

**ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE CIENCIAS**

**INTELIGENCIA COMPUTACIONAL EN LA MODELACIÓN DE SERIES  
FINANCIERAS: ENFOQUE DE LA LÓGICA DIFUSA**

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERO MATEMÁTICO**

**CARLOS PATRICIO FUENMAYOR VITERI**

**patricio.fuenmayor@gmail.com**

**DIRECTOR: Mat. LUIS HORNA HUARACA Ph.D**

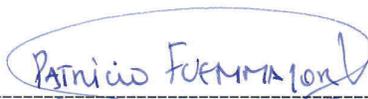
**luis.horna@epn.edu.ec**

**QUITO, Julio 2015**

## DECLARACIÓN

Yo, Carlos Patricio Fuenmayor Viteri, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



---

CARLOS PATRICIO FUENMAYOR VITERI

## CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Carlos Patricio Fuenmayor Viteri bajo mi supervisión.

A handwritten signature in blue ink, appearing to read 'Luis Horna Huaraca', is positioned above a horizontal dashed line.

Mat. LUIS HORNA HUARACA Ph.D

DIRECTOR

## **AGRADECIMIENTOS**

*Mis más sinceros agradecimientos:*

*A mis padres y hermana, por el apoyo y cariño incondicional.*

*A Hermann y Aston, por ser mis directrices en esta etapa de mi vida.*

*A la EPN y sus profesores, en especial al Dr. Luis Horna, por los conocimientos compartidos y el apoyo brindado.*

*Y a tod@s que forman parte de mis vivencias.*

# DEDICATORIA

A

*mi familia, amistades, conocidos y todos los seres que de una u otra forma han  
influenciado en mi vida...*

# ÍNDICE DE CONTENIDO

<b>LISTA DE FIGURAS</b>	<b>IV</b>
<b>LISTA DE TABLAS</b>	<b>VI</b>
<b>RESUMEN</b>	<b>VII</b>
<b>ABSTRACT</b>	<b>IX</b>
<b>1. INTRODUCCIÓN AL ANÁLISIS, MODELACIÓN Y OBTENCIÓN DE PRONÓSTICOS EN SERIES FINANCIERAS</b>	<b>1</b>
1.1. CARACTERÍSTICAS DE SERIES FINANCIERAS . . . . .	1
1.2. TÉCNICAS DE MODELAMIENTO . . . . .	2
1.3. TÉCNICAS DE PRONÓSTICOS . . . . .	5
<b>2. PRONÓSTICOS AUTOMÁTICOS</b>	<b>8</b>
2.1. FUNDAMENTOS TEÓRICOS DE SELECCIÓN Y ESTIMACIÓN DE PARÁMETROS . . . . .	9
2.1.1. SELECCIÓN DEL ORDEN DE LOS MODELOS USANDO PRUEBAS DE RAÍCES UNITARIAS Y AIC . . . . .	9
2.1.2. PROCEDIMIENTO PARA “ATRAVESAR” EL ESPACIO DE MODELOS . . . . .	10
2.2. IMPLEMENTACIÓN DE ALGORITMOS Y HERRAMIENTAS COMPUTACIONALES . . . . .	12
2.2.1. EL ENTORNO R . . . . .	12
2.2.1.1. Clases de objetos . . . . .	14
2.2.1.2. Procedimientos gráficos . . . . .	15

2.2.1.3. Extensiones y paquetes . . . . .	15
2.2.2. PAQUETE forecast PARA R, DE PRONÓSTICOS AUTOMÁTICOS DE SERIES DE TIEMPO . . . . .	16
<b>3. CONCEPTOS DE LÓGICA DIFUSA PARA AGRUPAMIENTO E INFERENCIA</b>	<b>18</b>
3.1. INTRODUCCIÓN AL ENFOQUE DE LA LÓGICA DIFUSA . . . . .	18
3.1.1. CONJUNTOS DIFUSOS Y FUNCIONES DE PERTENENCIA . . . . .	19
3.1.2. SISTEMAS DIFUSOS . . . . .	20
3.1.2.1. Sistemas difusos tipo Mamdani . . . . .	21
3.1.2.2. Sistemas difusos tipo Takagi-Sugeno . . . . .	22
3.1.3. INFERENCIA EN SISTEMAS DIFUSOS . . . . .	22
3.2. GENERACIÓN DE REGLAS POR AGRUPAMIENTO . . . . .	23
3.2.1. ALGORITMO DE AGRUPAMIENTO DIFUSO PARA LA GENERA- CIÓN DE REGLAS . . . . .	24
3.2.1.1. Elementos de la teoría de agrupamiento . . . . .	24
3.2.1.2. Partición fuerte . . . . .	25
3.2.1.3. Partición difusa . . . . .	26
3.2.1.4. Agrupamiento difuso <i>c</i> -medias . . . . .	26
<b>4. SISTEMA HÍBRIDO DIFUSO PARA PRONÓSTICOS DE SERIES FINANCIERAS</b>	<b>28</b>
4.1. DESCRIPCIÓN DE LAS SERIES FINANCIERAS . . . . .	29
4.2. MODELACIÓN TRADICIONAL DE LAS SERIES FINANCIERAS . . . . .	31
4.3. MODELACIÓN EN BASE AL SISTEMA HÍBRIDO DIFUSO DE LAS SERIES FINANCIERAS . . . . .	34
4.3.1. MODELACIÓN . . . . .	34
4.3.2. IMPLEMENTACIÓN . . . . .	38
4.4. COMPARACIÓN DE METODOLOGÍAS Y RESULTADOS . . . . .	42
4.4.1. SERIE CARTERA . . . . .	42

4.4.1.1.	Modelación Tradicional Cartera . . . . .	42
4.4.1.2.	Modelación Sistema Híbrido Difuso Cartera . . . . .	46
4.4.2.	SERIE MONETARIOS . . . . .	54
4.4.2.1.	Modelación Tradicional Monetarios . . . . .	54
4.4.2.2.	Modelación Sistema Híbrido Difuso Monetarios . . . . .	59
4.4.3.	SERIE AHORROS . . . . .	66
4.4.3.1.	Modelación Tradicional Ahorros . . . . .	66
4.4.3.2.	Modelación Sistema Híbrido Difuso Ahorros . . . . .	71
4.4.4.	SERIE PLAZO . . . . .	78
4.4.4.1.	Modelación Tradicional Plazo . . . . .	78
4.4.4.2.	Modelación Sistema Híbrido Difuso Plazo . . . . .	82

**5. CONCLUSIONES Y RECOMENDACIONES** **89**

## LISTA DE FIGURAS

4.4.1.1. Serie de cartera y funciones ACF y PACF . . . . .	43
4.4.1.2. Retardos de la serie de cartera . . . . .	43
4.4.1.3. Serie de cartera diferenciada de orden 1 . . . . .	44
4.4.1.4. Residuos y funciones ACF y PACF de cartera . . . . .	45
4.4.1.5. Pronósticos de la serie de cartera . . . . .	46
4.4.1.6. Ajustes de los modelos de cartera . . . . .	47
4.4.1.7. Residuos de los modelos de cartera . . . . .	48
4.4.1.8. Dispersión de residuos y residuos relativos de cartera . . . . .	49
4.4.1.9. Inferencia en los pronósticos de cartera . . . . .	50
4.4.1.10. Pruebas de pronósticos con datos reales de cartera . . . . .	52
4.4.1.11. Visualización conjunta de resultados de cartera . . . . .	53
4.4.2.1. Serie de monetarios y funciones ACF y PACF . . . . .	55
4.4.2.2. Retardos de la serie de monetarios . . . . .	55
4.4.2.3. Serie de monetarios diferenciada de orden 1 . . . . .	56
4.4.2.4. Residuos y funciones ACF y PACF de monetarios . . . . .	57
4.4.2.5. Pronósticos de la serie de monetarios . . . . .	58
4.4.2.6. Descomposición de la serie monetarios . . . . .	58
4.4.2.7. Ajustes de los modelos de monetarios . . . . .	60
4.4.2.8. Residuos de los modelos de monetarios . . . . .	61
4.4.2.9. Dispersión de residuos y residuos relativos de monetarios . . . . .	62
4.4.2.10. Inferencia en los pronósticos de monetarios . . . . .	63
4.4.2.11. Pruebas de pronósticos con datos reales de monetarios . . . . .	64
4.4.2.12. Visualización conjunta de resultados de monetarios . . . . .	65
4.4.3.1. Serie de ahorros y funciones ACF y PACF . . . . .	67

4.4.3.2. Retardos de la serie de ahorros . . . . .	67
4.4.3.3. Serie de ahorros diferenciada de orden 1 . . . . .	68
4.4.3.4. Residuos y funciones ACF y PACF de ahorros . . . . .	69
4.4.3.5. Pronósticos de la serie de ahorros . . . . .	69
4.4.3.6. Descomposición de la serie de ahorros . . . . .	70
4.4.3.7. Ajustes de los modelos de ahorros . . . . .	72
4.4.3.8. Residuos de los modelos de ahorros . . . . .	73
4.4.3.9. Dispersión de residuos y residuos relativos por modelo de ahorros . . .	74
4.4.3.10. Inferencia en los pronósticos de ahorros . . . . .	75
4.4.3.11. Pruebas de pronósticos con datos reales de ahorros . . . . .	76
4.4.3.12. Visualización conjunta de resultados de ahorros . . . . .	77
4.4.4.1. Serie de plazo y funciones ACF y PACF . . . . .	79
4.4.4.2. Retardos de la serie de plazo . . . . .	79
4.4.4.3. Serie de plazo diferenciada de orden 1 . . . . .	80
4.4.4.4. Residuos y funciones ACF y PACF de plazo . . . . .	81
4.4.4.5. Pronósticos de la serie de plazo . . . . .	81
4.4.4.6. Ajustes de los modelos de plazo . . . . .	83
4.4.4.7. Residuos de los modelos de plazo . . . . .	84
4.4.4.8. Dispersión de residuos y residuos relativos por modelo de plazo . . . .	85
4.4.4.9. Inferencia en los pronósticos de plazo . . . . .	86
4.4.4.10. Pruebas de pronósticos con datos reales de plazo . . . . .	87
4.4.4.11. Visualización conjunta de resultados de plazo . . . . .	88

## LISTA DE TABLAS

4.4.1.1. Resultados de pruebas en los pronósticos de cartera . . . . .	51
4.4.2.1. Resultados de pruebas en los pronósticos de monetarios . . . . .	62
4.4.3.1. Resultados de pruebas en los pronósticos de ahorros . . . . .	74
4.4.4.1. Resultados de pruebas en los pronósticos de plazo . . . . .	82

## RESUMEN

En el análisis financiero bancario existen muchos aspectos que deben manejarse eficientemente para obtener un buen desempeño. Una buena decisión, brinda siempre oportunidades de negocio y crecimiento. Generalmente, esta decisión es el resultado de un análisis objetivo de los factores relevantes en un estudio. El manejo de escenarios en la toma de decisiones es una herramienta necesaria para el tratamiento de incertidumbre. Estos escenarios consideran suposiciones, estimaciones, modelaciones, etc.; obtenidos en base a criterios técnicos y/o de negocio e intentan describir el comportamiento futuro de la institución en un ámbito específico.

Técnicamente, la elección del mejor modelo se basa en la optimización del criterio de información elegido como referente. Este mejor modelo se utiliza para realizar los pronósticos en un horizonte de tiempo determinado. En la práctica, los valores estimados no necesariamente son corroborados por el criterio de expertos; por esta razón la búsqueda de un modelo que conjugue el criterio de expertos, como también, un ajuste estadísticamente aceptado, constituye una tarea necesaria.

La flexibilidad en el criterio de información proporciona un conjunto de modelos posibles, cuyos ajustes y pronósticos en una visualización global pueden presentar una tendencia, la que es fácilmente analizada y descrita si se utiliza un lenguaje versátil y eficiente como es R, en el cual se han desarrollado procedimientos y funciones con este fin.

En el presente trabajo se propone modelar series financieras mediante la generación de una lista de modelos lineales que son elegidos en base a criterios de búsqueda y de discriminación por la minimización del criterio de información de Akaike. En base a esta lista se procede a realizar pronósticos para cada modelo, de tal forma que se puede tener un espectro visual global. Para caracterizar la tendencia y el patrón, se utilizan técnicas de agrupamiento difusas, dando como resultado un pronóstico medio.

El procedimiento es validado con las series financieras de: cartera, monetarios, ahorros y plazo, de una institución financiera; las mismas que por su naturaleza presentan comportamientos distintos y son utilizadas frecuentemente en escenarios de liquidez, de presupuestación, etc. Los resultados obtenidos son coherentes y coinciden con el punto de vista de los expertos. La metodología propuesta constituye un primer paso para la estructuración de los denominados sistemas híbridos, que se caracterizan por la complementación de dos o más técnicas entre sí. Como parte de este tipo de sistemas se tienen: los de enfoque neuro-difuso, algoritmos genéticos adaptativos, sistemas difusos envolventes, etc.; y en este caso, pronósticos automáticos y sistemas difusos. Esta metodología permite manejar varios escenarios tomando en cuenta múltiples criterios, y así disponer de mayor información para la toma de decisiones. Además, la automatización del proceso y la incorporación de parámetros que controlan el nivel de ajuste permiten manejar el sistema dinámicamente.

## ABSTRACT

In banking financial analysis there are many aspects that must be managed efficiently to achieve good performance. A proper decision always provides good business opportunities. The decision is usually making as the result of an objective function analysis of relevant some factors. Management scenarios for decision making is a necessary tool in the treatment of uncertainty. These scenarios considered assumptions, estimates, models, etc., which are obtained using technical criteria and/or business criteria, and try to forecast the behavior in a specific field. The choice of the best model is based on the optimization of the information criterion which was chosen as a reference. The best model is used to forecast a specific time horizon. In practice, estimated values are not necessarily the ones that experts expect. This is why searching for a model that combines the experts opinions as well as a statistically acceptable fit become a necessary task.

Adaptability in the model selection criterion provides a set of possible options. One of them is visualize the a tendency in fitting and forecasting. This trend is easily analyzed and described when using a versatile and efficient language such as R in which procedures and functions have been developed for this purpose.

In this thesis we propouse to generate a list of linear models which are chosen based on search criteria and discrimination by minimizing the Akaike information criterion. Based on this we proceed to make forecasts for each model, so we set a complete overview of the spectrum. To characterize the trend and pattern fuzzy clustering techniques are used. This result in an mean forecast.

Numerical experiments are performed with the financial series of: loans, accounts, savings accounts and certificates of deposit of a well known financial institution in the country, show good performace of the proposed method. The time series have different behaviors and they are frequently used in liquidity scenarios, budgeting, etc. The results are consistent with expert's opinion. The proposed methodology is a first step in the

structuring of so-called hybrid systems, which combine two or more techniques together, e.g., neuro-fuzzy approach, adaptive genetic algorithms, fuzzy systems enclosures, etc. In this case, automated forecasting and fuzzy systems. This methodology gives the possibility to handle multiple scenarios taking into account multiple criteria and thus have more information for decision making. Moreover, automation of the process and the incorporation of parameters that control the level of adjustment allows manage the system dynamically.

## INTRODUCCIÓN

En el análisis financiero bancario existen muchos aspectos que deben manejarse eficientemente para obtener un buen desempeño. Una buena decisión, brinda siempre oportunidades de negocio y crecimiento. Generalmente, esta decisión es el resultado de un análisis objetivo de los factores relevantes en un estudio. El manejo de escenarios en la toma de decisiones es una herramienta necesaria para el tratamiento de incertidumbre. Estos escenarios consideran suposiciones, estimaciones, modelaciones, etc.; obtenidos en base a criterios técnicos y/o de negocio e intentan describir el comportamiento futuro de la institución en un ámbito específico.

En lo que respecta a la modelación, se utilizan diversas teorías, técnicas, metodologías e incluso tan solo la opinión de expertos. Este conjunto de herramientas y criterios para modelar, desemboca en una diversidad de resultados que no necesariamente son los esperados, o que de una u otra forma, se desea que fuesen mejores. Técnicamente, la elección del mejor modelo se basa en la optimización del criterio de información elegido como referente. Este mejor modelo se utiliza para realizar los pronósticos en un horizonte de tiempo determinado. En la práctica, los valores estimados no necesariamente son corroborados por el criterio de expertos; por esta razón la búsqueda de un modelo que conjugue este criterio, y, un ajuste estadísticamente aceptado, se transforma en una tarea necesaria.

La flexibilidad en el criterio de información proporciona un conjunto de modelos posibles, cuyos ajustes y pronósticos en una visualización global pueden presentar una tendencia. El tratamiento conjunto de los datos obtenidos requiere de la utilización de técnicas diseñadas con este fin y que se enmarcan en el denominado mineo de datos, campo en el cual la inteligencia computacional ha dado resultados satisfactorios. Entre las técnicas utilizadas para obtener información de los datos, se tiene la de agrupamiento

difuso, la que clasifica a los datos en base a características similares, tomando en cuenta los conceptos de la lógica difusa.

La implementación del modelamiento y del tratamiento de los datos resulta fácil si se utiliza un lenguaje versátil y eficiente como es R, en el cual se han desarrollado procedimientos y funciones con este fin. Por ser R un proyecto de código abierto, es posible acceder a las fuentes, brindando la posibilidad de adaptar las funciones ya desarrolladas a las necesidades específicas del usuario.

En el presente trabajo se propone modelar series financieras mediante la generación de una lista de modelos lineales que son elegidos en base a criterios de búsqueda y de discriminación por la minimización del criterio de información de Akaike. En base a esta lista se procede a realizar pronósticos para cada modelo, de tal forma que se puede tener un espectro visual global. Para caracterizar la tendencia y el patrón, se utilizan técnicas de agrupamiento difusas, dando como resultado un pronóstico medio. Lo anterior permite entender los patrones de tendencia y estacionalidad de una mejor forma. Un aspecto fundamental de esta metodología, es la exclusión de los modelos que son considerados por los expertos como modelos de pronósticos no admisibles, debido a que el comportamiento no va acorde con características de la serie y/o del negocio, como son: la tendencia, estacionalidad, etc.

El procedimiento es validado con las series financieras de: cartera, monetarios, ahorros y plazo, de una institución financiera; las mismas que por su naturaleza presentan comportamientos distintos y son utilizadas frecuentemente en escenarios de liquidez, de presupuestación, etc. Los resultados obtenidos son coherentes y coinciden con el punto de vista de los expertos. La metodología propuesta constituye un primer paso para la estructuración de los denominados sistemas híbridos, que se caracterizan por la complementación de dos o más técnicas entre sí. Como parte de este tipo de sistemas se tienen: los de enfoque neuro-difuso, algoritmos genéticos adaptativos, sistemas difusos envolventes, etc.; y en este caso, pronósticos automáticos y sistemas difusos. Esta metodología permite manejar varios escenarios tomando en cuenta múltiples

criterios, y así disponer de mayor información para la toma de decisiones. Además, la automatización del proceso y la incorporación de parámetros que controlan el nivel de ajuste permiten manejar el sistema dinámicamente.

Para una mejor comprensión del sistema híbrido difuso para el pronóstico de series financieras propuesto, este trabajo está organizado de la siguiente manera: en el Capítulo 1 se realiza una revisión de los conceptos principales de la teoría de series temporales; específicamente lo que tiene que ver con la utilización de modelos SARIMA. En el Capítulo 2 se describe el proceso de obtención de pronósticos automáticos y la utilización de la implementación de estas metodologías en R; también se desarrolla una pequeña introducción a las características esenciales de este lenguaje. En el Capítulo 3 se revisan los algoritmos de agrupamiento difuso, que son el paso previo a la utilización de los criterios de ponderación para las inferencias. Luego en el Capítulo 4, que constituye el aporte fundamental de este trabajo, se presenta la modelación en base al sistema híbrido difuso en las cuatro series financieras nombradas anteriormente y se lo valida respecto a la modelación tradicional descrita en la primera parte del Capítulo. Finalmente conclusiones y recomendaciones se presentan en el Capítulo 5.

# 1 INTRODUCCIÓN AL ANÁLISIS, MODELACIÓN Y OBTENCIÓN DE PRONÓSTICOS EN SERIES FINANCIERAS

Analizar datos temporales y pronosticar sus valores futuros, son los más e importantes problemas que un analista tiene que enfrentar en distintos campos, que van desde: finanzas, economía, ingeniería, etc.; hasta administración, análisis político y social, psicología, etc. Consecuentemente, existe un gran grupo de personas en estos ámbitos que necesitan entender algunos conceptos básicos del análisis y pronósticos, motivo por el cual, la teoría de series de tiempo se ha desarrollado y popularizado grandemente en los últimos años. Sería redundante realizar su desarrollo teórico, por lo que este Capítulo está enfocado a dar una guía de los aspectos fundamentales que se deben considerar al momento de analizar series de tiempo.

## 1.1 CARACTERÍSTICAS DE SERIES FINANCIERAS

Una serie de tiempo es un conjunto de observaciones en una variable, llamada  $y$ , en intervalos consecutivos de tiempo. Si  $t = 1, \dots, T$  denota el tiempo, entonces  $y_t$  representa la realización de la variable  $y$  en el tiempo  $t$ . El conjunto de observaciones de 1 a  $T$  es escrita  $[y_1, \dots, y_T]$  y  $T$  representa el tamaño de la muestra.

