[1] # cuantas AR vienen en esta combinacion

```

```

klziNar <- length(klzNar) # numero de AR
}
# calcula
for(klziAR in klzNar) { # para cada combinacion de ARs haz:
for(klziPorMA in klzPorMA) { # para cada combinacion de MA
# prepara las MA
if(klziPorMA==0) { # no hay MA en esta iteracion
klzCombMA <- NULL # combinaciones no hay
klzNma <- c(0) # hay 0
klziNma <-0 # numero de AR 0
} else {
klzCombMA <- combinations(klzNumMA,klziPorMA,klz_ lasMA) # toma todas la combi-
naciones de MA con iPorMA elementos
klzNma <- 1:dim(klzCombMA)[1] # cuantas MA vienen en esta combinacion
klziNma <- length(klzNma)
}
if( klziNar > 0 || klziNma > 0) { # si hay AR o MA que calcular
for(klziMA in klzNma) {
klzVan <- klzVan + 1 # voy a probar 1 combinacion mas
if( klziNar > 0 && klziNma > 0) {
# calcula el modelo con ar y ma
klzARMA <- arma(klz_ yt,include.intercept=klz_ conC, lag=list(ar=klzCombAR[klziAR,],
ma=klzCombMA[klziMA,]))
} else if(klziNar>0) {
# calcula el modelo con ar
klzARMA <- arma(klz_ yt,include.intercept=klz_ conC, lag=list(ar=klzCombAR[klziAR,],ma=NULL))
# calcula el arma
} else {
# calcula el modelo con ma

```

```

klzARMA <- arma(klz_ym,include.intercept=klz_conC, lag=list(ar=NULL,ma=klzCombMA[klziMA,]))
# calcula el arma
}
# obten los datos basicos del modelo
klzARMASummary <- summary( klzARMA )
# checar modelo
# verificar la significancia de los coeficientes
# obtener la maxima probabilidad de significancia de todos los coeficientes
klzCheckCoefsTtest <- max( klzARMASummary$coef[,4] )
if(is.na(klzCheckCoefsTtest)) {
klzCheckCoefsTtest <- 99
}
# si se pudieron estimar los coeficientes y la maxima probabilidad de significancia es menor
a la deseada
# y se pudo calcular el AIC y el AIC es menor al mayor AIC seleccionado hasta este
momento entonces entra a los mejores
if( !is.na(klzARMASummary$coef[1]) && klzCheckCoefsTtest<klz_probCoef &&
!is.na(klzARMASummary$aic) && klzARMASummary$aic < klzMaxAIC ) { # modelo valido
# pon el modelo en los mejores evacuando el peor de los AICs
klzAIC <- klzARMASummary$aic
klzK <- 0 # inicializa el indice
for(klzTmpAIC in klzMinAIC) { # toma cada AIC minimo
klzK <- klzK + 1 # incrementa el indice
if(klzTmpAIC == klzMaxAIC) { # si el maximo es este elemento sustituelo, reporta y
continua
klzGustaron <- klzGustaron + 1 # indica uno mas gusto o entro a los minimos
# registra en mejores modelos
klzMinAIC[klzK] <- klzARMASummary$aic
klzMaxAIC <- max(klzMinAIC)
# reporta en la bitacora o log

```

```

# redacta las lineas a reportar en la bitacora
if( klziNar > 0 && klziNma > 0) {
  tmpARMA <- paste( paste("AR", klzCombAR[klziAR,],sep=" ", collapse=" "),
paste("MA",klzCombMA[klziMA,],sep=" ", collapse=" ") )
  tmpAR <- paste( klzCombAR[klziAR,],sep=" ", collapse=", ")
  tmpMA <- paste( klzCombMA[klziMA,],sep=" ", collapse=", ")
  tmpCoefAR <- klz_juntaEQ("AR(",klzCombAR[klziAR,],"",klzARMASummary$coef[0:length(
klzCombAR[klziAR,]),1])
  tmpCoefMA <- klz_juntaEQ("MA(", klzCombMA[klziMA,],"",klzARMASummary$coef[
(length(klzCombAR[klziAR,])+1) : length(klzARMASummary$coef[,1]) ,1])
  } else if(klziNar>0) {
  tmpARMA <- c( paste0("AR",klzCombAR[klziAR,],sep=" ", collapse=" ") )
  tmpAR <- paste( klzCombAR[klziAR,],sep=" ", collapse=", ")
  tmpMA <- "NULL"
  tmpCoefAR <- klz_juntaEQ("AR(",klzCombAR[klziAR,],"",klzARMASummary$coef[1,])
  tmpCoefMA <-
  } else {
  tmpARMA <- c( paste("MA",klzCombMA[klziMA,],sep=" ", collapse=") )
  tmpAR <- "NULL"
  tmpMA <- paste( klzCombMA[klziMA,],sep=" ", collapse=", ")
  tmpCoefAR <- ""
  tmpCoefMA <- klz_juntaEQ("MA(",klzCombMA[klziMA,],"",klzARMASummary$coef[,1])
  }
  tmpCall <- paste( " arma(yt, lag=list(ar=c(", tmpAR, "), ma=c(", tmpMA, ")),
include.intercept="klz_conC,")", sep=" ", collapse=" ")
  tmpCoef <- paste(tmpCoefAR,tmpCoefMA,sep=" ",collapse=" ")
  klzCombs[[klzK]] <- list( arma=klzARMA, summary=klzARMASummary,aic=klzAIC,call=tmpCall,
eq=tmpCoef )
  if(klz_saveFiles) { # si nos toca reportar escribe modelo en la bitacora de gustaron

```