Las series pueden ser clasificadas en base a sus características en:

- Continuas y discretas: si la información es constantemente almacenada, se denomina continua; y discreta, cuando la información es observada en intervalos de tiempo. Hay dos tipos de series discretas: si los intervalos de tiempo son equidistantes y si no lo son.

- Univariantes y multivariantes: si la serie consiste en una sola observación en cada punto de tiempo se denomina univariante, caso contrario, se denomina multivariantes.
- Estacionarias y no estacionarias: estacionaria, si la irregularidad de la serie se expresa por modelos estocásticos; si la estructura estocástica cambia en el tiempo es no estacionaria.
- Gaussianas y no gaussianas: si la distribución de la serie sigue una distribución normal, la serie es llamada gaussiana, caso contrario, no gaussiana.
- Lineal y no lineal: si la serie se expresa por medio de modelos lineales se denomina lineal, de otra forma, se denomina no lineal,
- Series con valores faltantes y/o atípicos: cuando los datos de la serie están incompletos, y si existe un comportamiento inusual en la serie.

En base a estas características se han desarrollado diferentes técnicas para modelarlas. Es de gran importancia poder identificar estas características en la serie que se va a analizar, ya que, en base a esto se puede elegir el método más adecuado y conveniente para modelarla.

Para mayor detalle, revisar [22].

## 1.2 TÉCNICAS DE MODELAMIENTO

Una hipótesis básica del análisis y modelamiento es suponer que cualquier aspecto del comportamiento pasado puede continuar y repetirse en el futuro; por este motivo todas las técnicas basan su desarrollo en el análisis de los datos para poder encontrar patrones de comportamiento, que puedan servir para explicar la serie.

La modelación tradicional contempla dos enfoques, el determinístico y el estocástico.

En el **enfoque determinístico**, un importante paso para el análisis de series de tiempo es considerar los tipos de patrones en los datos para aplicar el modelo adecuado

para estos patrones. Existen cuatro tipos de componentes que se pueden distinguir en las series:

- horizontal: cuando los datos fluctúan alrededor de un valor constante,
- tendencia: cuando hay largos períodos de crecimiento o decrecimiento en la serie,
- estacionalidad: cuando una serie es influenciada por factores estacionales y recurrentes en un periodo regular,
- cíclico: cuando los datos muestran crecimientos o caídas, pero no en periodos fijos.

Muchas de las series incluyen combinaciones de los patrones descritos. Luego de ser separados estos patrones en cualquier serie de tiempo, un patrón de residuos es identificable, que forma el componente aleatorio o error. Una cruda pero práctica forma de descomposición de la serie de datos (ignorando el componente cíclico) está dado por el modelo de descomposición de tipo aditivo:  $y_t = T_t + S_t + E_t$  o de tipo multiplicativo:  $y_t = T_t S_t E_t$  donde  $y_t$  es la serie de tiempo original, y:

- $T_t$  es el componente de tendencia, que es la dirección general de la variable en el periodo de observación, es decir el cambio a largo plazo de la media de la serie.
- $S_t$  es el componente de estacionalidad, que corresponde a fluctuaciones periódicas de la variable, en periodos relativamente cortos de tiempo.
- $E_t$  es el componente irregular, que después de extraer de la serie la tendencia y variaciones cíclicas, queda una serie de valores residuales, que pueden ser o no totalmente aleatorios.

Para más información ver [21].

En el **enfoque estocástico** se contempla la teoría de los modelos SARIMA o también llamados modelos lineales de Box-Jenkins (1970). Estos modelos se basan en los denominados modelos regresivos, que son modelos construidos usando análisis de

regresión, que es una colección de métodos para estudiar la relación entre las variables y para estimar y pronosticar los valores de una variable usando los valores de otras variables incorporadas en una serie de tiempo conjunta (Draper and Smith, 1981). Como parte de los modelos regresivos se tienen los siguientes tipos de modelos:

- Modelos autoregresivos (AR): estos expresan el valor actual de una serie de tiempo por un agregado finito lineal de los valores pasados y por un desplazamiento.
- Modelos medias móviles (MA): estos expresan el valor actual de una serie de tiempo por un promedio móvil simple de los errores pasados y por un desplazamiento.
- Modelos autoregresivos medias móviles (ARMA): que están constituidos por una componente autoregresiva y otra de media móvil.
- Modelos autorregresivos integrados de medias móviles (ARIMA): son modelos no estacionarios debido a que poseen raíces unitarias, y que pueden ser transformados a estacionarios, mediante diferenciaciones ( $\Delta$ ) de orden  $d$ ; una vez diferenciados tienen la forma ARMA.
- Modelos autorregresivos integrados de medias móviles estacionales (SARIMA): son la forma ampliada de los modelos ARIMA, que incorporan la presencia de factores estacionales. De igual manera, pueden ser transformados a la forma ARMA mediante diferenciaciones estacionales de orden  $D$  y de diferenciaciones estacionarias de orden  $d$ .

Existen funciones y estadísticas definidas en la teoría que son herramientas básicas para poder identificar estos modelos, se detallan las utilizadas en este trabajo:

- Función de autocorrelación (ACF): se define como la correlación cruzada de una serie consigo misma.

- Función de autocorrelación parcial (PACF): considera la correlación entre dos variables separadas entre sí “j” períodos y en función de los valores intermedios entre ellas.
- Prueba de raíces unitarias: para determinar si la serie de tiempo tiene una raíz unitaria. Las pruebas pueden ser: la de Dickey-Fuller, la de Phillips-Perron o la prueba KPSS.

Para mayor detalle, revisar [15, 16].

### **1.3 TÉCNICAS DE PRONÓSTICOS**

Existen dos grandes grupos de técnicas de pronósticos: las cualitativas y las cuantitativas.

- Cualitativas: son subjetivas en naturaleza y requiere el criterio de expertos; se usan en situaciones en que no se dispone de suficientes datos históricos en los que se base el pronóstico. Tal vez la técnica de predicción cualitativa más formal y ampliamente conocida es la del Método Delphi, la cual emplea un panel de expertos para analizar el problema.
- Cuantitativas: son las que hacen uso formal de los datos históricos y de un modelo de pronósticos. El modelo formalmente describe un patrón en los datos y expresa una relación estadística entre los valores pasados y presentes de la variable; entonces el modelo es usado para proyectar el patrón en el futuro.

Hay varios tipos de modelos de pronósticos de uso general; los tres más utilizados son los modelos de regresión, modelos de suavizado y modelos de series de tiempo.

- Los modelos de regresión o causales, hacen uso de las relaciones entre la variable de interés y una o más variables predictivas relacionadas. El método de los mínimos cuadrados es la base formal de la mayoría de los modelos de regresión.

- Los modelos de suavizado emplean típicamente una simple función de las observaciones anteriores para proporcionar un pronóstico de la variable de interés. Estos métodos pueden tener una base estadística formal, pero se usa frecuentemente una heurística justificada de fácil uso y que produce resultados satisfactorios.
- Los modelos de series de tiempo utilizan las propiedades estadísticas de los datos históricos para especificar un modelo formal y luego estimar el parámetro desconocido de este modelo (usualmente) por mínimos cuadrados.

Para más información ver [23].

El presente trabajo está enfocado en el uso modelos de series de tiempo para pronósticos.

Los pasos básicos en el modelado y en la obtención de pronósticos de una serie temporal son las siguientes:

1. Graficar la serie de tiempo y determinar sus características básicas, como la tendencia y/o el comportamiento estacional. Se pueden buscar posibles valores atípicos o cualquier indicación de que la serie de tiempo ha cambiado con respecto a sus características básicas sobre el período de tiempo observado.
2. Eliminar los componentes de tendencia o de temporada, ya sea por diferenciación o utilizando un modelo apropiado para los datos. También considerar el uso de transformaciones de datos en particular si la variabilidad en la serie de tiempo parece ser proporcional al nivel promedio de la serie. El objetivo de estas operaciones es producir un conjunto de residuos estacionarios.
3. Desarrollar un modelo de pronósticos de los residuos. No es raro encontrar que hay varios modelos plausibles y análisis adicionales se tendrán que realizar para determinar el mejor a ser implementado. Algunas veces los modelos potenciales

pueden ser eliminados en base a su ajuste a los datos históricos. Es poco probable que un modelo que se ajusta mal sirva para obtener buenos pronósticos.

4. Validar los resultados del modelo (o modelos) de la etapa anterior. Esto probablemente implicará algún tipo de división de la muestra o el procedimiento de validación cruzada. El objetivo de este paso es seleccionar un modelo para el uso del pronóstico.
5. También de interés son las diferencias entre la serie original y los valores pronosticados por el modelo en la escala original para esto se deben revertir las transformaciones y los ajustes de diferenciación hecha para eliminar las tendencias o los efectos estacionales.
6. Si los intervalos de confianza se definen para el pronóstico (se recomienda hacer esto), también hacerlo para los residuos y luego revertir las transformaciones hechas para producir los residuos como se describió anteriormente.
7. Desarrollar e implementar un procedimiento de control de los pronósticos para garantizar que el deterioro en el rendimiento serán detectados rápidamente. El seguimiento de los pronósticos se realiza mediante la evaluación de los errores de pronóstico que se experimenten.

## 2 PRONÓSTICOS AUTOMÁTICOS

El gran desarrollo computacional de la última década permite afrontar y manejar tareas exhaustivas en todos los ámbitos de la ciencia e ingeniería; este desarrollo se refleja también en el campo financiero; estructuras complejas de datos permiten el tratamiento de extensos volúmenes de información; concretamente, en la modelación de series financieras, la automatización de procesos de búsqueda de modelos óptimos es un campo de constante desarrollo; puesto que la propuesta de nuevos enfoques, algoritmos y procedimientos permiten obtener resultados precisos y eficientes.

En este campo han sido un gran aporte en el desarrollo de metodologías para la obtención de pronósticos automáticos, los trabajos que se resumen en [1] realizados por: Hannan-Rissanen (1982) [2], proponen la estimación de parámetros de un modelo autoregresivo de media móvil con la minimización de un criterio de información. Estos resultados fueron extendidos por Gómez (1998) a modelos ARIMA, el que conjuntamente con Maravall (1998) implementaron este procedimiento en el programa TRAMO SEATS. Lui (1989) propuso un método para la identificación de modelos SARIMA utilizando métodos de filtrado y reglas heurísticas (programa SCA-Expert). Existen otros algoritmos implementados en programas comerciales, cuyo desempeño es competitivo, los cuales no están documentados para el público; por ejemplo, Forecast Pro.

En este Capítulo se revisa el trabajo de Hyndman y Khandakar (2008), que consiste en la implementación de pronósticos automáticos de series de tiempo, en el paquete *forecast* para R, para más detalle revisar [1].

## 2.1 FUNDAMENTOS TEÓRICOS DE SELECCIÓN Y ESTIMACIÓN DE PARÁMETROS

### 2.1.1 SELECCIÓN DEL ORDEN DE LOS MODELOS USANDO PRUEBAS DE RAÍCES UNITARIAS Y AIC

Un proceso no estacional  $ARIMA(p, d, q)$  está dado por

$$\phi(B)(1 - B^d)y_t = c + \theta(B)\varepsilon_t,$$

donde, los  $\{\varepsilon_t\}$  son un proceso de ruido blanco con media cero y varianza  $\sigma^2$ ,  $B$  es el operador de retardos, y  $\phi(z)$ ,  $\theta(z)$  son polinomios de orden  $p$  y  $q$  respectivamente. Para asegurar la causalidad y la invertibilidad, se asume que  $\phi(z)$  y  $\theta(z)$  no tienen raíces en  $|z| < 1$  (Brockwell y Davis 1991). Si  $c \neq 0$ , hay un polinomio implícito de orden  $d$  en la función de pronóstico.

El proceso estacional de frecuencia  $m$ , denotado  $ARIMA(p, d, q)(P, D, Q)_m$  esta dado por

$$\Phi(B^m)\phi(B)(1 - B^m)^D(1 - B)^d y_t = c + \Theta(B^m)\theta(B)\varepsilon_t,$$

donde  $\Phi(z)$  y  $\Theta(z)$  son polinomios de ordenes  $P$  y  $Q$  respectivamente, que no contienen raíces unitarias en el círculo unidad. Si  $c \neq 0$ , existe un polinomio implícito de orden  $d + D$  en la función de pronóstico.

La tarea principal en un pronóstico automático ARIMA es seleccionar un modelo apropiado para los valores  $p, q, P, Q, D$  y  $d$ . Si  $d$  y  $D$  son conocidos, se puede seleccionar el orden de  $p, q, P$  y  $Q$  en base al criterio de información de Akaike (AIC):

$$AIC = -\log(L) + 2(p + q + P + Q + k),$$

donde,  $k = 1$  si  $c \neq 0$  y  $k = 0$  si  $c = 0$ ;  $L$  es la máxima verosimilitud del modelo ajustado para los datos *diferenciados*  $(1 - B^m)(1 - B)^d y_t$ .

Para la selección de los parámetros  $d$  y  $D$  se utiliza la prueba de raíces unitarias. Para

datos no estacionales, se considera el modelo  $ARIMA(p, d, q)$  donde  $d$  es seleccionado en base a sucesivas pruebas KPSS de raíces unitarias; es decir, si la prueba en los datos para una raíz unitaria resulta significativa, se realiza la prueba en los datos diferenciados; y así sucesivamente. El procedimiento se detiene cuando se obtiene el primer resultado insignificante.

Para datos estacionales, se considera el modelo  $ARIMA(p, d, q)(P, D, Q)_m$  donde  $m$  es la frecuencia de estacionalidad y  $D = 0$  o  $D = 1$  dependiendo de la prueba de Canova-Hansen (Canova y Hansen 1995). Después de seleccionar  $D$ , se escoge  $d$  aplicando sucesivas pruebas KPSS de raíces unitarias en los datos diferenciados estacionalmente (si  $D = 1$ ) o en los datos originales (si  $D=0$ ) para minimizar el AIC.

Para mayor detalle, revisar [1, 21].

### **2.1.2 PROCEDIMIENTO PARA “ATRAVESAR” EL ESPACIO DE MODELOS**

Se supone que se dispone de datos estacionales y se considera el modelo lineal  $ARIMA(p, d, q)(P, D, Q)_m$  donde  $p$  y  $q$  pueden tomar valores entre 0 y 3, y  $P$  y  $Q$  pueden tomar valores de 0 a 1. Cuando  $c = 0$  existe un total de 288 modelos posibles, y cuando  $c \neq 0$  existe un total de 192 posibles modelos, dando un total de 480 modelos. Si los valores  $p, d, q, P, D$  y  $Q$  permiten una gama más amplia, el número de modelos aumenta rápidamente. Consecuentemente, a menudo no es factible ajustar a todos los modelos posibles y elegir el que tiene el menor AIC. Se requiere una forma de atravesar el espacio de modelos eficientemente con el fin de llegar el modelo con menor valor de AIC.

Hyndman y Khandakar (2008) [1] proponen el siguiente algoritmo paso a paso:

- Paso 1: para iniciar se prueban los siguientes modelos
  - $ARIMA(2, d, 2)$  si  $m = 1$  y  $ARIMA(2, d, 2)(1, D, 1)$  si  $m > 1$ .
  - $ARIMA(0, d, 0)$  si  $m = 1$  y  $ARIMA(0, d, 0)(0, D, 0)$  si  $m > 1$ .
  - $ARIMA(1, d, 0)$  si  $m = 1$  y  $ARIMA(1, d, 0)(1, D, 0)$  si  $m > 1$ .
  - $ARIMA(0, d, 1)$  si  $m = 1$  y  $ARIMA(0, d, 1)(0, D, 1)$  si  $m > 1$ .

Si  $d+D \leq 1$ , estos modelos son ajustados con  $c \neq 0$ ; caso contrario, se fija  $c = 0$ . De estos cuatro modelos, se selecciona el de menor valor de AIC. Este es llamado el modelo "actual", es notado por  $ARIMA(p, d, q)$  si  $m = 1$  o  $ARIMA(p, d, q)(P, D, Q)_m$  si  $m \geq 1$ .

- Paso 2: se considera las trece variaciones del modelo actual:
  - donde solo uno de los valores de  $p, q, P$  y  $Q$  se le permite variar entre  $\pm 1$  para el modelo actual;
  - donde  $p$  y  $q$  varían en  $\pm 1$  para el modelo actual;
  - donde  $P$  y  $Q$  varían en  $\pm 1$  para el modelo actual;
  - donde la constante  $c$  es incluida si el modelo actual tiene  $c = 0$ , o se excluye si el modelo actual tiene  $c \neq 0$ .

Cada vez que un modelo con menor AIC es encontrado, se convierte en el nuevo modelo "actual" y se repite el procedimiento. Este proceso termina cuando no se puede encontrar un modelo con menor AIC que el actual.

Hay algunas restricciones sobre los modelos ajustados para evitar problemas de convergencia o de raíces unitarias. Las restricciones están dadas en forma general a continuación:

- Los valores de  $p$  y  $q$  no se les permite exceder los límites superiores (por defecto valores de 5 en cada caso).
- Los valores de  $P$  y  $Q$  no se les permite exceder los límites superiores (por defecto valores de 2 en cada caso).
- Se rechaza el modelo que esté “cerca” a la no invertibilidad o a la no causalidad.
- Si existe muchos errores generados en la rutina de optimización no lineal usada para la estimación, el modelo es rechazado. La razón es que si un modelo es difícil de ajustar quizá no es buen modelo para los datos.

El algoritmo garantiza encontrar un modelo válido porque el espacio de modelos es finito y al menos uno de los modelos de arranque será aceptado (el modelo sin parámetros para  $AR$  o  $MA$ ). El modelo seleccionado es usado para generar los pronósticos. Para mayor información ver [1].

## **2.2 IMPLEMENTACIÓN DE ALGORITMOS Y HERRAMIENTAS COMPUTACIONALES**

### **2.2.1 EL ENTORNO R**

R es un lenguaje y entorno de programación para el tratamiento de datos, realizar análisis estadístico y graficación. Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993. Su desarrollo actual es responsabilidad del R Development Core Team, [28].

Se trata de un proyecto de software libre, resultado de la implementación GNU del premiado lenguaje S. R y S-Plus (versión comercial de S) son, probablemente, los dos lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy populares en el campo de la investigación biomédica, la bioinformática y las

matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con finalidades específicas.

Entre otras características dispone de:

- almacenamiento y tratamiento efectivo de datos,
- operadores para cálculo sobre variables indexadas (arrays), en particular matrices,
- una amplia, coherente e integrada colección de herramientas para análisis de datos,
- posibilidades gráficas para análisis de datos, que funcionan directamente sobre pantalla o impresora, y
- un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y posibilidad de entradas y salidas. Debe destacarse que muchas de las funciones suministradas con el sistema están escritas en el lenguaje R.

El término “*entorno*” lo caracteriza como un sistema completamente diseñado y coherente, antes que como una agregación incremental de herramientas muy específicas e inflexibles, como ocurre frecuentemente con otros programas de análisis de datos. R es en gran parte un vehículo para el desarrollo de nuevos métodos de análisis interactivo de datos. Como tal es muy dinámico y las diferentes versiones no siempre son totalmente compatibles con las anteriores. Algunos usuarios prefieren los cambios debido a los nuevos métodos y tecnología que los acompañan, a otros sin embargo les molesta ya que algún código anterior deja de funcionar. Aunque R puede entenderse como un lenguaje de programación, los programas escritos en R deben considerarse esencialmente efímeros.

R es un lenguaje Orientado a Objetos, bajo este complejo término se esconde la simplicidad y flexibilidad de R. El hecho que R es un lenguaje de programación puede desanimar a muchos usuarios que piensan que no tienen “alma de programadores”. Esto no es necesariamente cierto por dos razones; primero R es un lenguaje interpretado

(como Java) y no compilado (como C, C++, Fortran, Pascal, etc), lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin necesidad de construir ejecutables.

Para mas detalle ver [24, 28].

### 2.2.1.1. Clases de objetos

Los vectores son el tipo básico de objeto en R, pero existen más tipos. Realizando un resumen de los tipos más importantes descritos en [24, 25, 28] se tienen:

- las matrices o, más generalmente, variables indexadas (*arrays*): son generalizaciones multidimensionales de los vectores; de hecho, son vectores indexados por dos o más índices y que se imprimen de modo especial,
- los factores: sirven para representar datos categóricos,
- las listas: son una forma generalizada de vector, en las cuales los elementos no tienen por qué ser del mismo tipo y a menudo son a su vez vectores o listas. Las listas permiten devolver los resultados de los cálculos estadísticos de un modo conveniente,
- las hojas de datos (*dataframes*): son estructuras similares a una matriz, en que cada columna puede ser de un tipo distinto a las otras. Las hojas de datos son apropiadas para describir “matrices de datos” donde cada fila representa a un individuo y cada columna una variable, las que pueden ser numéricas o categóricas. Muchos experimentos se describen muy apropiadamente con hojas de datos: los tratamientos son categóricos pero la respuesta es numérica,
- las funciones: son también objetos de R que pueden almacenarse en el espacio de trabajo, lo que permite extender las capacidades de R fácilmente.