```

write.table( data.frame( tmpARMA,klzARMASummary$aic,tmpCoef,tmpCall),klzFileOut,
append=TRUE,sep=" ",row.names = FALSE,col.names = FALSE)
}
break
} # end if(klzTmpAIC == klzAIC)
} # end for(klzTmpAIC in klzMinAIC)
} # end if modelo valido
} # for(klziMA in n)
} # end if( klzNar > 0 || klzNma > 0)
} # end for(klziPorMA in klzPorMA)
} # end for(i in klzNar)

if(klz_ReportaAvance) # si nos toca reportar el avance indica que ya se terminaron estas
combinaciones

write.table( data.frame(klzVan,klzGustaron),paste(klzFileOut,".van",sep="_"),
append=TRUE,sep=" ",row.names = FALSE,col.names = FALSE)

# regresa los modelos seleccionados
klzCombs
}

# funcion que manda ejecutar en paralelo todas las combinaciones de AR y MA para obtener
los numModelos (10) mejores modelos con minimo AIC que cumplan las condiciones
klzPAR <- function(klz_yl,klz_lasAR,klz_lasMA,klz_archivoReporte="klzPAR.csv",
klz_conC=FALSE,klz_probCoef=0.05,klz_ReportaAvance=FALSE,klz_saveFiles=FALSE) {
# inicializa variables
klzNumAR <- length(klz_lasAR) # numero de AR
klzPorAR <- 0:klzNumAR # r de combinaciones AR n,r de mayor a menor para que en
igual AIC prefiramos menos elemntos
klzNumMA <- length(klz_lasMA) # numero de MA
klzPorMA <- 0:klzNumMA # r de combinaciones MA n,r de mayor a menor para que en
igual AIC prefiramos menos elemntos
klziPorAR <- 0

```

```

klzModelos<-array(list(NULL)); # inicializa los modelos

# corre en paralelo la funcion klzPAR_arma usando un core por cada AR y sus combina-
ciones con las restantes AR y MA.

# Al terminar todas las combinaciones con esa AR el core queda libre para pasar al siguiente
AR aun no calculado ni en proceso de calculo

foreach(klziPorAR = klzPorAR, .inorder=FALSE, .export=c("klzPAR_arma","klz_juntaEQ"),
.packages=c("forecast","utils","gregmisc"), .verbose=FALSE )%dopar %

  klzPAR_arma( klz_yt,klz_archivoReporte,klz_conC,klz_probCoef,klz_lasAR,
klzNumAR,klz_lasMA,klzNumMA,klzPorMA,klziPorAR,
klz_ReportaAvance,klz_saveFiles)
}

# ejecuta en paralelo el calculo de las ecuaciones a evaluar

timea <- function () { klzPAR(yt,lasAR,lasMA,archivoReporte,conC,
probCoef,reportaAvance,reportaGustaron) }

```

# Anexo 3: Script en PHP de generación de combinaciones

Este script únicamente genera una parte del código de *E – Views* el cual posteriormente deberá ser insertado en el código final del paquete.

```
<?php
/**
 * Nombre del archivo: comb.php
 *
 * Pide los terminos ar y ma deseados
 * Escriba la "table" de entrada a copiar al programa de eviews que este corre todas las
combinaciones
 *
 * Nota: Correr en un webserver con php habilitado
 * 1.- pegar en un directorio del webserver
 * 2.- apuntar el browser al webserver/directorio/comb.php (ie http://localhost/comb.php)
 * Se despliega una pantalla de internet donde se solicita capturar los números de los terminos
ar y ma
 * 3.- Se llenan los campos y se hace click en "Generar programita"
 * 4.- Aparece el codigo a copiar en el programa de eviews
 *
 * @author Claudia Estela Ortiz de Dios
 */
set_time_limit(0); // tiempo ilimitado
?>
<!DOCTYPE html>
<html>
<head>
<title>KLZ ARIMAs</title>
</head>
```

```

<body>
<?php
if(param('ar')!=" " || param('ma')!=" ") // si llegaron terminos ar o ma genera y reporta la
tabla de terminos
combinaciones_ARMA();
?>
<!-- presenta la forma de terminos ar y ma a registrar -->
<div align=center">
<form method="get">
<table border="1">
<tr>
<th>AR</th>
<td><input size="60" name="ar" value="<?php echo param('ar'); ?>/>
<br /><i># separados por comas. ie 1,4</i>
<tr>
<th>MA</th>
<td><input size="60" name="ma" value="<?php echo param('ma'); ?>/>
<br /><i># separados por comas. ie 2,3</i>
<tr>
<td align=center colspan="2"><input type="submit" name="Calcular" value="Generar Programita/>
</table>
<i><ul><li>no poner "ni ' ni &gt; & &lt; en las cajitas!</ul></i>
</form>
</div>
</html>
</body>
<?php
/**
* combinaciones_ARMA()
* Despliega en pantalla en codigo de eviews de la tabla de terminos a combinar

```

```

* a sustiuir en la seccion pertinente del programa de eviews
*
* @author Claudia Estela Ortiz de Dios
* @return void
*/
function combinaciones_ARMA() {
$tmp=param('ar'); // obten las ar deseadas
if($tmp!="")
$in_ar=explode(',',$tmp); // si desea terminos ar, vuelvelas un array
else
$in_ar=array(); // no hay terminos ar haz un array vacio
$tmp=param('ma'); // obten las ma deseadas
if($tmp!="")
$in_ma=explode(',',$tmp); // si desea terminos ma, vuelvelas un array
else
$in_ma=array(); // no hay terminos ma haz un array vacio
hazTerminio($in_ar,'ar'); // cambia 1 por ar(1), 2 por ar(2) ...
hazTerminio($in_ma,'ma'); // cambia 1 por ma(1), 2 por ma(2) ...
$nar=sizeof($in_ar); // cuenta cuantas ar hay
$nma=sizeof($in_ma); // cuenta cuantas ma hay
$ar_ma_num=$nar+$nma; // total de terminos es numero de ar mas numero de ma
$com=array(); // incia el array final
if($nar) foreach($in_ar as $v) // si hay ar's pon unicamente cada ar
$com[]=$v;
if($nma) foreach($in_ma as $v) // si hay ma's pon unicamente cada ma
$com[]=$v;
$son=pow(2, $nar+$nma )-1; // calcula el numero total de combinaciones
// reporta que se va a realizar
echo "<ul>
<li>Combinaciones para

```

```

<ul>
<li>AR: ".print_r($in_ar,true)
."<li>MA; ".print_r($in_ma,true)
."</ul>
<li>Total combinaciones: ".number_format($son,0," ',';)
."<li>Programita de eviews<pre>";
// escribe la tabla de entrada de terminos a pegar en el programa de eviews
// no escribimos todas las combinaciones pues no soporta una tabla tan grande el eviews
echo <<< EVIEWS
'tabla de ar y ma seleccionadas
!arArrNum=$ar_ma_num
table(!arArrNum,1) { %arArr}
EIEWS;
foreach($com as $k=>$v)
echo PHP_EOL."{ %arArr}("$k+1).",1)=\trim($v)."\;
echo PHP_EOL;
echo "</pre><p>&nbsp;</p><hr>";
}
/**
* hazTerminio()
* Pasa de 1,2 a ar(1), ar(2) done ar es el valor de $ter
*
* @param array $arr vector de numeros a pasar a ar(#)
* @param termino $ter ar o ma
* @return
*/
function hazTerminio(&$arr,$ter) {
foreach($arr as $k=>$v)
$arr[$k]=$ter."("$v." ) ";
}

```

```

////////////////////////////////////
/// utilerias de php ///
////////////////////////////////////
/**
 * regresa el parametro llamado $name en post en string o array, si no existe busca en get,
 * de no estar regresa $dfft
 * @param string $nombre nombre del parametro a regresar
 * @param variable $dfft=" el valor a regresar de no existir el parametro. default "
 * @return string el parametro llamado $name en post o get array si es array, de no estar
regresa $dfft
 */
function param($name,$dfft=") { return param_post($name, param_get($name,$dfft) ); }
/**
 * regresa el parametro llamado $name en get en string o array, de no estar regresa $dfft
 * @param string $nombre nombre del parametro a regresar
 * @param variable $dfft=" el valor a regresar de no existir el parametro. default "
 * @return string el parametro llamado $name en get, de no estar regresa $dfft
 */
function param_get($name,$dfft=") {
    if( array_key_exists($name,$_GET) && !is_array($_GET[$name]) )
        return trim(urldecode($_GET[$name]));
    if( !array_key_exists($name,$_GET))
        return $dfft;
    $arr=$_GET[$name];
array_walk_recursive($arr,'arrayWalk_param_post_encode');
return $arr;
}
/**
 * regresa el parametro llamado $name en post en string o array, de no estar regresa $dfft
 * @param string $nombre nombre del parametro a regresar