### 2.2.1.2. Procedimientos gráficos

Las posibilidades gráficas son un componente de R muy importante y versátil. Es posible utilizarlas para mostrar una amplia variedad de gráficos estadísticos y también para construir nuevos tipos de gráficos. Las órdenes gráficas se dividen en tres grupos básicos:

- alto nivel: son funciones que crean un nuevo gráfico, posiblemente con ejes, etiquetas, títulos, etc.,
- bajo nivel: son funciones que añaden información a un gráfico existente, tales como puntos adicionales, líneas y etiquetas,
- interactivas: son funciones que permiten interactuar con un gráfico, añadiendo o eliminando información, utilizando un dispositivo apuntador, como un ratón.

Además, en R existe una lista de *parámetros gráficos* que pueden utilizarse para adaptar los gráficos.

### 2.2.1.3. Extensiones y paquetes

El sistema de paquetes de R ha sido uno de los factores clave del éxito global del proyecto. Los paquetes permiten una fácil, transparente y multi-plataforma extensión del sistema de base R. Un paquete de R puede ser considerado como el equivalente en *software* de un artículo científico: los artículos son el estándar para comunicar los resultados científicos, y los lectores esperan que estén en un formato determinado. Los paquetes de R son entonces (después de una fase de aprendizaje cortas) una manera cómoda para mantener las colecciones de funciones y conjuntos de datos. En un artículo se distribuye ideas científicas, en un paquete de R se distribuye la metodología estadística. Los paquetes son también una manera conveniente para mantener las funciones privadas y compartirlas con colegas. Por ser R parte de un proyecto colaborativo y abierto, sus usuarios pueden publicar paquetes que extienden

su configuración básica. Existe un repositorio oficial de paquetes cuyo número superó en otoño de 2009 la cifra de los 2000. Dado el enorme número de nuevos paquetes, éstos se han organizado en vistas (o temas), que permiten agruparlos según su naturaleza y función; por ejemplo, hay grupos de paquetes relacionados con estadística bayesiana, econometría, series temporales, etc. Para facilitar el desarrollo de nuevos paquetes, se ha puesto a servicio de la comunidad una forja de desarrollo que facilita las tareas relativas a dicho proceso.

### **2.2.2 PAQUETE *forecast* PARA R, DE PRONÓSTICOS AUTOMÁTICOS DE SERIES DE TIEMPO**

Los creadores de este paquete son: Rob J. Hyndman y Yeasmin Khandakar, ambos de la Universidad de Monash en Australia. La creación del paquete *forecast*, según sus autores, surge por la necesidad de generar pronósticos automáticos de grandes cantidades de series univariantes, que son requeridos en los negocios, convirtiéndose de esta manera en una herramienta esencial. Los algoritmos automáticos para pronósticos pueden determinar un apropiado modelo para la serie de tiempo, estimando los parámetros y calculando los pronósticos. Los más populares algoritmos para pronósticos automáticos están basados en suavizamiento exponencial o en modelos ARIMA, para mayor detalle sobre el contenido del paquete y sus especificaciones técnicas, revisar [1] y la ayuda del paquete<sup>1</sup>.

Para el presente trabajo es de interés la implementación realizada utilizando modelos ARIMA, ya que estos modelos tienen una estructura paramétrica, que es más flexible para ser incorporada en un sistema. Las tres funciones utilizadas y sus parámetros más relevantes son las siguientes:

- `auto.arima()`: la función realiza una búsqueda del mejor modelo posible en el marco de orden y condiciones establecidas. Los parámetros básicos que requiere son:

---

<sup>1</sup>R permite visualizar las ayudas proporcionadas en los paquetes, la función a utilizar es : `help()`

- x: serie de tiempo univariante
  - ic: criterio de información a ser usado en la selección del modelo. Los criterios disponibles son: “aic”, “aicc” y “bic”; y que corresponden a los criterios: de Akaike, al de Akaike de segundo orden y Bayesiano.
  - trace: si es *TRUE*, la lista de modelos ARIMA considerados es reportada.
- Arima(): ajusta un modelo ARIMA a una serie de tiempo univariante. Los parámetros básicos que requiere son:
    - x: serie de tiempo univariante
    - order: una especificación de la parte no estacional del modelo ARIMA de la forma  $(p, d, q)$ .
    - seasonal: una especificación de la parte estacional del modelo ARIMA más el período, expresado a manera de lista:  $(P, D, Q)$ , período.
    - include.drift: si el modelo ARIMA puede incluir un término de desplazamiento lineal. Por defecto *FALSE*.
  - forecast(): es una función genérica para calcular los pronósticos en base a modelos. Los parámetros básicos que requiere son:
    - object: una serie de tiempo o un modelo de serie de tiempo para el cual los pronósticos son requeridos.
    - h: número de períodos para pronosticar.
    - level: nivel de confianza para intervalos de pronósticos.

Existen varias funciones adicionales que brindan facilidad en el cálculo y en el tratamiento de datos, y que su adecuada utilización depende de la experiencia del usuario.

### **3 CONCEPTOS DE LÓGICA DIFUSA PARA AGRUPAMIENTO E INFERENCIA**

El ser humano debido a su constante deseo de entender su entorno, ha estudiado los sistemas naturales y ha aprendido mucho de ellos; de manera que, en base a los principios y características esenciales de estos sistemas, se han desarrollado nuevos algoritmos y modelos para resolver problemas complejos que se presentan en la ciencia y en la tecnología, problemas que con los métodos matemáticos convencionales resultan difíciles de tratar.

Metodologías que se han inspirado en diversos sistemas naturales, tales como: las redes neuronales biológicas, los procesos evolutivos, el comportamiento de organismos sociales, también llamados enjambres, y el proceso del pensamiento humano, los que respectivamente han dado los fundamentos para el desarrollo de conceptos como: las redes neuronales artificiales, la computación evolutiva o genética, la inteligencia de enjambres o colonias y los sistemas difusos, ver [6, 8]. El conjunto de estas metodologías se entiende como “inteligencia computacional”<sup>1</sup>.

#### **3.1 INTRODUCCIÓN AL ENFOQUE DE LA LÓGICA DIFUSA**

Tradicionalmente la teoría de conjuntos requiere que un elemento cualquiera sea, o no, parte de un conjunto. De la misma manera, la lógica binaria requiere que el valor del parámetro sea 0 o 1, con similares restricciones en los resultados de un sistema de inferencia. El razonamiento humano, sin embargo, no es siempre exacto y bivalente, debido a que la observación y el razonamiento incluyen una medida de incertidumbre.

Los conjuntos difusos y la lógica difusa permiten que se trate un razonamiento

---

<sup>1</sup>Términos utilizados por Bezdek en 1992

aproximado, con alguna incertidumbre. En los conjuntos difusos, un elemento pertenece a un conjunto con un cierto grado de certeza. La lógica difusa permite razonamientos e inferencias considerando este factor de incertidumbre.

Una parte de la metodología planteada está basada en la teoría de la lógica difusa. El término “difuso” fue introducido por Lotfi A. Zadeh (1965) en su trabajo sobre conjuntos difusos y su propuesta de una lógica multivaluada denominada lógica difusa. En la actualidad se ha convertido en un campo de investigación muy importante debido a la amplia gama de problemas que se pueden resolver eficientemente con esta técnica.

La lógica difusa modela sistemas bajo incertidumbre, como por ejemplo: ¿Es una persona delgada?, ¿A qué edad se le considera a una persona vieja?, etc. Estos problemas no pueden ser tratados por la lógica tradicional (bivalente) pues es necesario un espectro más amplio de valores, por ejemplo la premisa “yo estudiaré mañana cálculo” puede ser parcialmente verdadera o parcialmente falsa. Uno de los precursores de la lógica difusa, Zimmermann (1993) [14], propuso utilizar técnicas difusas para modelar sistemas complejos bajo incertidumbre, sistemas no lineales, o en general sistemas gobernados por reglas heurísticas, es decir, “criterio de expertos”.

En la literatura especializada existen muchas aplicaciones de estas técnicas a problemas de diferente naturaleza, como por ejemplo: control de sistemas (tráfico, centrales térmicas, ascensores), optimización (fenómenos físicos, geofísicos, sociales), en reconocimiento de patrones, sistemas de información (bases de datos, sistemas expertos), etc. Sin embargo, la opinión general de los expertos es que estas técnicas no han sido suficientemente explotadas en finanzas, área en la cual los resultados obtenidos hasta ahora son particularmente buenos.

### **3.1.1 CONJUNTOS DIFUSOS Y FUNCIONES DE PERTENENCIA**

Para poder introducir el concepto de un conjunto difuso, se debe tener claro el significado de un conjunto clásico.

**Conjunto puntual o clásico:** surge de forma natural, por la necesidad del ser humano de clasificar objetos y conceptos. Dentro de este conjunto se define una *función de pertenencia*  $\mu_C(x)$ ,  $x \in X$ , siendo  $X$  el universo de información (discurso). La restricción de la función es de que  $\mu_C : X \rightarrow \{0, 1\}$ . Expresado matemáticamente como

$$\mu_C(x) = \begin{cases} 1 & \text{si } x \in C \\ 0 & \text{si } x \notin C \end{cases}$$

**Conjunto difuso:** se define de forma similar que el conjunto clásico, con la diferencia de que se relaja la restricción de la función, esto es que  $\mu_F : X \rightarrow [0, 1]$ . Cuando más cerca esté  $\mu_F(x)$  del valor 1, mayor será la pertenencia del objeto  $x$  al conjunto  $F$ , es decir

$$\mu_F(x) = \begin{cases} 1 & \text{si } x \in F \text{ (pertenece completamente)} \\ ]0, 1[ & \text{si } x \in F \text{ (pertenece parcialmente)} \\ 0 & \text{si } x \notin F \text{ (no pertenece totalmente)} \end{cases}$$

La *función de pertenencia* es la idea clave introducida en la teoría de conjuntos difusos para medir el grado en el cual los elementos de un conjunto difuso coinciden en propiedades específicas, generándose de esta manera las denominadas *reglas*, que son un conjunto de condicionantes que se aplican a las variables. Existen varios tipos o formas para las funciones de pertenencia, tales como: Triangular, Gamma, Gaussiana, etc. También dentro de esta definición de conjuntos, se ha desarrollado toda una teoría, desde operaciones, normas, relaciones, etc., hasta sistemas basados en esta lógica, para mayor detalle se puede revisar [5].

### 3.1.2 SISTEMAS DIFUSOS

Los sistemas difusos, también llamados sistemas de lógica difusa, tienen una directa relación con los conceptos difusos, y son únicos en el sentido de que pueden procesar datos numéricos y lingüísticos al mismo tiempo. Desde el punto de vista matemático, un sistema difuso es una asignación no lineal de un vector de características de *entrada*

(datos) en una *salida* que en general es escalar, ver [5, 7].

Un sistema difuso toma una variable de entrada ( $X$ ) y la “codifica” o “difumina”, es decir, se determina su grado de pertenencia en el conjunto difuso de entrada. Esto alimenta al sistema de inferencia difuso, que en base al conjunto de reglas difusas guardadas, que son condicionantes de la forma *SI-ENTONCES* (que es lo que caracteriza al comportamiento del sistema), genera salidas difusas, que luego de ser “decodificadas”, dan lugar a las variables de salida del sistema ( $Y$ ). En la mayoría de aplicaciones la salida del sistema corresponde a un escalar, en general la salida está representada por un elemento difuso.

El *sistema de inferencia difuso* es el núcleo del sistema difuso, ya que es la forma por la cual se obtiene la *base de reglas difusas*, ver [4, 5]. En la práctica, los sistemas de inferencia difusos más utilizados son los desarrollados por:

- Mamdani
- Takagi-Sugeno

### **3.1.2.1. Sistemas difusos tipo Mamdani**

Mamdani (1977) propuso el primer sistema de inferencia difuso con el objetivo de controlar una máquina de vapor y su caldero, basado en un conjunto de reglas lingüísticas de control extraídas de las pruebas de laboratorio. Este tipo de reglas son conocidas como *reglas difusas tipo Mamdani*. Las características principales de estas reglas, son que tanto las premisas como las consecuencias son difusas.

A modo de ejemplo, se detalla la aplicación a la ley de Boyle de los gases, que considera que los estados de la presión ( $P$ ) de una masa de gas dada, varía inversamente proporcional al volumen ( $V$ ), manteniendo la temperatura ( $T$ ) constante.

Regla-1: **SI**  $P$  es BAJO y  $T$  es CONSTANTE, **ENTONCES**  $V$  es ALTO

Regla-2: **SI**  $P$  es MEDIO y  $T$  es CONSTANTE, **ENTONCES**  $V$  es MEDIO

Regla-3: **SI**  $P$  es ALTO y  $T$  es CONSTANTE, **ENTONCES**  $V$  es BAJO

Para más información, ver [4, 5]

### 3.1.2.2. Sistemas difusos tipo Takagi-Sugeno

En las reglas difusas de los sistemas Takagi-Sugeno, las premisas son difusas, mientras que las consecuencias corresponden a una función escalar, por ejemplo, una ecuación lineal. La ley de los gases, propuesta en la subsección anterior, se escribe como:

Regla-1: **SI**  $P$  es BAJO y  $T$  es CONSTANTE, **ENTONCES**  $V = a_1P + b_1T + c_1$

Regla-2: **SI**  $P$  es MEDIO y  $T$  es CONSTANTE, **ENTONCES**  $V = a_2P + b_2T + c_2$

Regla-3: **SI**  $P$  es ALTO y  $T$  es CONSTANTE, **ENTONCES**  $V = a_3P + b_3T + c_3$

donde las constantes  $a_i, b_i, c_i$  con  $i = 1, 2, 3$ , son parámetros de las correspondientes reglas  $Regla_i, i = 1, 2, 3$ .

Para más información, ver [4, 5]

### 3.1.3 INFERENCIA EN SISTEMAS DIFUSOS

Inferencia es el proceso de generación de salidas de conjuntos difusos cuando las reglas son difusas y las entradas están dadas. Las salidas difusas pueden ser decodificadas si se requiere valores de salidas puntuales; para este propósito preferentemente se usa el método de centro de gravedad (CG). Dado un conjunto difuso  $F$  representado de la forma:

$$F = \{\mu_F(x_1)/x_1, \mu_F(x_2)/x_2, \dots, \mu_F(x_p)/x_p\},$$

el método CG ayuda a calcular las coordenadas de  $x$  en el centro de gravedad del conjunto  $F$  de la siguiente manera:

$$x' = \frac{\sum_{i=1}^p \mu_F(x_i) \cdot x_i}{\sum_{i=1}^p \mu_F(x_i)}.$$

Específicamente los modelos de inferencia se desarrollan en el producto cartesiano  $X \times Y$  (entradas  $\times$  salidas), tomando en cuenta las diferentes implicaciones tanto para los sistemas Mamdani como Takagi-Sugeno. Estos tipos de inferencias se vuelven un problema cuando el número de datos es grande (mayores a 500), por el hecho de que la generación automática de las reglas, debe realizar un particionamiento equidistante para definir los subconjuntos difusos y sus respectivas funciones de pertenencia. Se han propuesto varias formas de reducir esta generación masiva de funciones de pertenencia, desde el uso de funciones más estructuradas, hasta un trato a priori del conjunto de entradas.

Una de las técnicas usadas para poder reducir los problemas en la generación de las funciones de pertenencia, es la de agrupamientos. El detalle de esta forma de generación de reglas automáticas es el tema a tratar en la siguiente sección.

Para mayor información ver [4].

### 3.2 GENERACIÓN DE REGLAS POR AGRUPAMIENTO

Los métodos automáticos que conducen a la generación de reglas trabajan considerablemente bien para la modelación y predicción de series de tiempo no lineales. Sin embargo, las reglas bases difusas generadas de esta manera, son muy extensas, porque cada par de *entrada-salida* genera una regla difusa; esto limita el uso de datos extensos, ya que se requiere altas características computacionales. Como alternativa se plantea el *algoritmo de agrupamiento difuso*, el cual usa pocas reglas difusas para el modelamiento, ver [4, 5].

### 3.2.1 ALGORITMO DE AGRUPAMIENTO DIFUSO PARA LA GENERACIÓN DE REGLAS

Los algoritmos de agrupamiento son herramientas matemáticas usadas para la identificación natural de grupos en los datos, basados en comunes similitudes, para que de un conjunto extenso de datos, se pueda obtener una representación concisa del comportamiento de este sistema. Uno de los algoritmos es conocido como el de *c-medias*, revisar [10].

#### 3.2.1.1. Elementos de la teoría de agrupamiento

Estas técnicas agrupan los datos en espacios característicos, es decir datos con características similares pertenecen a la misma clase. Los métodos de agrupamiento son independientes de la naturaleza de los datos; es decir, se pueden aplicar a datos cualitativos, cuantitativos y mixtos. En el presente trabajo nos enfocaremos en el agrupamiento de datos cuantitativos. Se asumirá que cada observación consiste de  $n$  variables, agrupadas en vectores columna  $n$ -dimensionales

$$\mathbf{Z}_s = [Z_{1s}, Z_{2s}, \dots, Z_{ns}]^T, \mathbf{Z}_s \in \mathbb{R}^n.$$

Un conjunto de  $N$  observaciones está descrito por  $\mathbf{Z} = \{Z_s | s = 1, 2, \dots, N\}$ , formándose una matriz de  $n \times N$  denominada *matriz de patrones* o *matriz de datos*  $Z$ . A las filas y columnas de esta matriz se les denomina *características* o *atributos*, y *patrones* u *objetos* respectivamente.

Un grupo se define como un subconjunto de objetos similares entre si (Bezdek (1981); Jain and Dube (1988)), donde el término “similar” se define matemáticamente. Por ejemplo, en espacios métricos, la similitud está definida como el promedio de distancias dada una cierta métrica, tomada entre los datos, o entre los puntos y un centro de referencia o de grupos. Los centros de los grupos usualmente no son conocidos, por lo que deben también ser calculados por el algoritmo.

Varios algoritmos de agrupamiento han sido propuestos. Los algoritmos de clasifi-

cación definidos en base a la teoría tradicional de conjuntos, clasifican a los objetos individualmente de acuerdo a su cercanía al grupo, lo cual se denomina *agrupamiento fuerte*. Esta partición de datos es tal que un objeto puede ser miembro de un solo grupo en particular.

Los algoritmos de agrupamiento difuso, sin embargo, permiten que los objetos pertenezcan a varios grupos a la vez, pero con diferentes grados de pertenencia, que es una forma más natural a comparación del agrupamiento fuerte.

### 3.2.1.2. Partición fuerte

La partición fuerte puede ser considerada como un grupo de subconjuntos formulados en los términos de un conjunto clásico. El objetivo del agrupamiento fuerte es partir el conjunto de datos  $Z = \{Z_1, Z_2, \dots, Z_N\}$  en  $c$  subgrupos denominados *clases*. Se asume que el valor de  $c$  es conocido. Usando conjuntos clásicos, una partición fuerte de  $Z$  puede ser definida como una familia de subconjuntos  $\{A_g | 1 \leq g \leq c\}$  con las siguientes propiedades:

$$\begin{aligned} \bigcup_{g=1}^c A_g &= Z, \\ A_g \cap A_h &= \emptyset, \quad 1 \leq g \neq h \leq c, \\ \emptyset &\subset A_g \subset Z, \quad 1 \leq g \leq c. \end{aligned}$$

En términos de funciones de pertenencia, una partición puede ser representada en una matriz de partición:  $U = [\mu_{sg}]_{c \times N}$ . Los elementos de esta matriz deben cumplir las siguientes condiciones:

$$\begin{aligned} \mu_{gs} &\in \{0, 1\}, \quad 1 \leq g \leq c; \quad 1 \leq s \leq N; \\ \sum_{g=1}^c \mu_{gs} &= 1, \quad 1 \leq s \leq N; \\ 0 < \sum_{s=1}^N \mu_{gs} &< N, \quad 1 \leq g \leq c. \end{aligned}$$

El espacio de todas las posibles matrices de partición fuerte de  $Z$  llamado espacio de

particionamiento fuerte se define por

$$M_{hc} = \left\{ U \in \mathbb{R}^{c \times N} \mid \mu_{gs} \in \{0, 1\}, \forall g, s; \sum_{g=1}^c \mu_{gs} = 1, \forall s; 0 < \sum_{s=1}^N \mu_{gs} < N, \forall g \right\}.$$

Para más información, ver [5, 10]

### 3.2.1.3. Partición difusa

Una partición difusa puede ser considerada como una generalización de una partición fuerte y esto conlleva directamente a que los valores de  $\mu_{gs}$  estén en el intervalo  $[0, 1]$  (Babuška, 1996). Al igual que para la matriz de partición fuerte se detalla las condiciones que esta debe cumplir:

$$\begin{aligned} \mu_{gs} &\in [0, 1], \quad 1 \leq g \leq c; \quad 1 \leq s \leq N; \\ \sum_{g=1}^c \mu_{gs} &= 1, \quad 1 \leq s \leq N; \\ 0 < \sum_{s=1}^N \mu_{gs} &< N, \quad 1 \leq g \leq c. \end{aligned}$$

La  $g$ -ésima fila de la matriz de partición  $U$  contiene los valores de la función de pertenencia  $\mu_s$  del subconjunto  $A_g$  de  $Z$ . El espacio de particionamiento difuso de  $Z$  es

$$M_{fc} = \left\{ U \in \mathbb{R}^{c \times N} \mid \mu_{gs} \in [0, 1], \forall g, s; \sum_{g=1}^c \mu_{gs} = 1, \forall s; 0 < \sum_{s=1}^N \mu_{gs} < N, \forall g \right\}.$$

Para más información, ver [5, 10]

### 3.2.1.4. Agrupamiento difuso $c$ -medias

El algoritmo de agrupamiento difuso  $c$ -medias es uno de los más populares para la generación automática de reglas difusas. La minimización del funcional  $c$ -medias representa un problema de optimización no lineal que en primera instancia puede ser resuelto con el método simple de iteración de Picard conocido como el algoritmo FCM.