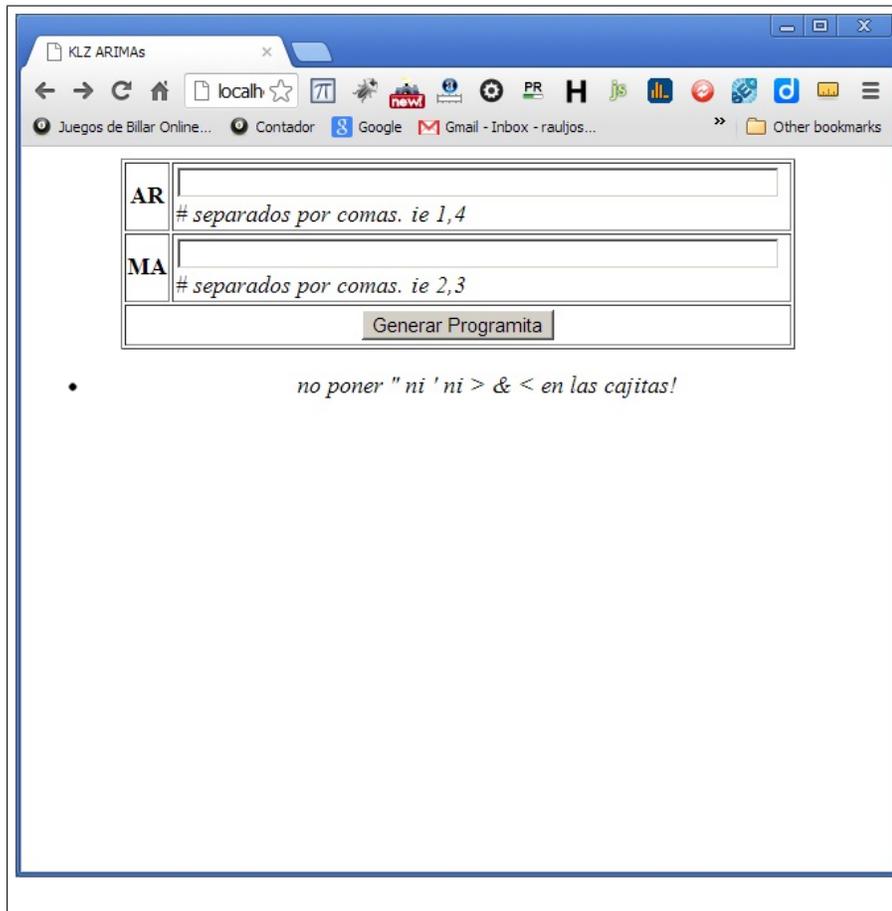
```

```

* @param variable $dflt=" el valor a regresar de no existir el parametro. default "
* @return string el parametro llamado $name en post, de no estar regresa $dflt
*/
function param_post($name,$dflt=") {
    if( array_key_exists($name,$_POST) && !is_array($_POST[$name]) )
        return trim(html_entity_decode($_POST[$name],ENT_QUOTES));
    if( !array_key_exists($name,$_POST) )
        return $dflt;
    $arr=$_POST[$name];
    array_walk_recursive($arr,'arrayWalk_param_post_encode');
    return $arr;
}
/**
* arrayWalk_param_post_encode()
*
* @param mixed $item
* @param mixed $key
* @return
*/
function arrayWalk_param_post_encode(&$item,$key) {
    $item=trim(html_entity_decode($item,ENT_QUOTES));
}
?>

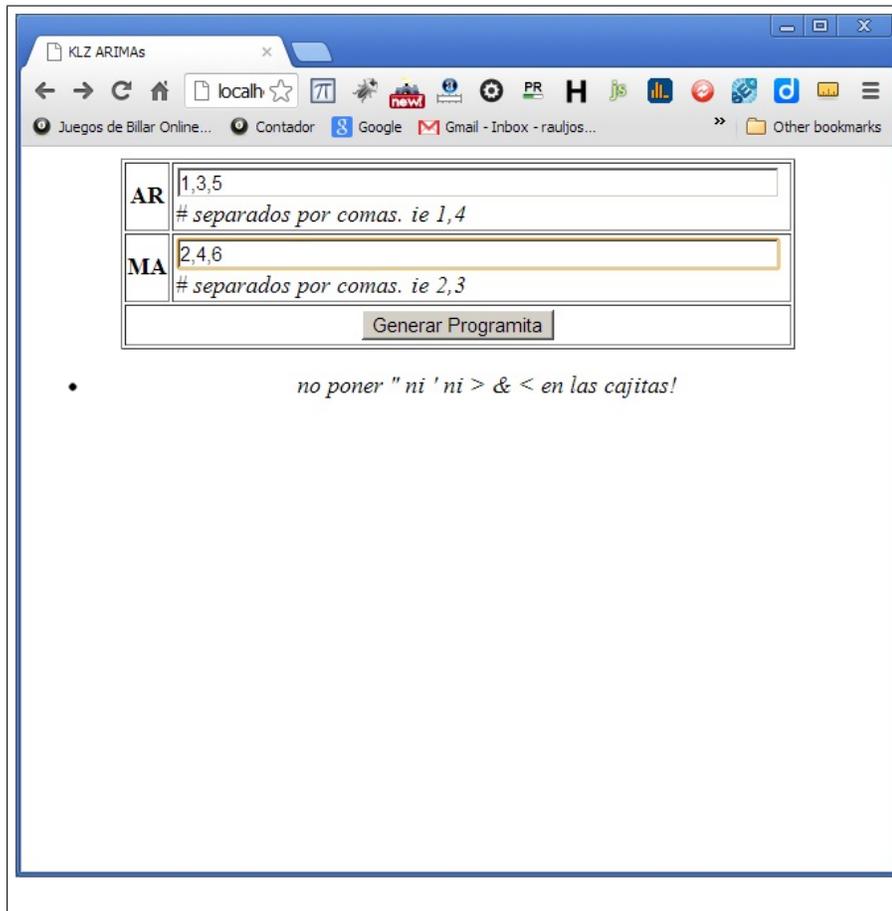
```

Una imagen de la vista de captura es:



Muestra de la vista en donde se captura la información

A continuación se muestra una vista con información antes de generar el programa:



Vista de captura de términos AR y MA a combinar

Finalmente muestro una pantalla de salida con el código generado y que se debe sustituir en el programa de *E – Views*.

La única parte que se copia y pega en el scrip del Anexo 4 en la sección que especifica pegar aquí es la parte que se encuentra debajo del bullet *Programita de E – Views*.





```

'Entre las comillas va el nombre de la variable dependiente
%dep = "DVOLTRIP" 'varibale dependiente
'Entre las comillas se ponen los nombres de las variables independientes (explicativas)
%regs = ' variables independientes separadas por espacios. SI HAY C
' debe ser la PRIMERA!
!testC=1 ' 1= no hay C o expulsa si Prob(C)>!probMin
' 2.- Si ya hay C y no pruebas Prob(c)
!probMin = 0.05 ' Probabilidad maxima aceptada del coeficiente
""""""""""
[PEGAR AQUI EL RESULTADO DE comb.php]
'OJO: No deben quedar ni los corchetes, ni esta nota.
""""""""""
""""""""""
' no cambiar
' inicializar las variables
!combnum=2^!arArrNum-1 ' numero de combinaciones
%armabest= ' equacion del mejor modelo hasta el momento
!mejorcrit = 1e11 ' valor grande para iniciar
!mincrit = 1e12 ' valor grande para iniciar
'tabla mejores
%mejoresTabla = "las_Mejores"
if @isobject(%mejoresTabla) then
delete { %mejoresTabla}
endif
table(!mejoresNum,2) { %mejoresTabla}
' vector mejores modelos
%tmp="mejoresVec"
if @isobject(%tmp) then
delete { %tmp}
endif

```

```

vector(!mejoresNum) mejoresVec
for !j=1 to !mejoresNum
{ %mejoresTabla}(!j,1)=
{ %mejoresTabla}(!j,2)=
mejoresVec(!j)!=mejorcrit
next
!da=0 ' auxiliar para bitwise and
"""
' subrutinas
' regresa 0 si bitwise de a y b da cero otro regesa uno
subroutine local klz_bit_and(scalar ainis,scalar binis,scalar da)
scalar a=ainis
scalar b=binis
' ambos nones? si si es true, pues ambos tienen el primer bit en 1
if @mod(a,2)=1 AND @mod(b,2)=1 then
da = 1
return 1
endif
' son iguales? si si es true, son los mismos bits
if a=b then
da = 1
return
endif
' 0 and lo que sea es cero, regresa cero
if a=0 OR b=0 then
da = 0
return
endif
vector(32) bbits
vector(32) abits

```

vector(32) bits ' potencias de dos

bits(32)=4294967296

bits(31)=2147483648

bits(30)=1073741824

bits(29)=536870912

bits(28)=268435456

bits(27)=134217728

bits(26)=67108864

bits(25)=33554432

bits(24)=16777216

bits(23)=8388608

bits(22)=4194304

bits(21)=2097152

bits(20)=1048576

bits(19)=524288

bits(18)=262144

bits(17)=131072

bits(16)=65536

bits(15)=32768

bits(14)=16384

bits(13)=8192

bits(12)=4096

bits(11)=2048

bits(10)=1024

bits(9)=512

bits(8)=256

bits(7)=128

bits(6)=64

bits(5)=32

bits(4)=16