Dado el conjunto de datos  $Z = \{Z_1, Z_2, \dots, Z_N\}$ , se deben seleccionar: el número de grupos  $1 < c < N$ , el factor de peso exponencial (también llamado exponente difuso)  $m > 1$ , el nivel de tolerancia  $\varepsilon > 0$  y la matriz de norma-inducida  $A$ . Iniciar la matriz de partición aleatoriamente, tal que  $U^{l=0} \in M_{fc}$ . Repetir las iteraciones para  $l = 1, 2, 3, \dots$

- **Paso 1:** calcular los grupos prototipos o grupos centros (medias)

$$v_g^{(l)} = \frac{\sum_{s=1}^N (\mu_{gs}^{(l-1)})^m \cdot Z_s}{\sum_{s=1}^N (\mu_{gs}^{(l-1)})^m}; 1 \leq g \leq c.$$

- **Paso 2:** calcular las distancias:

$$D_{gsA}^2 = (Z_s - v_g^{(l)})^T A (Z_s - v_g^{(l)}), 1 \leq g \leq c; 1 \leq s \leq N;$$

- **Paso 3:** actualizar la matriz de partición. Si  $D_{gsA} > 0$  para todo  $g = 1, 2, 3, \dots, c$ .

$$\mu_{gs}^{(l)} = \frac{1}{\sum_{h=1}^c (D_{gsA} / D_{hsA})^{2/(m-1)}}, 1 \leq g \leq c; 1 \leq s \leq N,$$

caso contrario,

$$\mu_{gs}^{(l)} = 0 \text{ y } \mu_{gs}^{(l)} \in [0, 1], \text{ con } \sum_{gs}^c \mu_{gs}^{(l)} = 1$$

hasta que  $\|U^{(l)} - U^{(l-1)}\| < \varepsilon$ .

Una forma de detectar grupos de diferentes formas geométricas en un conjunto de datos, es mediante una extensión al algoritmo FCM empleando una norma de distancia aditiva (Gustafson y Kessel, 1975), ver [10].

La implementación en R del algoritmo de agrupamiento difuso  $c$ -medias se encuentra en la librería *e1071*, función *cmeans*.

En esta sección se han presentado los conceptos principales del enfoque difuso necesarios para desarrollar la aplicación en el modelamiento de las series financieras, que es el objetivo de este trabajo. Para más información, consultar [4, 5, 10]

## **4 SISTEMA HÍBRIDO DIFUSO PARA PRONÓSTICOS DE SERIES FINANCIERAS**

Como se describió en la Introducción, el análisis del comportamiento de las series financieras para la obtención de pronósticos, es de gran importancia debido a su utilidad. Las técnicas hasta ahora desarrolladas, han sido un gran aporte para el entendimiento de los fenómenos de distinta naturaleza; sin embargo con el pasar del tiempo, estos fenómenos se han vuelto más complejos, debiéndose buscar nuevas alternativas para tratarlos. Una vez que se ha revisado parte de lo que tiene que ver con la modelación de series financieras utilizando las metodologías de descomposición y de modelos SARIMA, y su respectiva automatización (Capítulos 1 y 2); y también habiendo descrito los conceptos principales de inteligencia computacional, específicamente lo que tiene que ver con la lógica difusa en técnicas de agrupamiento de datos e inferencia (Capítulo 3), en este Capítulo se desarrolla un sistema para el pronóstico de series. El cual, básicamente está compuesto por los procesos de obtención de pronósticos automáticos y de la identificación de patrones.

La metodología propuesta constituye un primer paso para la estructuración de un sistema híbrido difuso. Estos sistemas se caracterizan por la complementación de dos o más técnicas entre sí, como pueden verse en los trabajos realizados por: Halgamuge - Wang (2005) [13], que realizan una compilación interesante de aplicaciones en diferentes campos y dedican un capítulo a los sistemas híbridos; y Palit - Popovic (2005) quienes detallan la utilización de inteligencia computacional y de sistemas híbridos para pronósticos en Control Industrial [4]. Esta metodología permite manejar múltiples escenarios tomando en cuenta varios criterios, y así disponer de mayor información en la toma de decisiones. Además, la automatización del proceso y la incorporación de

parámetros que controlan el nivel de ajuste permiten manejar el sistema dinámicamente, ver [3].

El presente Capítulo se organiza de la siguiente manera. En primer lugar, se hace una descripción de las series en las que se va a validar la metodología. Luego, se detalla el proceso de modelamiento tradicional y ciertas caracterizaciones puntuales para cada etapa. Se describe la metodología propuesta, cuyas etapas son esencialmente las partes que conforman el sistema híbrido difuso y constituye el aporte de este trabajo. Finalmente, se analiza el modelamiento tradicional y el propuesto en forma conjunta, aplicados a cada una de las series.

## **4.1 DESCRIPCIÓN DE LAS SERIES FINANCIERAS**

La metodología propuesta es probada en cuatro series financieras bancarias<sup>1</sup> correspondientes a cartera, monetarios, ahorros y plazo de una institución financiera<sup>2</sup>, las mismas que por su naturaleza presentan comportamientos distintos, y son utilizadas frecuentemente en escenarios de liquidez, de presupuestación, etc<sup>3</sup>. Las series tienen periodicidad mensual y se toma en cuenta para este análisis un histórico a partir de enero 2004 hasta febrero 2011, de los cuales, los últimos tres meses son usados para las pruebas en los pronósticos.

A continuación, se detalla el concepto de cada una de las series y su identificación en el Catálogo Único de Cuentas (CUC) de la Superintendencia de Banco y Seguros del Ecuador (SBS).

### **Cartera**

La cartera o cartera de créditos es una parte de los activos bancarios que son los recursos propios de la actividad de las instituciones financieras. El crédito es el uso de

---

<sup>1</sup>Datos obtenidos de los boletines mensuales entre los años 2004 y 2011, publicados por la Superintendencia de Bancos y Seguros del Ecuador.

<sup>2</sup>Se escogió al Banco Pichincha C.A. por ser el de mayor participación dentro del sistema financiero.

<sup>3</sup>Una aplicación en pronósticos de impuestos se la puede ver en [31]

un capital ajeno por un tiempo determinado a cambio del pago de una cantidad de dinero que se conoce como interés. De acuerdo al CUC, la cartera corresponde a la cuenta 14. Por cuestiones de análisis, no se consideran las provisiones<sup>4</sup> (cuenta 1499).

## **Monetarios**

Los depósitos monetarios o también llamadas cuentas corrientes. La cuenta corriente es un contrato (escrito) por el cual un banco se obliga a cumplir las órdenes de pago que emite el cuenta-correntista (cliente o titular de la cuenta corriente), en forma de cheques (documentos que constituyen un medio de pago) u otros mecanismos de pago y de registro, hasta el límite de la cantidad de dinero que tal cliente haya depositado en dicha cuenta, o del crédito que se haya estipulado entre las partes. Por esta condición de disponibilidad inmediata de fondos, se le considera de vencimiento incierto (también a los ahorros), lo que obliga a las instituciones financieras a realizar estimaciones con respecto a su comportamiento para la optimización de los fondos. Las cuentas consideradas como monetarios son: depósitos monetarios que generan intereses (cuenta 210105), depósitos monetarios que no generan intereses (cuenta 210110) y depósitos monetarios de instituciones financieras (cuenta 210115).

## **Ahorros**

El ahorro es aquella parte de la renta que no es gastada en los bienes de consumo y servicios. Tal reserva tiene como finalidad la previsión de eventualidades futuras o la realización de una inversión; constituye la diferencia entre ingresos y egresos. Estos valores generalmente se depositan en instituciones financieras a cambio de un interés y son exigibles mediante la presentación de libretas de ahorro u otros mecanismos de pago y registro. La banca utiliza tales depósitos para otorgar préstamos o hacer inversiones. La teoría económica supone que existe una relación entre el ahorro, la inversión y la

---

<sup>4</sup>Lo que se pretende estudiar es el comportamiento del fenómeno sin el efecto de medidas contingentes financieras para mitigar un posible riesgo crediticio.

formación de capital. La cuenta correspondiente es la 210135.

### **Plazo**

Los depósitos a plazo son aquellos que sólo pueden ser exigidos luego de un período especificado de por lo menos mayor a treinta días; ganan un interés mayor que el de las cuentas de ahorro debido a que quedan inmovilizados para el depositante durante el plazo fijado (puede ser requerido antes del vencimiento, previo acuerdo entre las partes). Para la institución financiera, constituyen obligaciones que se esperan sean requeridas en un plazo medio o largo. Este rubro se registra en la cuenta 2103.

## **4.2 MODELACIÓN TRADICIONAL DE LAS SERIES FINANCIERAS**

La búsqueda de regularidades y de patrones ha sido siempre una de las tareas básicas en Finanzas, y muchas veces se descubren simetrías que sirven de fundamento para pronosticar el comportamiento de los fenómenos, incluso antes de que se entienda la razón o causa que justifica esa regularidad. En la práctica, la mayoría de las series financieras son no estacionarias, es decir, tienen tendencia en media y muchas veces en varianza. Previamente se habrá sometido la serie a una depuración o control de calidad para garantizar la fiabilidad de los datos y la homogeneidad del intervalo en el que se toman los datos sucesivos. También es importante conocer los motivos que justifiquen cambios inusuales en el comportamiento de las series. Por lo general, estos cambios son debido a factores externos que influyen directamente en el fenómeno, como son: decisiones de negocio, entorno económico, competitividad, etc.

El paso obligatorio para analizar una serie temporal es graficar la evolución de la variable a lo largo del tiempo. El siguiente paso consiste en determinar si la secuencia de valores es completamente aleatoria o si, por el contrario, se puede encontrar algún patrón a lo largo del tiempo, pues sólo en este caso se puede seguir con el análisis. Como se detalló en el primer Capítulo, la modelación tradicional contempla dos enfoques, el

determinístico y el estocástico.

El **enfoque determinístico** se basa en descomponer la serie en varias partes: tendencia, variación estacional o periódica, y otras fluctuaciones irregulares o residuos, descritas brevemente a continuación:

1. Tendencia: es la dirección general de la variable en el periodo de observación, es decir el cambio a largo plazo de la media de la serie.
2. Estacionalidad: corresponde a fluctuaciones periódicas de la variable, en periodos relativamente cortos de tiempo.
3. Otras fluctuaciones irregulares: después de extraer de la serie la tendencia y variaciones cíclicas, queda una serie de valores residuales, que pueden ser o no totalmente aleatorios.

Para más información ver [21].

En el **enfoque estocástico** se desarrolla la teoría de los modelos SARIMA o también llamados modelos lineales de Box-Jenkins (1970), los cuales en el proceso de construcción asumen que la serie de tiempo a ser modelada es estacionaria. La metodología comprende las siguientes etapas:

1. Identificación del modelo: en la cual se tiene una idea del número de parámetros y de sus valores como resultado del cálculo previo de algunos estadísticos. El modelo tentativo es el punto de partida del análisis en el que las herramientas gráficas y el criterio del analista intervienen en su mayor parte.
2. Estimación del modelo: en la cual los valores de los parámetros del modelo tentativo seleccionado son estimados minimizando la suma de los errores cuadráticos o en base al uso de criterios de información.
3. Validación del modelo: en la cual el ajuste del modelo es revisado y se establece un posible mejoramiento al mismo. El análisis de los residuos se enfoca en demostrar

que estos son aleatorios, por lo que el análisis de sus funciones ACF y PACF es fundamental en esta etapa.

4. Pronósticos del modelo: en la cual el modelo es usado para el cálculo de los pronósticos y sus límites de confianza.

En la etapa de identificación de los modelos es necesario diferenciar la forma de identificar los parámetros<sup>5</sup> entre la modelación de series no estacionales y estacionales.

- Identificación de modelos no estacionales

- a) Seleccionando  $d$ : graficar la ACF de  $y_t$ . Si esta decae rápidamente, tomar  $d = 0$ . Si no, graficar la ACF de  $\Delta y_t$ ; si esta converge rápidamente,  $d = 1$ . Si no, revisar la ACF de  $\Delta^2 y_t$ ; si esta converge rápidamente, tomar el valor de  $d = 2$ . De otra forma, revisar la serie para ver un eventual problema antes de considerar  $d = 3$ ; que es raramente encontrado en la práctica.
- b) Seleccionando  $p$  y  $q$ : calcular y graficar las funciones ACF y PACF de  $\Delta^d y_t$ . Si la ACF se trunca después de algún retardo  $q$  pequeño ( $q \leq 3$ ), seleccionar un  $MA(q)$ . Si la PACF se trunca después de algún retardo pequeño  $p$  ( $p \leq 3$ ), seleccionar  $AR(p)$ . Si ninguna de las funciones ACF o PACF se trunca después de un pequeño valor del retardo, una mezcla de modelos puede ser considerada. En ese caso es mejor seguir buscando modelos simples; es decir, modelos con polinomios de orden inferior.

- Identificación de modelos estacionales

- a) Seleccionando  $D$  y  $d$ : graficar la ACF de  $y_t$ . Si una larga y persistente autocorrelación en los retardos 12, 24, 36 pueden verse<sup>6</sup>, diferenciar  $y_t$ :  $w_t = \Delta_{12} y_t$ . De otra manera  $D = 0$ . A continuación mirar  $d$ , el número de diferenciaciones regulares, como en los modelos no estacionales.

---

<sup>5</sup>cuya terminología y nomenclatura se definió en el Capítulo 1

<sup>6</sup>se consideran estos valores debido a que la periodicidad es anual en las series analizadas

- b) Seleccionando  $p, P, q, Q$ : la ACF y la PACF son usadas de igual forma que en los modelos no estacionales. Si ACF de  $w_t$  muestra largas autocorrelaciones en los retardos 12, esto da una idea del valor de Q, de igual manera para la función PACF. Para la parte regular del modelo ( $p$  y  $q$ ), la identificación es como se describió anteriormente.

El esquema del proceso se describe a continuación:

---

### **Modelamiento SARIMA**

---

1. Graficar serie financiera y funciones ACF y PACF.
2. Graficar retardos e identificar grado de diferenciación.
3. Seleccionar los valores de los parámetros y estimar los coeficientes.
4. Graficar los residuos y sus respectivas funciones ACF y PACF para validación.
5. Calcular pronósticos y graficarlos.
6. Analizar la descomposición de la serie.
7. Si el modelo requiere modificación, volver al paso 3.

## **4.3 MODELACIÓN EN BASE AL SISTEMA HÍBRIDO DIFUSO DE LAS SERIES FINANCIERAS**

### **4.3.1 MODELACIÓN**

El sistema híbrido difuso (SHD) que se propone está compuesto de tres etapas, en las cuales se crearon funciones que internamente utilizan algoritmos para análisis específicos y rutinas para la estructuración de datos, facilitando y optimizando el tratamiento de los mismos. Para completar el sistema se creó una función de graficación, la cual toma los resultados relevantes de cada etapa y los muestra en una gráfica apropiada.

La primera etapa comprende la generación de las listas de parámetros de los modelos probados en la búsqueda del mejor modelo, que se elige en base a la minimización

del criterio de información (CI), para la cual se utiliza la función *auto.arima2* que es una modificación de la función *auto.arima* de la librería *forecast* disponible en lenguaje R. Para poder abarcar un mayor número de modelos, se utilizan los tres criterios de información<sup>7</sup> disponibles como parámetros en la función, generándose para cada uno de ellos su respectiva lista de parámetros.

La función realiza primeramente un análisis de los parámetros a ser utilizados (específicamente los valores de  $d$  y  $D$ ), e inicia las iteraciones en busca del mejor modelo variando los parámetros  $(p, d, q, P, D, Q)$  y seleccionando los modelos según el CI elegido como discriminante. Internamente la función genera una lista en la que se registran los parámetros de todos los modelos probados, evitando que un modelo sea considerado nuevamente. Los modelos que generan indeterminación ( $CI \geq 1e20$ ) son parte de la información, a pesar de no ser seleccionados. Esta lista de parámetros no es parte de la salida de la función original, motivo por el cual se modificó el código fuente para poder acceder explícitamente a la lista. Un aspecto fundamental de esta etapa, es la exclusión de los modelos que son considerados por los expertos como modelos de pronósticos no admisibles<sup>8</sup>. Estos modelos son el resultado del análisis y del proceso reiterativo realizado con el sistema. Los modelos son archivados e incorporados en el proceso a manera de filtro, de tal forma que la lista de parámetros resultante no los contiene.

El nivel de control en el criterio de información viene dado por el valor del percentil; se quiere que el CI de los modelos sea menor o igual a este. Luego de las eliminaciones y de los filtrados, las tres listas se unifican y se procede a obtener modelos únicos, es decir, que la combinación de todos los parámetros se presente una sola vez. La lista de modelos es guardada en la lista de salida, conjuntamente con los resultados del mejor modelo obtenido para cada CI.

En la segunda etapa, la lista de modelos se utiliza para realizar los pronósticos. La función *forecast* requiere como entradas: la serie de datos, los órdenes de los parámetros

---

<sup>7</sup>Criterios de: Akaike o AIC, el de Akaike de segundo orden o AICc y Bayesiano o BIC

<sup>8</sup>Debido a que el comportamiento no va acorde con características de la serie, como son: la tendencia, estacionalidad, etc.

$(p, d, q, P, D, Q)$  del modelo, y si este tiene o no desplazamiento. La función genera una estructura de datos que contiene: el modelo, las estimaciones, los ajustes, los residuos, los pronósticos, etc.; permitiendo generar las estructuras respectivas de estos valores, de todos los modelos considerados. También los resultados para el mejor modelo según el criterio de AIC. A más de estos resultados, se incluye en la lista de salida, estructuras de ordenamiento para los valores de CI, que serán elementos importantes en la siguiente etapa.

La etapa final utiliza la matriz de pronósticos como base para el análisis y la inferencia. Gráficamente los pronósticos pueden mostrar un patrón de comportamiento, es así que el sistema híbrido difuso se basa en la identificación de dicho patrón. Para esto se analiza cada mes independientemente, con el fin de encontrar un valor (pronóstico medio) que represente a los pronósticos obtenidos para el respectivo mes. Un patrón puede ser interpretado como la similitud, o cercanía entre sí, de los elementos de un conjunto o agrupación. Para obtener estos conjuntos se utiliza el algoritmo de agrupamiento difuso o también llamado *c-medias*, el que está basado en los conceptos de la lógica difusa, permitiendo así manejar un nivel de pertenencia de los elementos a cada conjunto más flexible, lo que es ideal para temas de incertidumbre. Como se explicó en el Capítulo 3, el algoritmo requiere de ciertos parámetros para ser ejecutado, es así que se define que halle tres centros, ya que los pronósticos pueden ser clasificados en: estimación exacta, sub estimación y sobre estimación. Los conjuntos generados, y el número de elementos pertenecientes a cada uno de ellos, identifican al patrón resultante en cada mes.

La inferencia definida para la obtención del pronóstico medio, está basada en la implementación de dos criterios de ponderación para los centros de cada agrupamiento realizado en los pronósticos correspondientes a cada mes. El primer criterio considera el número de elementos en cada agrupación, y el segundo criterio se basa en el número de elementos en cada agrupación que provienen de modelos de bajo IC. A continuación se detalla la definición formal de estos dos criterios.

Sea  $1_A(x)$  la función indicadora tal que:

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

En el primer criterio, se definen las ponderaciones  $w_{mh}^1$  en las  $A_{mh}$  agrupaciones generadas en cada  $h$  mes que forma parte del horizonte de pronóstico, de la siguiente manera:

$$w_{mh}^1 = \frac{\sum_{i=1}^n 1_{A_{mh}}(p_{ih})}{n},$$

donde  $n$  es el número total de modelos que intervienen en la inferencia y  $m$  el número de centros definidos para el agrupamiento. Para los  $c_j$  centros ( $j = 1, \dots, m$ ) generados en cada  $h$  mes de pronóstico, se tiene que el pronóstico medio es:

$$\hat{p}_h^1 = \sum_{j=1}^m c_{jh} w_{jh}^1.$$

Sea  $mdl_i$  el modelo que genera el pronóstico  $p_i$  para el mes  $h$ , se define el tercio del rango de valores para el AIC de los modelos como:

$$b = \frac{\max(AIC(mdl_i)) - \min(AIC(mdl_i))}{3}.$$

Se define el conjunto  $B$  de valores bajos de los AIC como:

$$B = \{[\min(AIC(mdl_i)), \min(AIC(mdl_i)) + b]\}.$$

Con esto se puede definir la ponderación  $w_{mh}^2$  para el segundo criterio y como la correspondencia es unívoca entre el modelo  $mdl_i$  y el pronóstico  $p_{ih}$  en el mes  $h$ , se puede realizar la notación siguiente:

$$w_{mh}^2 = \frac{\sum_{i=1}^n 1_{B_{mh}}(p_{ih})}{n}.$$

Para los  $c_j$  centros ( $j = 1, \dots, m$ ) generados en los  $h$  meses del horizonte de pronóstico se tiene que el pronóstico medio según el segundo criterio se define como:

$$\hat{p}_h^2 = \sum_{j=1}^m c_{jh} w_{jh}^2.$$

El pronóstico medio resultante para cada  $h$  mes está dado por:

$$\hat{p}_h = \frac{\hat{p}_h^1 + \hat{p}_h^2}{2}.$$

Con las listas de resultados de cada etapa, se procede a realizar las gráficas donde claramente se visualizan los resultados<sup>9</sup> tanto de: ajustes, residuos, pronósticos, los pronósticos medios y su respectiva comparación con el mejor modelo SARIMA.

Como parte de la validación de los resultados, se realiza la prueba de pronósticos con datos reales de tres meses, donde los resultados obtenidos en las series analizadas, validan la hipótesis planteada para este trabajo.

#### 4.3.2 IMPLEMENTACIÓN

El aporte de este proyecto es la implementación de los conceptos e ideas revisadas anteriormente, en una herramienta versátil y funcional, de tal forma que el sistema híbrido difuso desarrollado pueda ser administrado eficientemente; y con esto lograr, que sea utilizado para la obtención de resultados necesarios e importantes en diferentes ámbitos del respectivo negocio.

Para la obtención de pronósticos mediante el sistema híbrido difuso<sup>10</sup>, se crearon las funciones que se detallan a continuación; mencionando su definición, las entradas requeridas, la lista de salida o de resultados y el pseudo código. El código fuente de las funciones se encuentra en el Anexo 1.

---

**shd\_md1:** genera lista de objetos resultantes de la modelación automática: lista de modelos y mejores modelos en base a los criterios de información AIC, BIC y AICc.

---

- Entradas:
  - dat: datos o serie financiera.
  - mdlexc: modelos excluidos.
  - quaprb: probabilidad para seleccionar los modelos con  $CI \leq$  percentil correspondiente.
  - trace: imprimir en consola los modelos probados. Por defecto FALSE.
  - fil: nombre del archivo para guardar resultados (por defecto rstsel.Rdata).

---

<sup>9</sup>Un análisis más detallado del comportamiento de los residuos en el ajuste se puede revisar en [31].

<sup>10</sup>Los scripts del sistema pueden ser revisados gratuitamente en [32].

- Salidas (lstmmdl):
  - mtzmdl: lista de modelos.
  - mdl\_aic: mejor modelo en base a AIC.
  - mdl\_bic: mejor modelo en base a BIC.
  - mdl\_aicc: mejor modelo en base a AICc.
- Algoritmo 1:
  1. Generar listas de modelos en base a los criterios de información AIC, BIC y AICc, utilizando la función *auto.arima2* (función modificada de la librería *forecast*).
  2. Cargar datos de cada resultado y filtrar los modelos que generaron indeterminación en el CI.
  3. Seleccionar los modelos cuyo CI es menor o igual al valor correspondiente al percentil ingresado.
  4. Excluir modelos enlistados.
  5. Combinar las tres listas de resultados en una sola.
  6. Seleccionar los modelos únicos.
  7. Guardar los resultados.
  8. Crear lista de salida.

---

**shd\_prn:** genera lista con matrices de ajustes, residuos y pronósticos. A más de esto, guarda los resultados del mejor modelo obtenido en base al AIC.

---

- Entradas:
  - lstmmdl: salida de *shd\_mdl* (lista de modelación).
  - hrzprn: horizonte de pronóstico (h en la función *forecast*).
  - nivitv: nivel de confianza de los intervalos en la función *forecast*.
  - fil: nombre del archivo para guardar resultados. Por defecto (*rstprn.Rdata*).
- Salidas (lstpnr):
  - mtzaju: matriz de ajustes.
  - mtzrsd: matriz de residuos.
  - mtzrel: matriz de residuos relativos.
  - mtzprn: matriz de pronósticos.
  - mtzaic: matriz de valores de AIC.
  - mtzajumej: matriz de ajuste del mejor modelo.
  - mtzrsdmej: matriz de residuos del mejor modelo.
  - mtzrelmej: matriz de residuos relativos del mejor modelo.
  - mtzprnmej: matriz de pronósticos del mejor modelo.
  - dat: datos o serie financiera.
  - mtzmdl: lista de modelos.
- Algoritmo 2:
  1. Extraer lista de modelos de la lista de modelación.

2. Obtener las dimensiones de las matrices y generar listas y matrices nulas.
3. Generar las estimaciones para cada modelo.
4. Generar los pronósticos para cada modelo.
5. Crear las matrices de ajuste, residuos, residuos relativos, pronósticos y de criterio de información.
6. Generar las salidas para el mejor modelo.
7. Guardar los resultados de los pronósticos.
8. Crear lista de salida.

---

**shd\_cdp:** genera agrupamientos difusos de los pronósticos.

---

- Entradas:
  - `lstprn`: salida de la función `shd_prn` (lista de ajustes, residuos y pronósticos).
  - `fil`: nombre del archivo para guardar resultados. Por defecto (`rstcdp.Rdata`).
- Salidas (`lstcdp`):
  - `lstclu`: lista de resultados del algoritmo *c-medias*.
  - `mtzclu`: matriz de agrupamiento de los pronósticos.
  - `mtzcen`: matriz de centros.
  - `mtzele`: matriz de número de elementos por agrupamiento.
  - `mtzifr`: matriz resultante de la inferencia.
- Algoritmo 3:
  1. Extraer la matriz de pronósticos y sus dimensiones.
  2. Generar listas y matrices nulas.
  3. Crear los agrupamientos difusos para cada mes de pronóstico, previa validación de la dispersión.
  4. Crear las matrices de agrupamientos, centros y número de elementos de cada agrupación.
  5. Calcular la regla de inferencia en base a los dos criterios de ponderación.
  6. Crear lista de salida.
  7. Guardar los resultados.

---

**shd\_grf:** genera gráficas.

---

- Entradas:
  - `lstprn`: salida de `shd_prn`.
  - `lstcdp`: salida de `shd_cdp`.
  - `pre`: prefijo para nombre de gráficas.
- Salidas:
  - `_dat.pdf`: gráfica de datos y de las funciones ACF y PACF.
  - `_lag.pdf`: gráfica de los retardos de la serie.

- `_prn.pdf`: gráfica de los pronósticos del mejor modelo.
- `_rsd.pdf`: gráfica de los residuos del mejor modelo.
- `_dcpstl.pdf`: gráfica de descomposición stl (método *loess*).
- `_dcpmul.pdf`: gráfica de descomposición multiplicativa.
- `_dcpadd.pdf`: gráfica de descomposición aditiva.
- `_dcpmul.s.pdf`: gráfica del componente estacional multiplicativo.
- `_dcpadd.s.pdf`: gráfica del componente estacional aditivo.
- `_dcpadd_sr.pdf`: gráfica del componente estacional y del residuo (aditivos).
- `_mtzaju.pdf`: gráfica de la matriz de ajustes.
- `_mtzrsd.pdf`: gráfica de la matriz de residuos.
- `_mtzifr.pdf`: gráfica de la matriz de pronósticos e inferencia.
- `_mtztst.pdf`: gráfica de la matriz de inferencia con pruebas de pronósticos e identificadores de los modelos.
- `_mtzdat.pdf`: gráfica de la matriz de datos conjuntos (reales, pronósticos e inferencia).
- `_boxrsd.pdf`: gráfica en *boxplot* de los residuos.
- `_boxrel.pdf`: gráfica en *boxplot* de los residuos relativos.
- `_disrsd.pdf`: gráfica de dispersión de los residuos y residuos relativos por modelo.

■ Algoritmo 4:

1. Extraer las matrices a ser graficadas.
2. Calcular dimensiones de las estructuras requeridas en las gráficas.
3. Graficar matrices de datos.
4. Guardar en formato pdf las gráficas.

El esquema del proceso general para la obtención de los pronósticos mediante el sistema híbrido difuso que utiliza las funciones mencionadas es el siguiente:

---

**proceso general**

---

1. Ejecutar `shd_mdl`: etapa de selección de modelos. Algoritmo 1.
2. Ejecutar `shd_prn`: etapa de cálculo de: ajustes, residuos, residuos relativos y pronósticos. Algoritmo 2.
3. Ejecutar `shd_cdp`: etapa de cálculo de agrupamientos difusos e inferencias. Algoritmo 3.
4. Ejecutar `shd_grf`: generación de gráficas. Algoritmo 4.
5. Analizar gráficas e identificar modelos que deben ser excluidos por generar pronósticos no admisibles.

## 4.4 COMPARACIÓN DE METODOLOGÍAS Y RESULTADOS

### 4.4.1 SERIE CARTERA

#### 4.4.1.1. Modelación Tradicional Cartera

La modelación inicia con las gráficas de la evolución de la serie y de sus funciones ACF y PACF. La Figura 4.4.1.1 muestra que la serie presenta tres características importantes:

- Tiene una tendencia creciente regular en la mayor parte de los siete años de tiempo observado.
- Presenta un incremento de la tendencia a mediados del 2008 y a finales del 2010, el primero debido a estrategias de negocio y el segundo por la compra de cartera de Lloyds Bank (aproximadamente US\$100 millones).
- Muestra un decaimiento en los primeros seis meses del año 2009, debido a la incertidumbre que se tuvo con respecto a la dolarización, en la que los bancos comenzaron a restringir los créditos y a priorizar la liquidez.

En la Figura 4.4.1.1, la función ACF sugiere que la serie debe ser diferenciada estacionariamente; esto es ratificado por la gráfica de los retardos (Figura 4.4.1.2), donde la correlación es la más alta en la diferenciación de primer orden, y por el resultado de la función *ndiffs* de la librería *forecast* que realiza una prueba de raíz unitaria ( $ndiffs(datcar) = 1$ ).

Se presenta la gráfica de la serie diferenciada de orden 1 y sus respectivas funciones ACF y PACF (Figura 4.4.1.3), en base al análisis de estas funciones, se plantea el modelo inicial y se realiza la búsqueda del modelo que minimice el CI<sup>11</sup>.

Al terminar el proceso de búsqueda se obtiene el modelo con menor AIC siguiente:

---

<sup>11</sup>para este trabajo se eligió al AIC como discriminante.

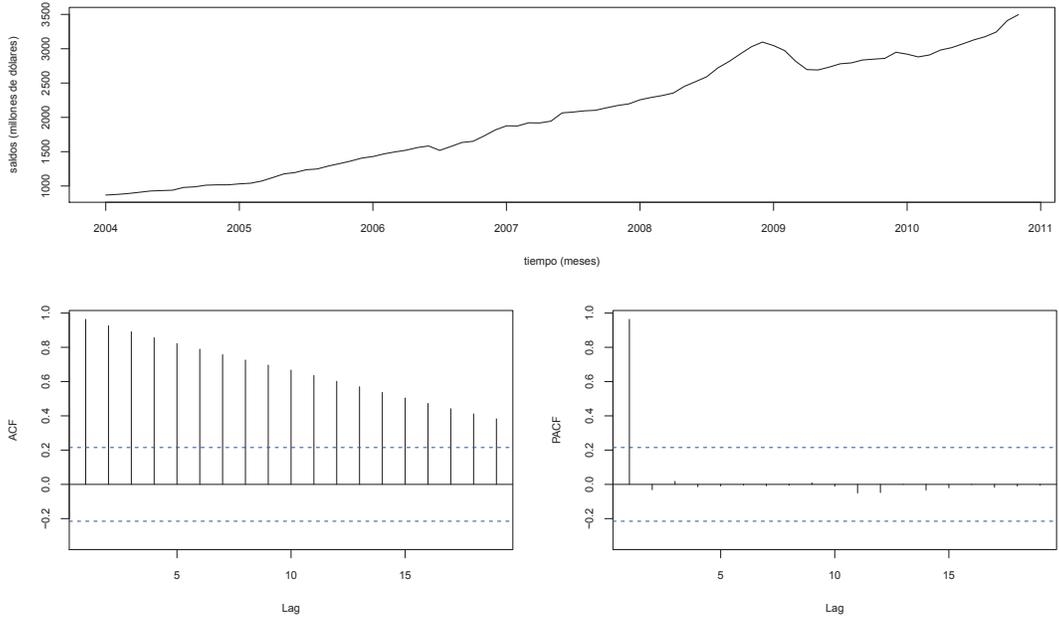


Figura 4.4.1.1: serie de cartera y funciones ACF y PACF

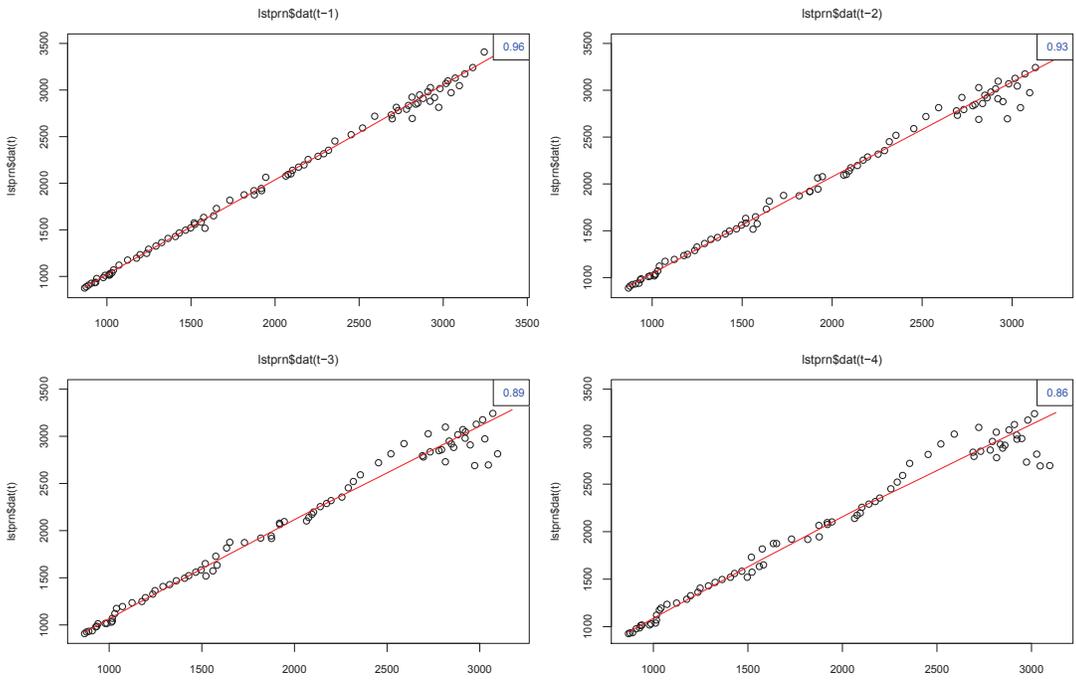


Figura 4.4.1.2: retardos del orden 1 al 4

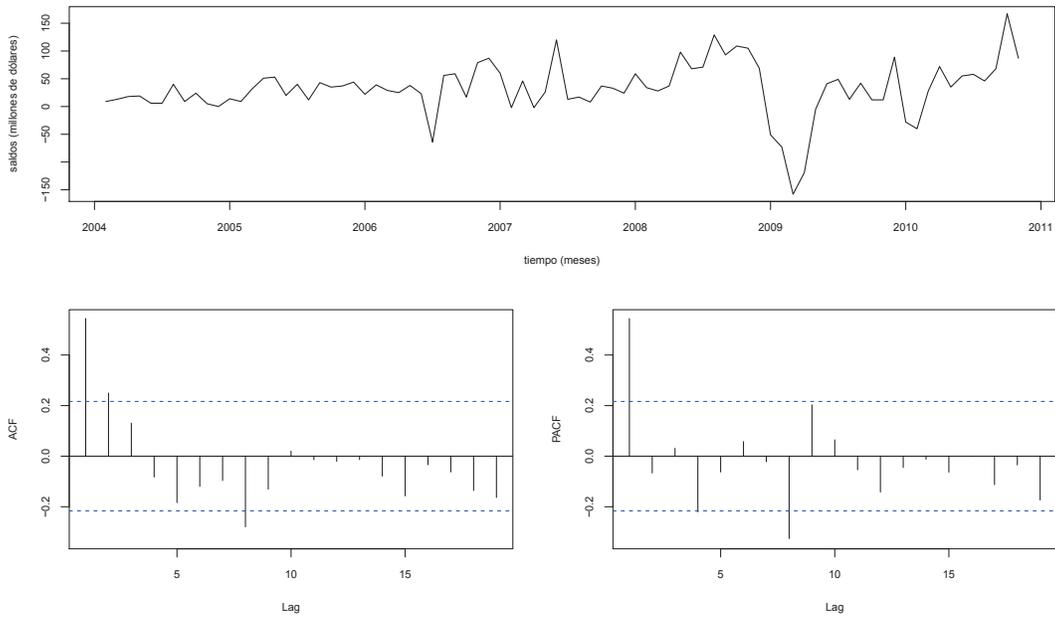


Figura 4.4.1.3: serie cartera diferenciada de orden 1

```

> mdlcar
Series: datcar
ARIMA(1,1,0) with drift
Call: arima(x = datcar)
Coefficients:
      ar1    drift
 0.5530 33.1286
s.e. 0.0909 9.6181
sigma^2 estimated as 1579: log likelihood = -423.59
AIC = 853.18  AICc = 853.48  BIC = 860.43

```

Para la validación del modelo se analizan las gráficas de las funciones ACF y PACF de los residuos (Figura 4.4.1.4), que no muestran correlación en los retardos de menor grado, por lo que el modelo es aceptado. El octavo retardo excede el límite de confianza, se analizó la inclusión de este parámetro, pero no se obtuvo una mejora significativa, por

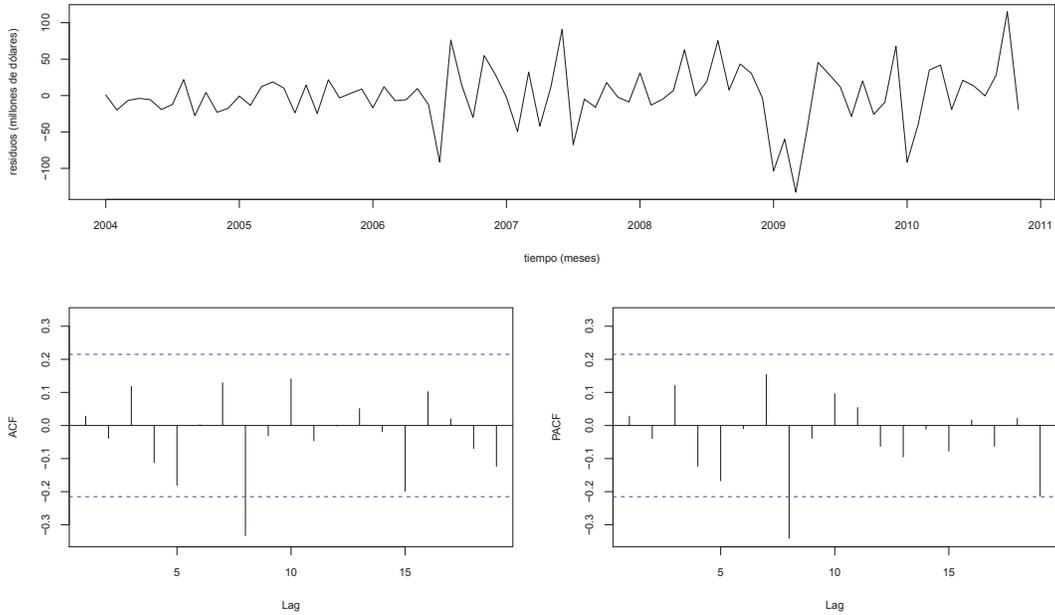


Figura 4.4.1.4: residuos y funciones ACF y PACF (cartera)

lo que no se tomó en cuenta en el modelo.

El modelo obtenido es usado para generar los pronósticos, los cuales mantienen la característica creciente lineal que es predominante en la serie (Figura 4.4.1.5).

Debido a que la serie presenta un alto grado de linealidad, no es necesario el análisis de descomposición. Por practicidad, la validación de los pronósticos se la realiza conjuntamente con el sistema híbrido difuso.