```

bits(3)=8
bits(2)=4
bits(1)=2
scalar ib=0
' pasamos a a binario en abits y b a binario en bbits
for ib=32 to 1 step -1
if a >= bits(ib) then
abits(ib)=1
a = a - bits(ib)
else
abits(ib)=0
endif
if b >= bits(ib) then
bbits(ib)=1
b = b - bits(ib)
else
bbits(ib)=0
endif
next ib
'si el mismo bit esta prendido en ambos es true
for ib=1 to 32
if abits(ib)=1 AND bbits(ib)=1 then
da=1
return
endif
next ib
' no hubo el mismo bit prendido en ambos, es false
da=0
endsub
' calcula la arima con las ar y ma seleccionadas, si es el mejor criterio

```

```

' ponlo en el mejor modelo y si esta dentro de los mejores
' ponlo en la lista de los mejores
subroutine klz_ arima_ calcula
' inicializa las variables
!crit = 0 ' criterio de la ecuacion
!errorsVan=@errorcount ' registra errores al inicio
' calcula la ecuacion
equation { %eqname}.ls { %dep} { %regs} { %armarstring}
' determina si hubo error al calcular la ecuacion
!errorsSon=@errorcount
if !errorsSon > !errorsVan then
' hubo error reportalo
logmsg !van,0,0, %armarstring , !crit, Error
else
' no hubo, prosigue
' prepara las variables
' criterio de seleccion Schwarz
!crit = { %eqname}.@schwarz
' asume la ecuacion cumple con los requisitos
!ok=1
' numero de observaciones
!nobs = { %eqname}.@regobs
' obten informacion de los coeficientes
!ncf = { %eqname}.@ncoef
for !k=!testC to !ncf
!errorsVan=@errorcount
!tt = 99
!tt = { %eqname}.@tstats(!k)
!errorsSon=@errorcount
if !errorsSon > !errorsVan then

```

```

' hubo errores marca no cumple con los requisitos
!ok=0
' prosigue con la siguiente
exitloop
else
if @tdist(!tt,!nobs) > !probMin then
' coeficiente no significativo
' marca no cumple con los requisitos
!ok=0
' prosigue con la siguiente
exitloop
endif
' verifica no hubieron errores
!errorsSon=@errorcount
if !errorsSon > !errorsVan then
' hubo errores marca no cumple con los requisitos
!ok=0
' prosigue con la siguiente
exitloop
endif
endif
next
if !ok=1 then
' la ecuacion paso los criterios
' es el mejor?
if !crit < !mincrit then
' es la mejor hasta el momento
%armabest= %armarstring
' pon el minimo criterio en mincrit
!mincrit = !crit

```

```

' pon el schwarz en el mejor criterio
!mejorcrit = !crit
endif
' entra en los n (!mejoresNum) mejores?
' pon en maxj el indice del peor modelo
' de los n mejores
!maxj=1
for !j=2 to !mejoresNum
if mejoresVec(!j) > mejoresVec(!maxj) then
!maxj=!j
endif
next
if !crit < mejoresVec(!maxj) then
' si entra en los mejores
' escribe en el log que entro a los mejores
logmsg !van,1,1, %armarstring , !crit
' registrala en los mejores sobre la peor que habia
{%mejoresTabla}(!maxj,1)= %armarstring
{%mejoresTabla}(!maxj,2)=!crit
mejoresVec(!maxj)=!crit
else
' no entro solo reportala
logmsg !van,0,1, %armarstring , !crit
endif
else
' no cumplio los requisitos reportala
logmsg !van,0,0, %armarstring , !crit
endif
endif
' cada 2000 modelos guarda el log

```

```

if @mod(!van,2000)=0 then
logsave c:\tesis\log\klz_arima_log_04.txt
endif
' limpia la memoria
delete %eqname
endsub
' va armando cada combinacion y la manda calcular
subroutine klz_arima_best
' reporta las columnas de la bitacora o log
logmsg #,Entra a Min,Coefs Ok,ar/ma,Schwartz,Error,dep:,%dep,regs:,%regs, combina-
ciones:!,combnun
' inicializa las variables
!van=0 ' modelos calculados
!xi=0
!xj=0
!tmpi=0
!tmpj=0
!jklz=0
%klztmps =
%klzarmaterm =
!iklzLim = 2 ^!arArrNum ' combinaciones a correr +1
!iklz=1 ' contador
' bucle (loop) para calcular todas las combinaciones
while !iklz < !iklzLim
' inicializa los terminos de la equacion
%klzarmaterm =
' inicializa el contador
!jklz=0
' pon los terminos que corresponden a esta combinacion
while !jklz < !arArrNum

```

```

call klz_bit_and(!klz, 2 ^!jklz,!da)
' da trae 1 si en esta combinacion toca este termino
' da trae 0 si a esta combinacion no le toca el termino
if !da = 1 then
' incluye el termino en esta ecuacion
%klztmps = { %arArr}(!jklz + 1,1)
%klzarmaterm = %klzarmaterm + + %klztmps
endif
!jklz = !jklz + 1
wend
' actualiza cuantas van
!van = !klz
' guarda en armastring los terminos de esta ecuacion
%armarstring = %klzarmaterm
' calcula este modelo e incluye en los mejores en su caso
call klz_arima_calcula
' incrementa el contador
!klz = !klz + 1
wend
endsub
""""""""""""""""""""
' procesa
' muestra la tabla de los mejores
show { %mejoresTabla}
' guarda errores en el log
logmode error 1
' inicia el proceso
call klz_arima_best
' guarda la bitacora o log
logsave c:\tesis\log\klz_arima_log_04.txt

```