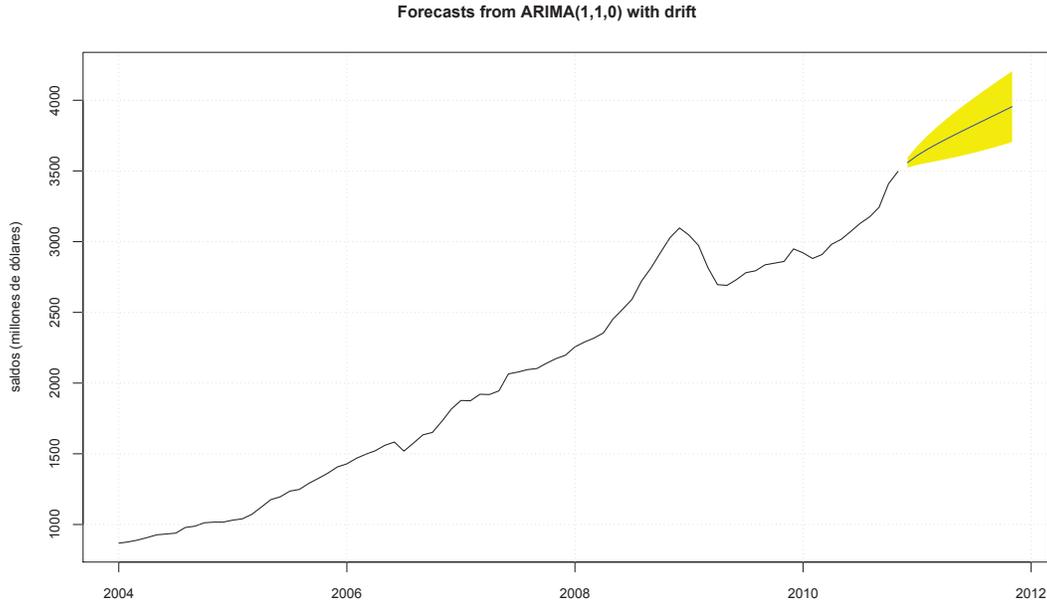


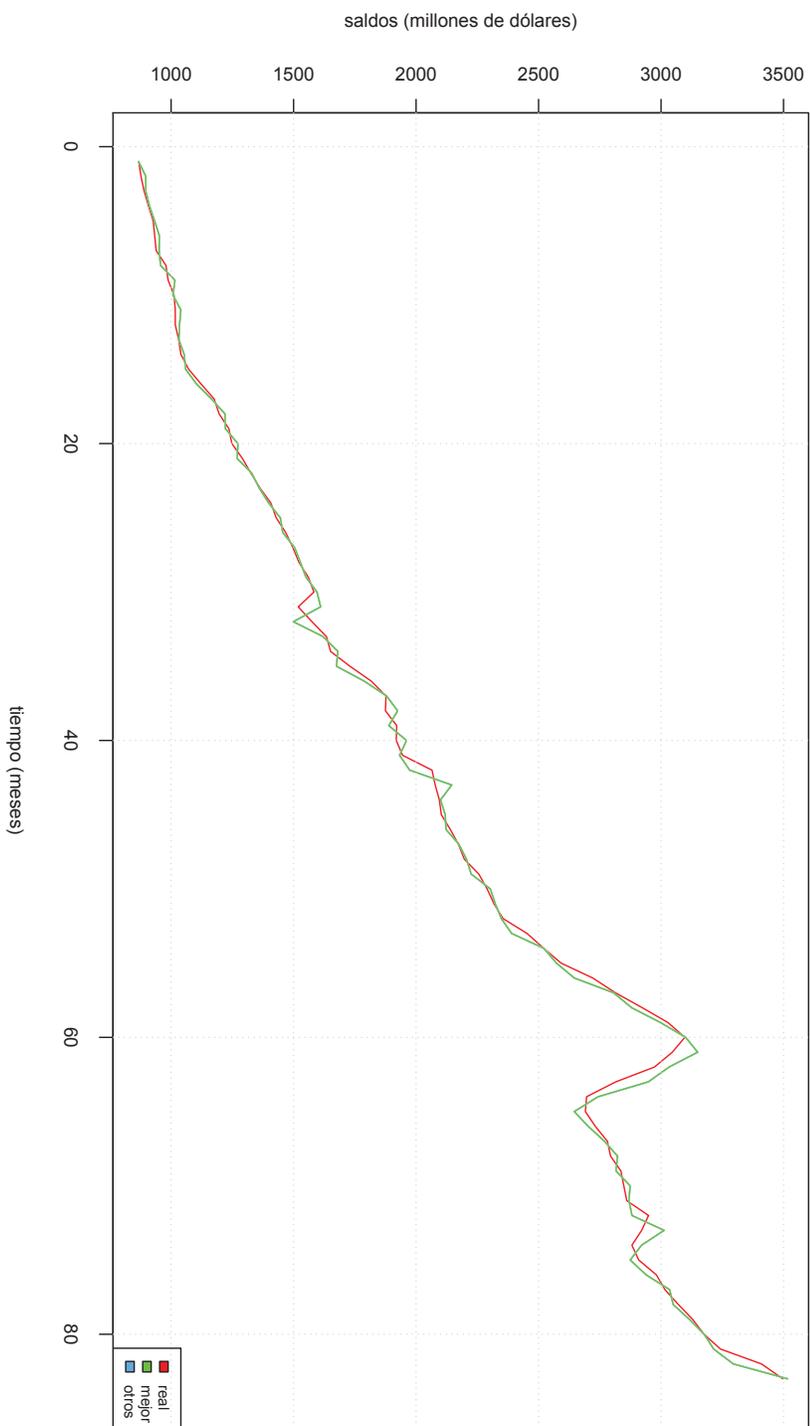
Figura 4.4.1.5: pronósticos de la serie de cartera

#### 4.4.1.2. Modelación Sistema Híbrido Difuso Cartera

Al iniciar la ejecución del sistema, la primera etapa genera once modelos, de los cuales tres resultan ser los más representativos al final del análisis. La Figura 4.4.1.6 muestra los ajustes de los modelos, notándose la similitud con el comportamiento de la serie original. Los modelos capturan los cambios en la tendencia debido a su componente autoregresivo.

La Figura 4.4.1.7 de los residuos (en millones de dólares), muestra que las desviaciones llegan a los  $\pm 20$  millones los tres primeros años, los siguientes años en promedio llegan a los  $\pm 50$  millones, exceptuando al valor de los primeros meses del año 2009 que llega a un valor atípico aproximado de -140 millones. Las gráfica de los residuos relativos muestra los mismos conceptos pero a manera de porcentaje del residuo respecto al valor real en el correspondiente mes. El porcentaje fluctúa entre  $\pm 3\%$  aproximadamente en los tres modelos y llegando a un nivel  $4\%$  y de  $-6\%$  en los atípicos.

Figura 4.4.1.6: ajustes de los modelos (cartera)



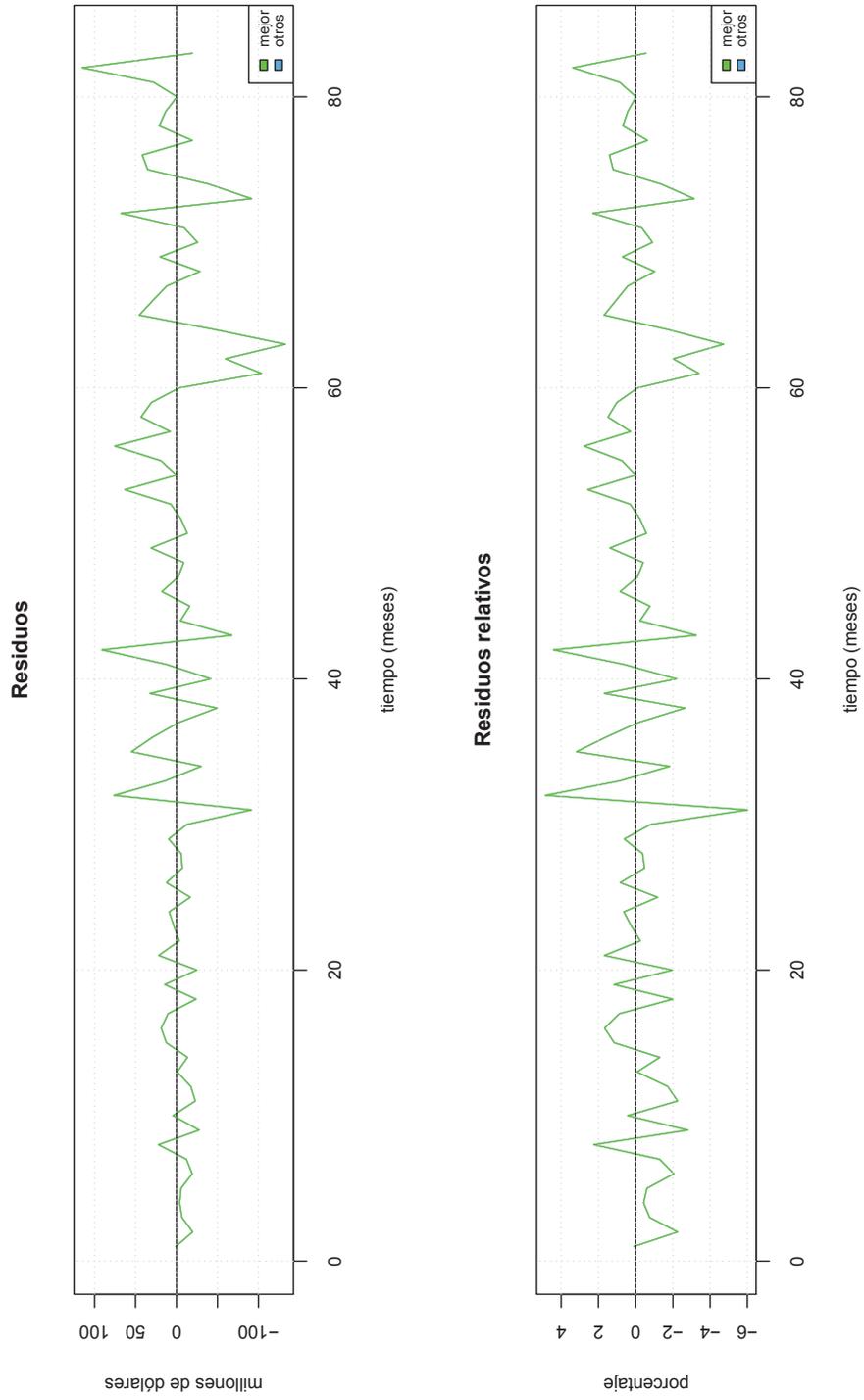


Figura 4.4.1.7: residuos de los modelos (cartera)

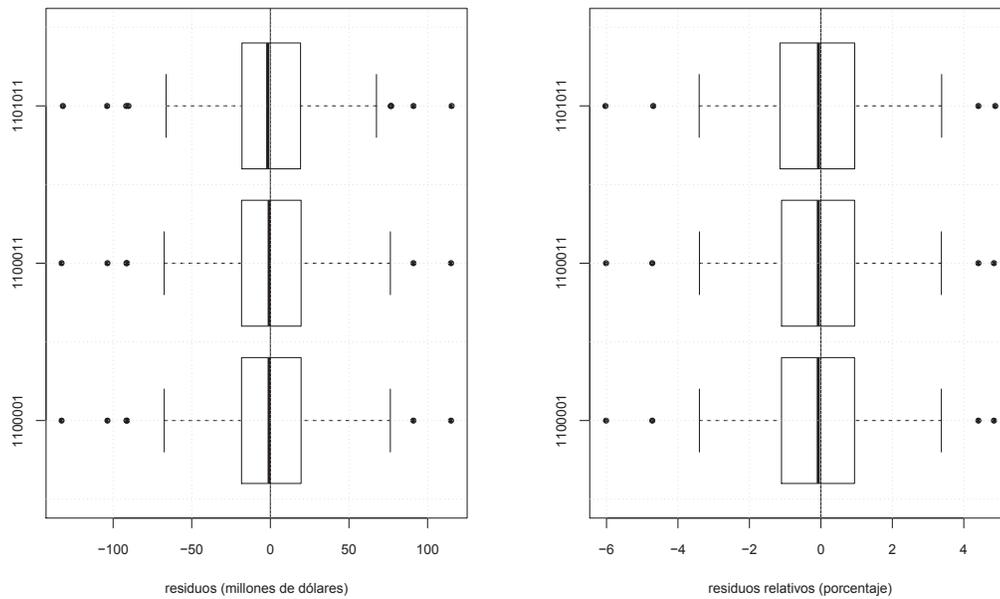


Figura 4.4.1.8: dispersión de residuos y residuos relativos por modelo (cartera)

La Figura 4.4.1.8 muestra claramente la dispersión de los residuos por modelo, los que promedian el valor cero. La presencia de datos atípicos, significa que ninguno de los tres modelos puede recoger los cambios bruscos en tendencia.

En la Figura 4.4.1.9 superior se muestran los pronósticos para un horizonte de tiempo de doce meses. La tendencia creciente se mantiene. Los tres modelos tienen forma similar, presentándose una pequeña variación a partir del cuarto mes, llegando a ser esta diferencia de dos millones al final del horizonte. En los pronósticos generados por los tres modelos se utiliza la técnica de agrupamiento, pero debido a la similitud en su comportamiento, en los ocho primeros meses el algoritmo no se ejecuta, mostrando como dato representativo la media de los pronósticos. A partir del noveno mes, el algoritmo se ejecuta y es posible diferenciar los grupos. La inferencia es aplicada en los pronósticos y se obtiene el pronóstico medio, que se representa en la Figura 4.4.1.9 inferior.

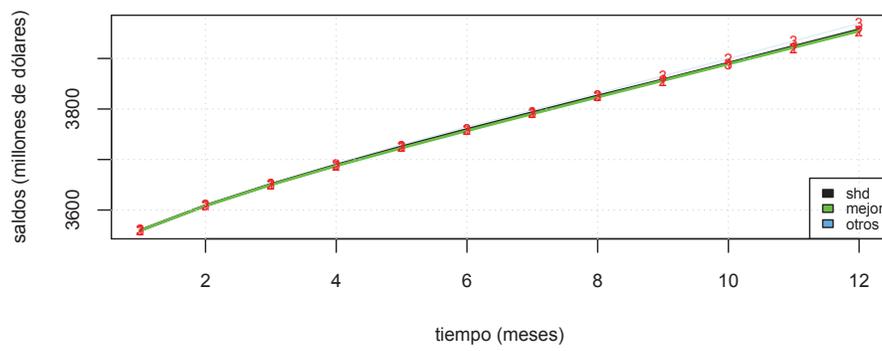
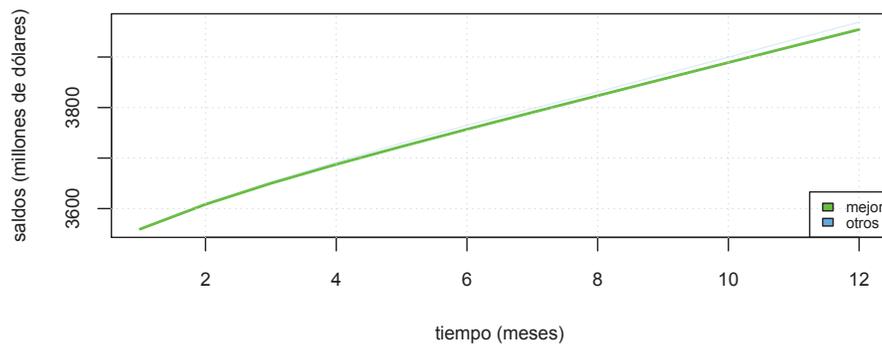


Figura 4.4.1.9: inferencia en los pronósticos (cartera)

La Figura 4.4.1.10 muestra la comparación de los pronósticos del SHD y del mejor modelo SARIMA con respecto a tres meses reales. También se definió una zona de confianza <sup>12</sup> del  $\pm 3\% \sqrt{h}$ . Como se puede visualizar, el modelo SHD tiene más cercanía con los datos reales. La Tabla 4.4.1.1 muestra el cálculo de comparación de ajuste entre los modelos para la prueba de pronósticos.

Como medidas de ajuste se consideran las siguientes:

- $RMSE = (\sum_i^h e_i^2/n)^{1/2}$ , raíz del error medio cuadrático (*root mean square error*).
- $MAE = \sum_i^h |e_i|/n$ , media absoluta del error (*mean absolute error*).
- $MAPE = \sum_i^h |e_i/x_i|/n$ , media del error absoluto porcentual (*mean absolute percentage error*).

Las comparaciones se las presentan en la Tabla 4.4.1.1, muestran que el valor de las tres medidas de ajuste para las pruebas de los pronósticos. Realizando la comparación, los valores de las medidas en el sistema SHD son menores que en el modelo SARIMA.

CARTERA				RMSE		MAE		MAPE	
mes	real	sarima	shd	sarima.e2	shd.e2	sarima.abs	shd.abs	sarima.relabs	shd.relabs
2010-12	3577	3559	3560	324	289	18	17	0.50	0.48
2011-01	3608	3608	3609	0	1	0	1	0.00	0.03
2011-02	3671	3650	3651	441	400	21	20	0.57	0.54
<b>ESTADÍSTICOS</b>				<b>15.97</b>	<b>15.17</b>	<b>13.00</b>	<b>12.67</b>	<b>0.36</b>	<b>0.35</b>

Tabla 4.4.1.1: resultados de pruebas en los pronósticos (cartera)

Para culminar con la modelación, se presenta una gráfica que muestra: los datos reales, los ajustes y los resultados de los pronósticos en forma conjunta (Figura 4.4.1.11).

<sup>12</sup>el factor  $\sqrt{h}$  es considerado como la incertidumbre en el tiempo, también llamada “la regla de la raíz cuadrada del tiempo” (“square root of time rule”) propuesta por RiskMetrics en los pronósticos de volatilidades, ver [29]

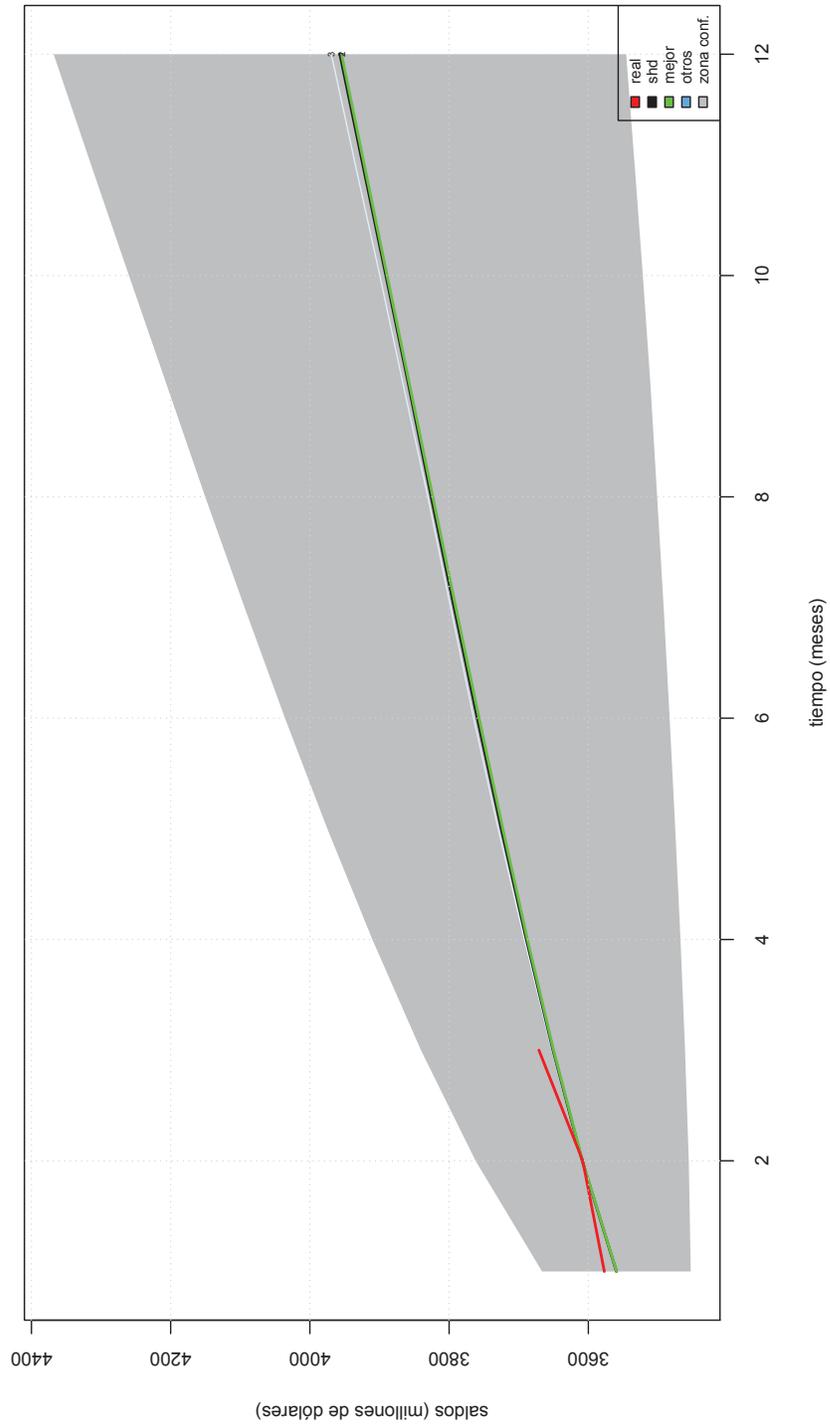


Figura 4.4.1.10: pruebas de pronósticos con datos reales (cartera)

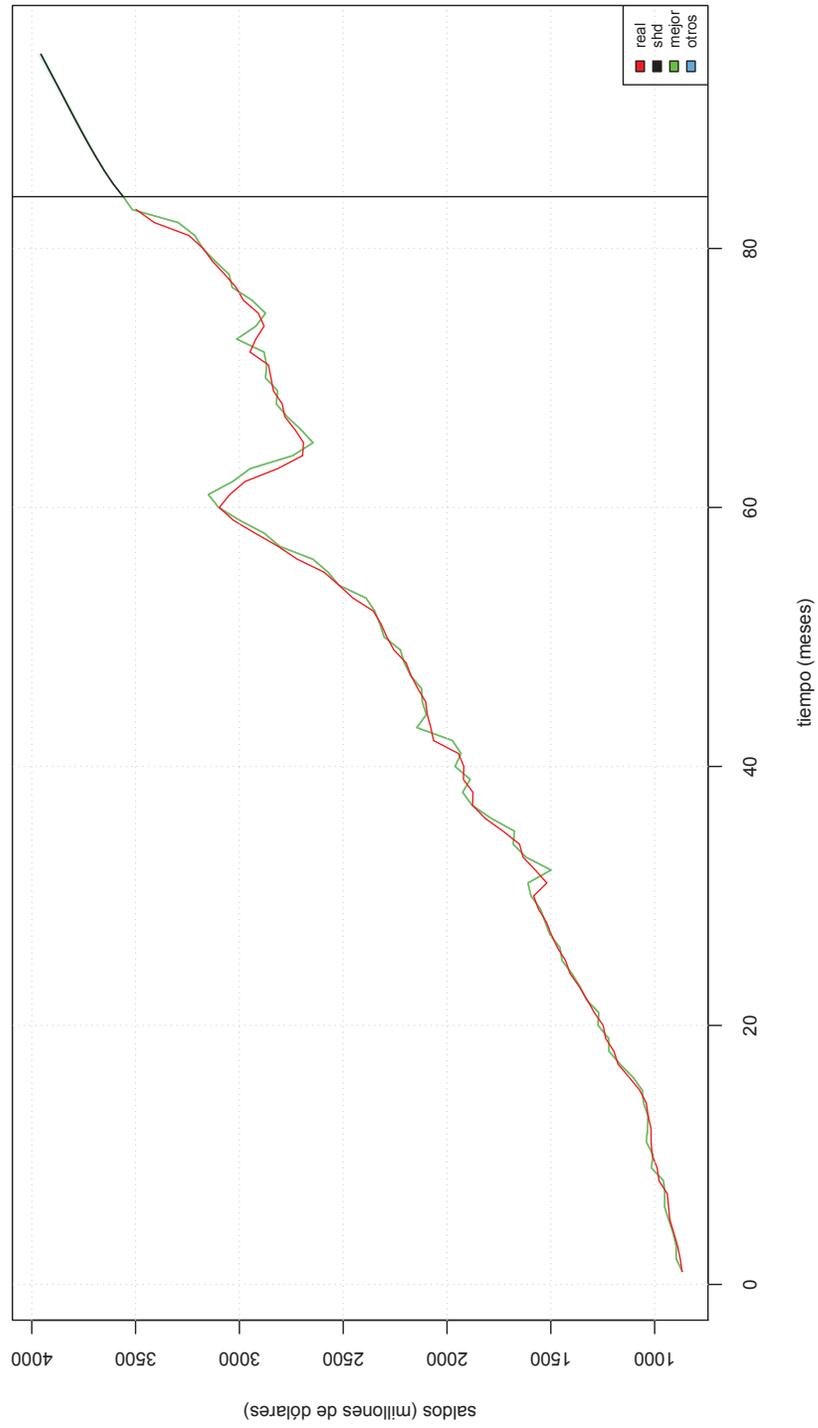


Figura 4.4.1.11: visualización conjunta de resultados (cartera)

## 4.4.2 SERIE MONETARIOS

### 4.4.2.1. Modelación Tradicional Monetarios

Al inicio se muestran las gráficas de la evolución de la serie y de sus funciones ACF y PACF. En la Figura 4.4.2.1 se pueden notar los siguientes aspectos relevantes de la serie:

- Tiene una tendencia creciente en la mayoría de años, con excepción del año 2009 que presenta un comportamiento constante e incluso decreciente en los primeros meses. Como se mencionó en la modelación anterior, esto se debe a la incertidumbre acerca de política económica presente en esa época.
- Esta serie tiene como característica la presencia de pequeñas variaciones periódicas en la tendencia que pueden ser causadas por la existencia de estacionalidad, la misma que es analizada posteriormente en base a la descomposición de la serie.
- En el mes de octubre de 2004 se presenta una caída brusca en la tendencia debido a una corrida de depósitos que sufrió la institución por rumores de iliquidez.

En la Figura 4.4.2.1, la función ACF sugiere que la serie debe ser diferenciada estacionariamente, esto lo corroboran: la Figura 4.4.2.2 de los retardos, que muestran la correlación más alta en la diferenciación de primer orden; y también, el resultado de la función  $ndiffs(ndiffs(datmon) = 1)$ .

Se presenta la gráfica de la serie diferenciada de orden 1 y sus respectivas funciones ACF y PACF (Figura 4.4.2.3), en base al análisis de estas funciones se plantea el modelo inicial y se realiza la búsqueda del modelo que minimice el AIC. Como se anotó al inicio del análisis, la serie presenta estacionalidad por lo que se incluyen parámetros estacionales.

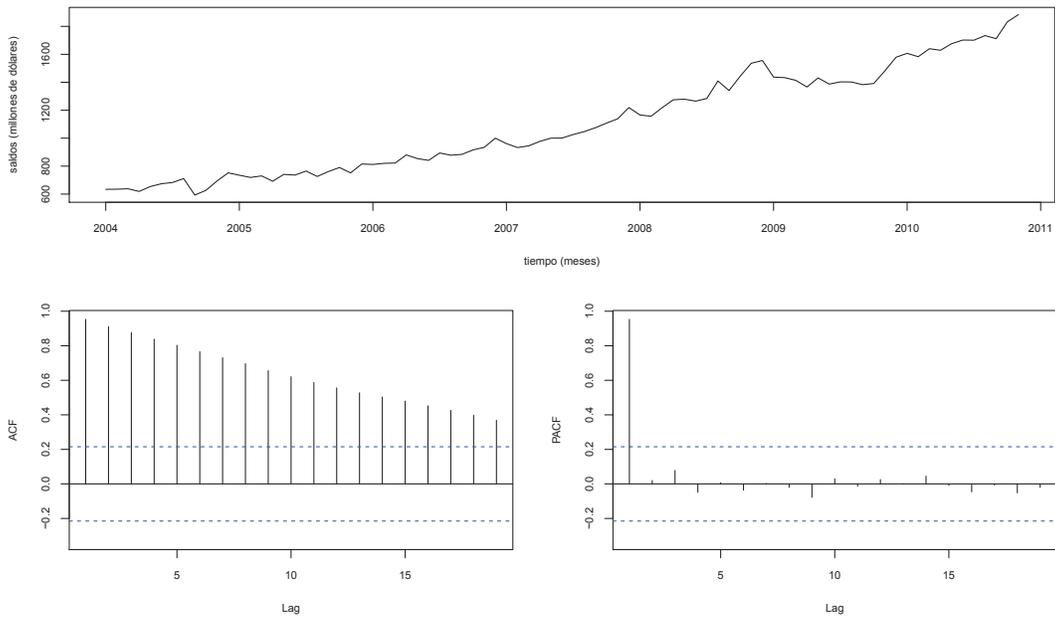


Figura 4.4.2.1: serie de monetarios y funciones ACF y PACF

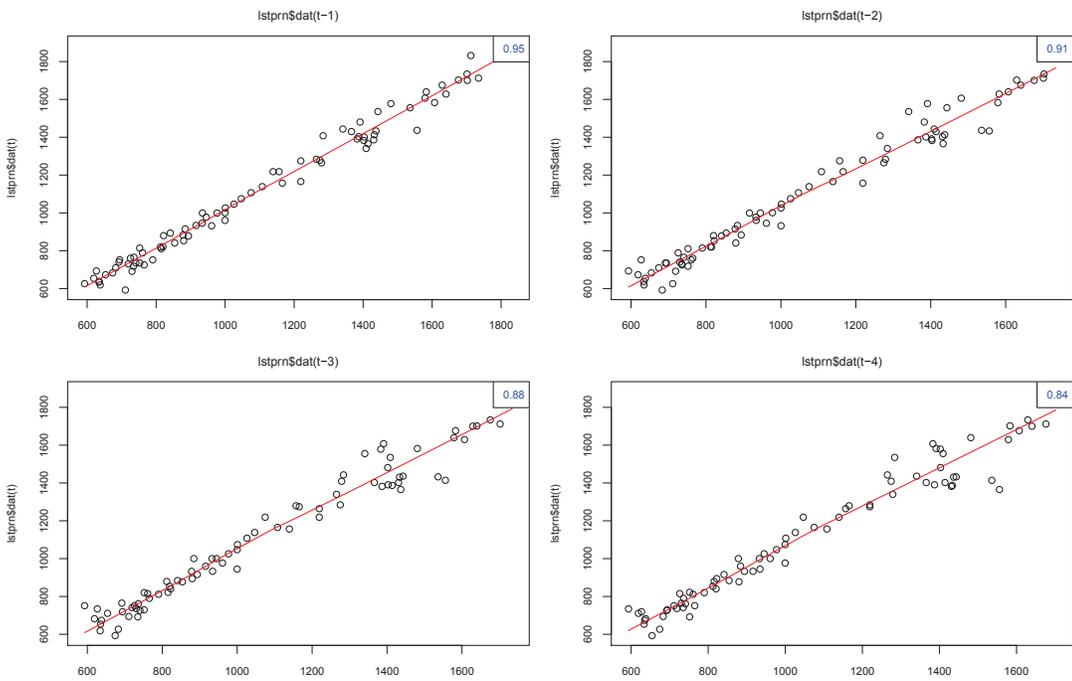


Figura 4.4.2.2: retardos del orden 1 al 4

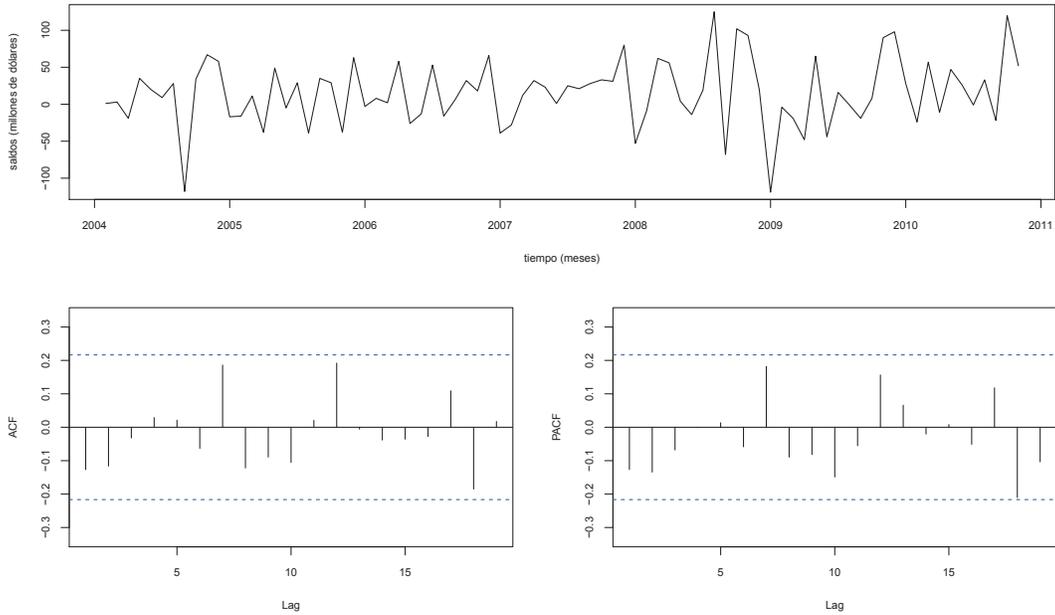


Figura 4.4.2.3: serie monetarios diferenciada de orden 1

Al terminar el proceso de búsqueda se obtiene el siguiente modelo:

```
> mdlmon
Series: datmon
ARIMA(1,1,0)(2,0,0)[12] with drift
Call: arima(x = datmon)
Coefficients:
          ar1      sar1      sar2      drift
      -0.1669  0.1407  0.3500  14.3769
s.e.    0.1091  0.1081  0.1263   6.2746
sigma^2 estimated as 1674:  log likelihood = -422.57
AIC = 855.15   AICc = 855.94   BIC = 867.18
```

Para la validación del modelo se analizan las gráficas de las funciones ACF y PACF de los residuos (Figura 4.4.2.4), los que no muestran correlación en los retardos de menor

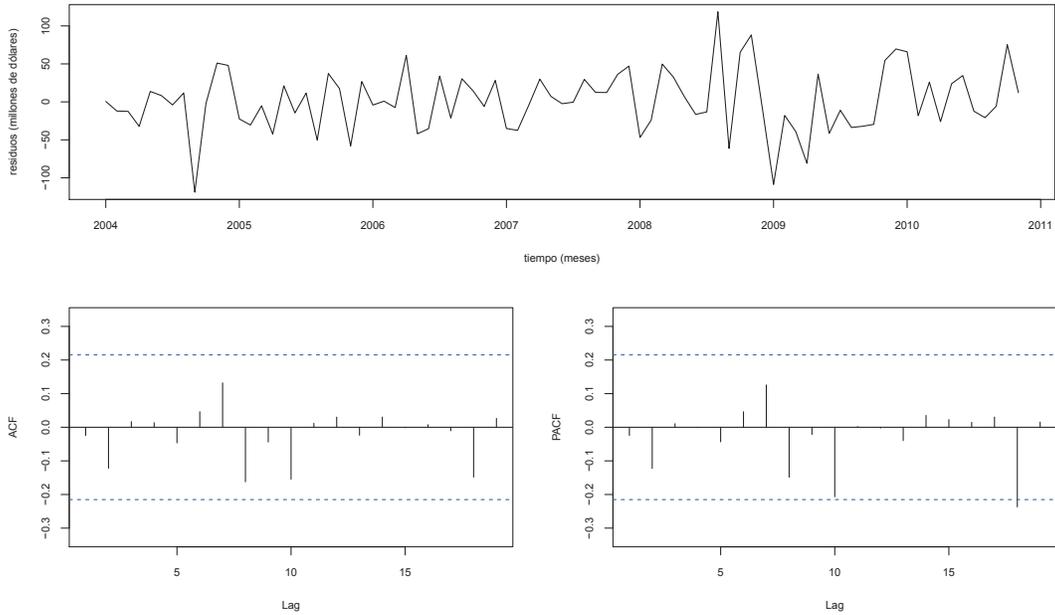


Figura 4.4.2.4: residuos y funciones ACF y PACF (monetarios)

grado, por lo que el modelo es aceptado. Los valores de los residuos más grandes se dan en los dos eventos señalados al inicio del análisis.

El modelo obtenido es usado para generar los pronósticos, los cuales mantienen la característica creciente y el comportamiento estacional (Figura 4.4.2.5).

En este análisis se incluye la modelación de la serie por descomposición para identificar su comportamiento estacional. La Figura 4.4.2.6 presenta los componentes de: tendencia, estacionalidad y residuos, al analizar estas gráficas, los residuos no tienen un comportamiento aleatorio, esto es un indicio de que la estacionalidad es débil. El componente estacional está marcado por el incremento en el mes de diciembre que se caracteriza por altos niveles de consumo, por ende, el incremento en las cuentas: corporativas, empresariales y/o de negocios; que son los sectores que utilizan en su mayor parte las cuentas monetarias.

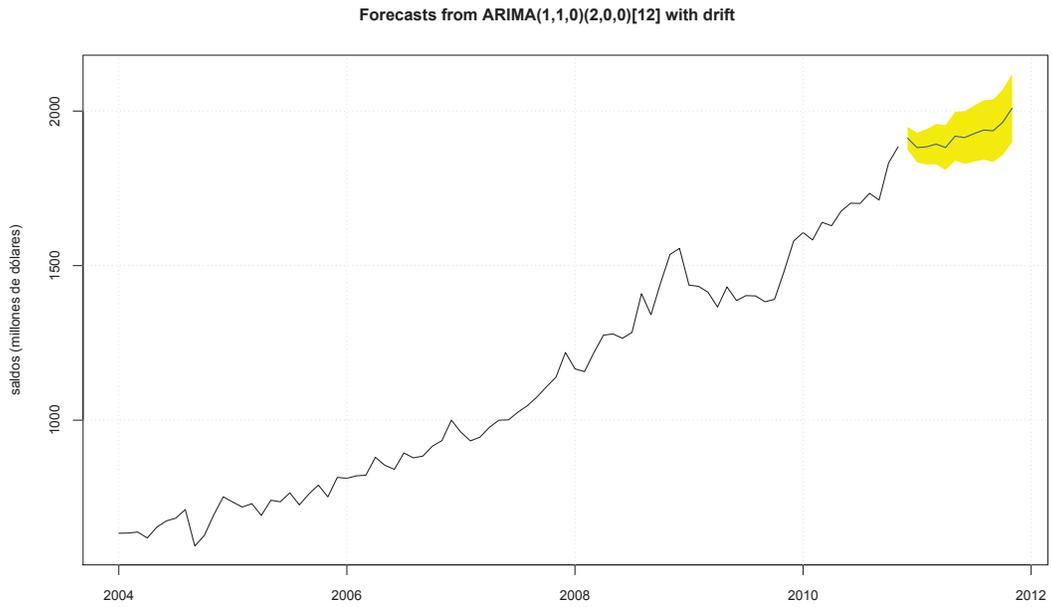


Figura 4.4.2.5: pronósticos de la serie de monetarios

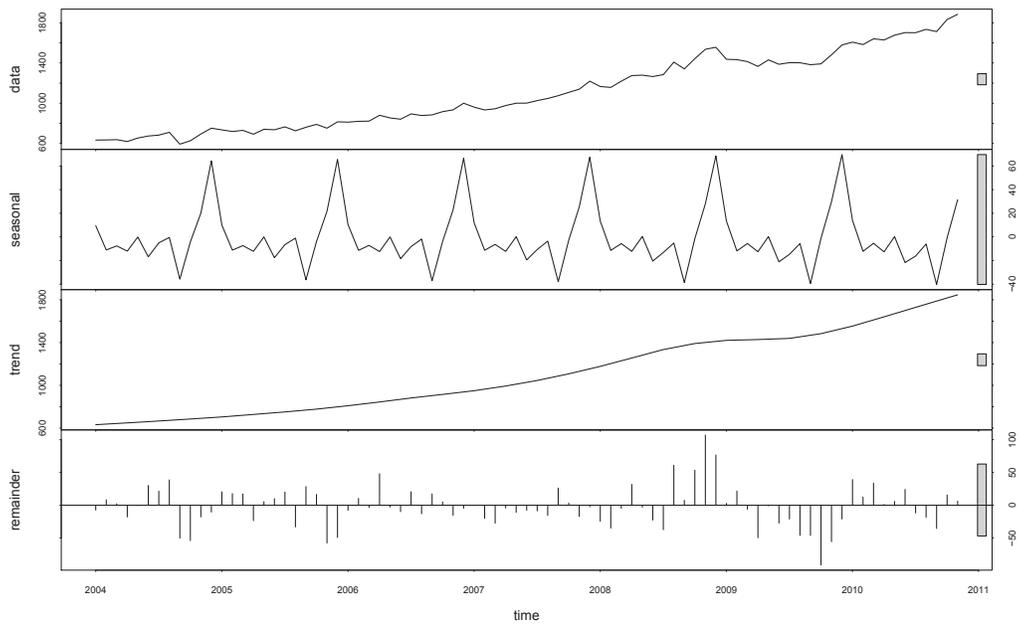


Figura 4.4.2.6: descomposición de la serie monetarios

#### 4.4.2.2. Modelación Sistema Híbrido Difuso Monetarios

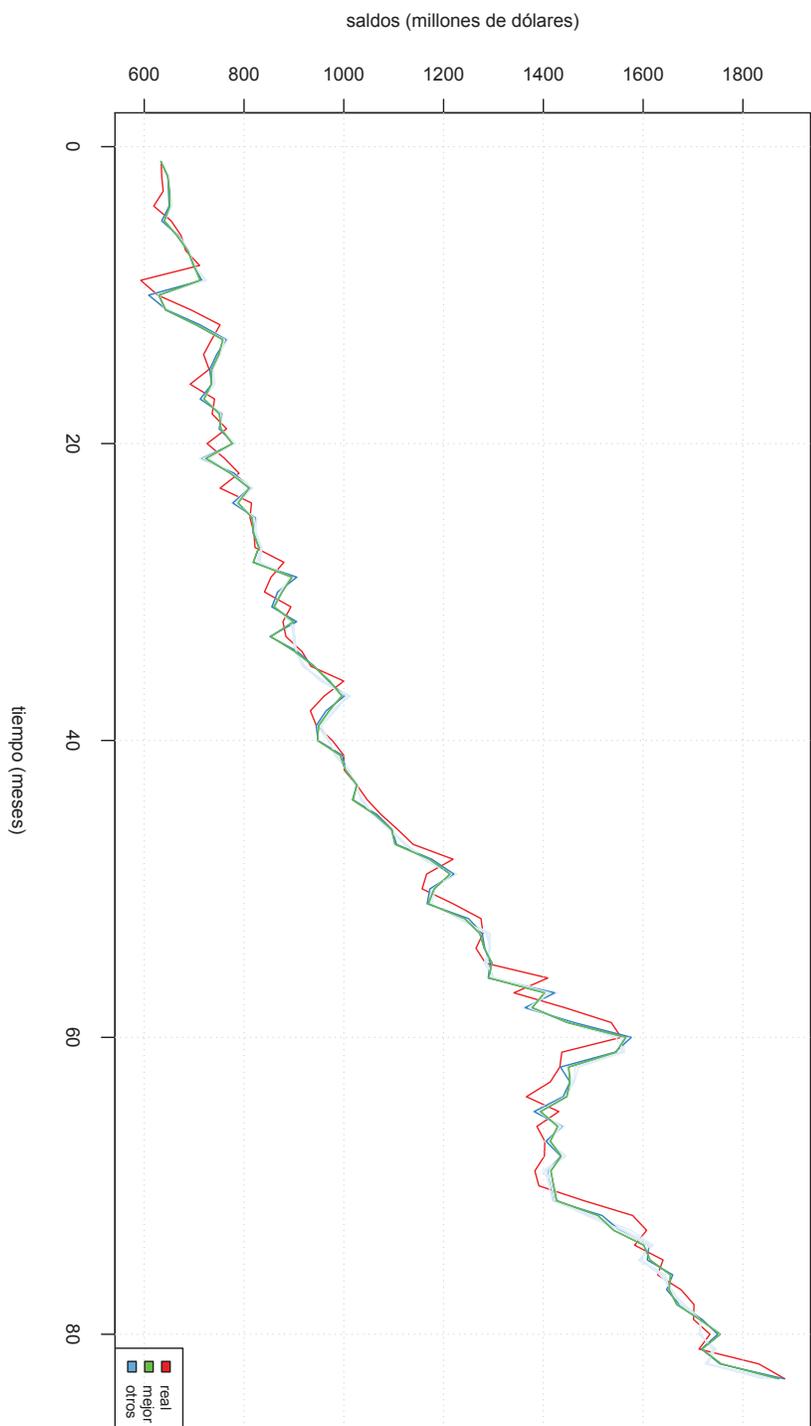
La primera etapa de la modelación genera doce modelos, de los cuales seis resultan ser los más representativos al final del análisis. La Figura 4.4.2.7 muestra los ajustes dichos modelos.