```

' guarda los resultados
output(c) c:\tesis\log\klz_arima_mejores_04.csv
show { %mejoresTabla}
' termina de registrar en el log
output off
' limpia la memoria
logclear
' reporta al log los mejores modelos
for !j=1 to !mejoresNum
%arma_son={ %mejoresTabla}(!j,1)
%arma_crit={ %mejoresTabla}(!j,2)
logmsg %arma_son, %arma_crit
next
' guarda el reporte en el log
logsave c:\tesis\log\klz_arima_logMejores_04.txt
' muestra al usuario los mejores modelos
show { %mejoresTabla}

```

# Anexo 5: Script de MATLAB modelo RNA univariado

El script se corre directo en la ventana de comandos de *MATLAB*, después de haber importado las matrices de las series con las que se va a trabajar.

Los renglones con el signo % implican que son comentarios y no se encuentran activos al momento de correr el script.

```
%% 1. Importing data
%S = load('magdata');
T = con2seq(voltript);
%Donde voltript es la variable de entrada y salida
%% 2. Data preparation
N = 12; % Numero de predicciones a realizar
% Input and target series are divided in two groups of data:
% 1st group: used to train the network
targetSeries = T(1:end-N);
% 2nd group: this is the new data used for simulation. inputSeriesVal will
% be used for predicting new targets. targetSeriesVal will be used for
% network validation after prediction
targetSeriesVal = T(end-N+1:end); % This is generally not available
%% 3. Network Architecture
delay =12;
neuronsHiddenLayer =2;
% Network Creation
net = narnet(1:delay,neuronsHiddenLayer);
%% 4. Training the network
[Xs,Xi,Ai,Ts] = preparets(net,{}, {},targetSeries);
% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
%net.divideFcn = 'dividerand'; % Divide data randomly
```

```

%net.divideMode = 'time'; % Divide up every value
%net.divideParam.trainRatio = 75/100;
%net.divideParam.valRatio = 15/100;
%net.divideParam.testRatio = 10/100;
% Choose a Training Function
% For a list of all training functions type: help ntrain
net.trainFcn = 'trainlm'; % Levenberg-Marquardt
% Choose a Performance Function
% For a list of all performance functions type: help nperformance
net.performFcn = 'mse'; % Mean squared error
net = train(net,Xs,Ts,Xi,Ai);
view(net)
Y = net(Xs,Xi,Ai);
% Performance for the series-parallel implementation, only
% one-step-ahead prediction
perf = perform(net,Ts,Y);
%% 5. Multi-step ahead prediction
targetSeriesPred = [targetSeries(end-delay+1:end), con2seq(nan(1,N))];
netc = closeloop(net);
view(netc)
[Xs,Xi,Ai,Ts] = preparets(netc,{}, {},targetSeriesPred);
yPred = netc(Xs,Xi,Ai);
perf = perform(net,yPred,targetSeriesVal);
figure;
plot([cell2mat(targetSeries),nan(1,N);
nan(1,length(targetSeries)),cell2mat(yPred);
nan(1,length(targetSeries)),cell2mat(targetSeriesVal)]')
legend('Original Targets','Network Predictions','Expected Outputs')

```

# Anexo 6: Script de MATLAB modelo RNA multivariado

El script se corre directo en la ventana de comandos de *MATLAB*, después de haber importado las matrices de las series con las que se va a trabajar.

Los renglones con el signo % implican que son comentarios y no se encuentran activos al momento de correr el script.

Este script es prácticamente igual al anterior, salvo que si existe variable *input (baset)* y la red no es *NAR* sino *NARX*.

```
%% 1. Importing data
%S = load('magdata');
X = con2seq(baset);
T = con2seq(voltript);
% Donde baset es la matriz de entradas con las variables explicativas.
% Donde voltript es la matriz de salidas (variable dependiente)
%% 2. Data preparation
N = 12; % Numero de predicciones a realizar
% Input and target series are divided in two groups of data:
% 1st group: used to train the network
inputSeries = X(1:end-N);
targetSeries = T(1:end-N);
% 2nd group: this is the new data used for simulation. inputSeriesVal will
% be used for predicting new targets. targetSeriesVal will be used for
% network validation after prediction
inputSeriesVal = X(end-N+1:end);
targetSeriesVal = T(end-N+1:end); % This is generally not available
%% 3. Network Architecture
% Delay es la cantidad de rezagos a utilizar
delay =12;
% NeuronsHiddenLayer es la cantidad de neuronas
```

```

neuronsHiddenLayer = 7;
% Network Creation
net = narxnet(1:delay,1:delay,neuronsHiddenLayer);
%% 4. Training the network
[Xs,Xi,Ai,Ts] = preparets(net,inputSeries,{},targetSeries);
% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
%net.divideFcn = 'dividerand'; % Divide data randomly
%net.divideMode = 'time'; % Divide up every value
%net.divideParam.trainRatio = 75/100;
%net.divideParam.valRatio = 15/100;
%net.divideParam.testRatio = 10/100;
% Choose a Training Function
% For a list of all training functions type: help ntrain
net.trainFcn = 'trainlm'; % Levenberg-Marquardt
% Choose a Performance Function
% For a list of all performance functions type: help nperformance
net.performFcn = 'mse'; % Mean squared error
net = train(net,Xs,Ts,Xi,Ai);
view(net)
Y = net(Xs,Xi,Ai);
% Performance for the series-parallel implementation, only
% one-step-ahead prediction
perf = perform(net,Ts,Y);
%% 5. Multi-step ahead prediction
inputSeriesPred = [inputSeries(end-delay+1:end),inputSeriesVal];
targetSeriesPred = [targetSeries(end-delay+1:end), con2seq(nan(1,N))];
netc = closeloop(net);
view(netc)
[Xs,Xi,Ai,Ts] = preparets(netc,inputSeriesPred,{},targetSeriesPred);

```

```
yPred = netc(Xs,Xi,Ai);
perf = perform(net,yPred,targetSeriesVal);
figure;
plot([cell2mat(targetSeries),nan(1,N);
nan(1,length(targetSeries)),cell2mat(yPred);
nan(1,length(targetSeries)),cell2mat(targetSeriesVal)]')
legend('Original Targets','Network Predictions','Expected Outputs')
```

# Bibliografía

- [1] Time Series Analysis: Forecasting and Control.  
G. P. E. Box y G. M. Jenkins.  
Revised ed., Holden Day,  
San Francisco, 1978
  
- [2] Econometría.  
Quinta edición.  
Damodar N. Gujarati.  
McGraw-Hill.  
2010.
  
- [3] An Introduction to Econometrics.  
Michael Pokorny.  
Basil Blackwell.  
New York, 1987
  
- [4] A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity.  
H. White.  
*Econometrica*, vol. 48, 1980.

- [5] Análisis Econométrico.  
Tercera edición.  
William H. Greene.  
Prentice Hall.  
2000
- [6] La Estructura industrial de México en 1960.  
Banco de México, S. A.  
Investigaciones industriales.  
Oficina de economía industrial.  
Departamento de investigaciones industriales.  
Oficina editorial 1967.
- [7] Matriz de insumo producto de México año 1970.  
Tomo 3 Actividades primarias y servicios.  
Secretaría de programación y presupuesto.
- [8] Distribution of the Estimators for Autoregressive Time Series with a Unit Root.  
Dickey, D.A. and W.A. Fuller.  
Journal of the American Statistical Association, Vol. 74.  
1979.
- [9] Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order.  
Said E. and David A. Dickey.  
Biometrika, Vol. 71  
1984.
- [10] Basic Econometrics.  
Fourth edition.  
Damodar N. Gujarati.  
McGraw-Hill.  
2003.

- [11] A logical calculus of the ideas immanent in neurons activity.  
W. S. McCulloch y W. Pitts.  
Bull. Math. Biophys.  
1943.
- [12] Neural Network Learning and Expert Systems.  
Third edition.  
Stephen I. Gallant.  
The MIT Press.  
1995.
- [13] The Perceptron a perceiving and recognizing automaton.  
Rosenblatt, Frank.  
Report 85-460-1.  
Cornell Aeronautical Laboratory.  
1957.
- [14] Neural Network Learning: Theoretical Foundations.  
Martin Anthony, Peter L. Bartlett.  
Cambridge University Press.  
2009.
- [15] Redes de Retropropagación.  
[http://rna.50webs.com/tutorial/RNA\\_backprop.html](http://rna.50webs.com/tutorial/RNA_backprop.html)
- [16] Approximations by superpositions of sigmoidal functions.  
Cybenko., G.  
Mathematics of Control, Signals, and Systems, 2 (4).  
1989.

- [17] MATLAB Student Version R2012a.  
Help Section  
Neural Network Toolbox  
Varias secciones.  
2012.
- [18] A Method for the Solution of Certain Non-Linear Problems in Least Squares.  
Kenneth Levenberg.  
Quarterly of Applied Mathematics 2.  
1944.
- [19] An Algorithm for Least-Squares Estimation of Nonlinear Parameters.  
Donald Marquardt.  
SIAM Journal on Applied Mathematics 11 (2)  
doi:10.1137/0111030.  
1963.
- [20] Neural Network - Multi Step Ahead Prediction.  
[www.mathworks.com/matlabcentral/answers/14970](http://www.mathworks.com/matlabcentral/answers/14970).