Las Figuras 4.4.2.8 y 4.4.2.9 muestran un patrón similar en todos los modelos. Los residuos atípicos están entre los  $\pm 110$  millones. Se puede visualizar que la caída del saldo en el mes de octubre del 2004 fue alrededor del 20 % del saldo total, la cual es una cifra considerable. La mayor cantidad de los residuos están en el rango de  $\pm 25$  millones que representa aproximadamente el  $\pm 2$  % en cifras relativas.

La Figura 4.4.2.10 muestra claramente uno de los propósitos de la modelación mediante el sistema híbrido difuso. La diferencia entre los pronósticos es muy marcada entre los generados por modelos de alto AIC y los de bajo AIC, a pesar de que los AIC estén en un intervalo pequeño [855.15,860.91]. El pronóstico medio tiene un comportamiento similar a los pronósticos de modelos con bajo AIC y además se desplaza levemente hacia los de alto AIC que influyen por ser más numerosos. El mejor modelo SARIMA es el que tiene los menores pronósticos en todos los meses y su diferencia con los modelos de pronósticos altos, es alrededor de cien millones, que como se vio en los residuos, es un valor que se puede presentar. También es notorio el comportamiento estacional de la serie que se caracteriza por comportamientos crecientes en los últimos meses de año.

La Figura 4.4.2.11 muestra la comparación de los pronósticos del SHD y del mejor modelo SARIMA, con respecto a tres meses reales y se muestra la respectiva zona de confianza. El último dato real muestra una tendencia inusual de crecimiento debido a la estructura de productos utilizada por el banco. La Tabla 4.4.2.1 muestra las diferencias de los pronósticos con respecto a los datos reales, mostrando mejores resultados el SHD.

Figura 4.4.2.7: ajustes de los modelos (monetarios)



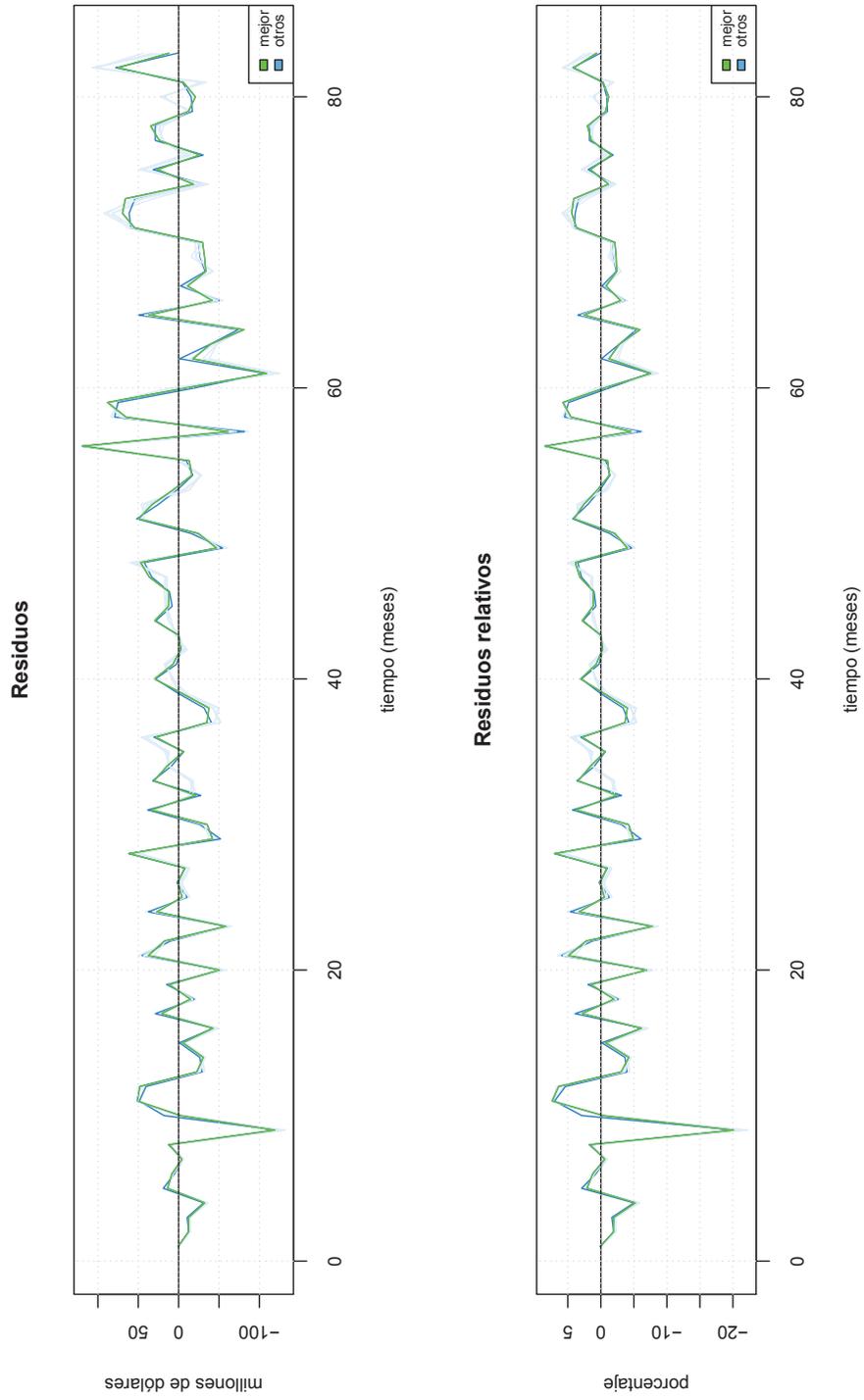


Figura 4.4.2.8: residuos de los modelos (monetarios)

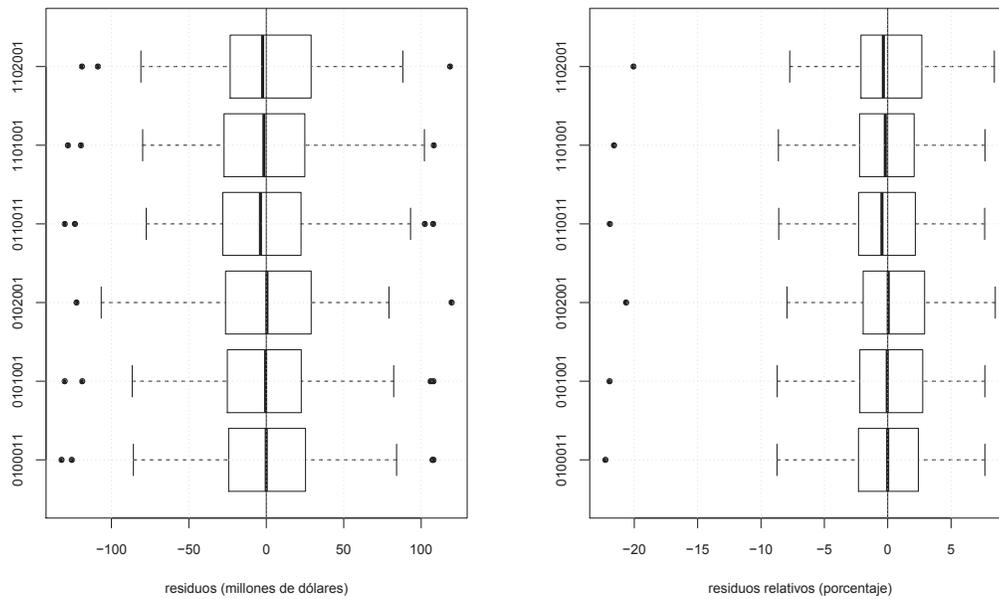


Figura 4.4.2.9: dispersión de residuos y residuos relativos por modelo (monetarios)

mes	MONETARIOS			RMSE		MAE		MAPE	
	real	sarima	shd	sarima_e2	shd_e2	sarima_abs	shd_abs	sarima_relabs	shd_relabs
2010-12	1925	1912	1913	169	144	13	12	0.68	0.62
2011-01	1874	1882	1900	64	676	8	26	0.43	1.39
2011-02	1981	1884	1904	9409	5929	97	77	4.90	3.89
<b>ESTADÍSTICOS</b>				<b>56.69</b>	<b>47.43</b>	<b>39.33</b>	<b>38.33</b>	<b>2.00</b>	<b>1.97</b>

Tabla 4.4.2.1: resultados de pruebas en los pronósticos (monetarios)

Como parte final, se presenta la gráfica que muestra: los datos reales, los ajustes y los resultados de los pronósticos en forma conjunta (Figura 4.4.2.12).

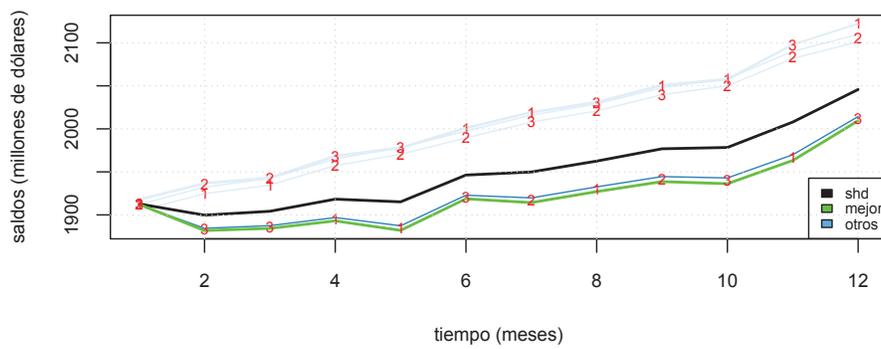
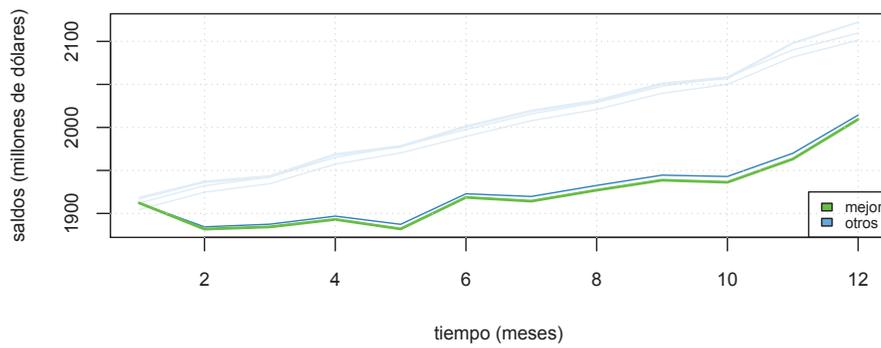


Figura 4.4.2.10: inferencia en los pronósticos (monetarios)

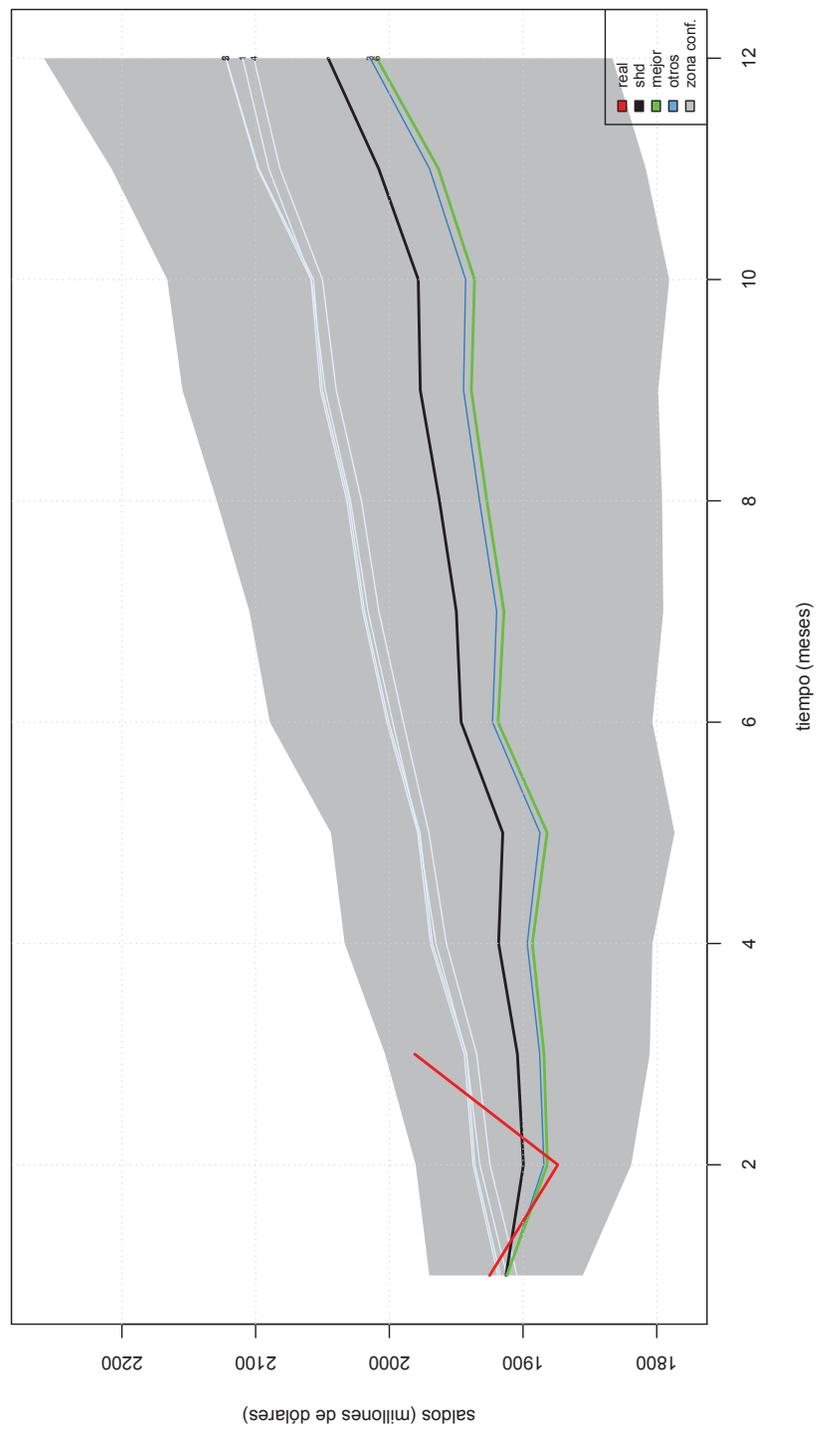


Figura 4.4.2.11: pruebas de pronósticos con datos reales (monetarios)

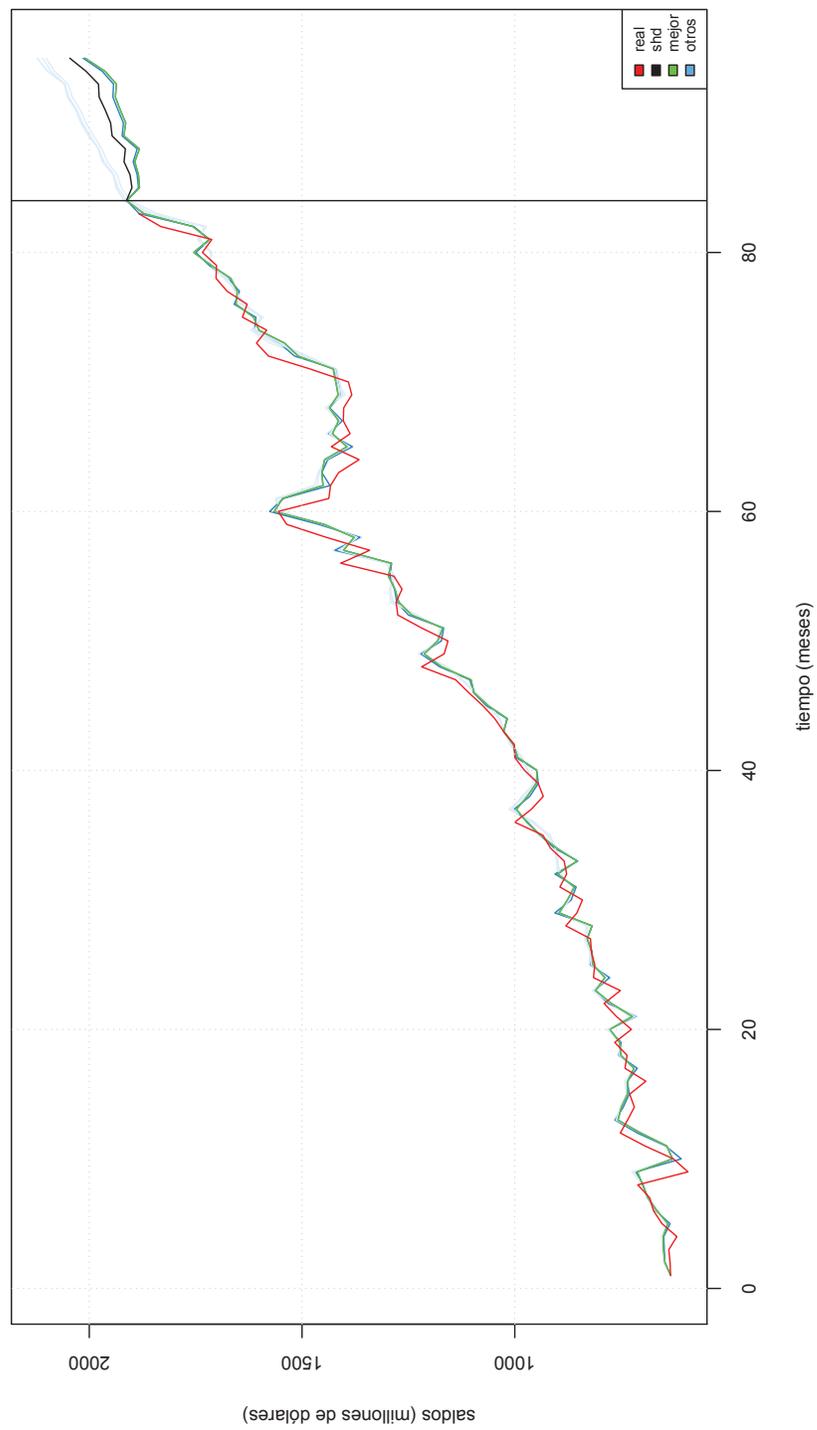


Figura 4.4.2.12: visualización conjunta de resultados (monetarios)

### 4.4.3 SERIE AHORROS

#### 4.4.3.1. Modelación Tradicional Ahorros

La evolución de la serie y sus funciones ACF y PACF son presentadas en una primera instancia. En la Figura 4.4.3.1 se pueden notar los siguientes aspectos:

- La serie tiene una tendencia creciente en la mayoría de años con excepción del año 2009, que presenta un comportamiento decreciente en los primeros meses. El efecto causado por el tema de la dolarización también está presente en esta serie.
- Al igual que en la serie de monetarios, esta presenta pequeñas variaciones periódicas en la tendencia a causa de una posible estacionalidad, la que será analizada por descomposición más adelante.
- La corrida de depósitos del mes de octubre del 2004 también es notoria en esta serie.

El decaimiento rápido de la función ACF en la Figura 4.4.3.1, significa que la serie debe ser diferenciada estacionariamente. De igual manera se llega a esta conclusión en la gráfica de los retardos (Figura 4.4.3.2), que muestra la correlación más alta en la diferenciación de primer orden, como también se concluye con el resultado de la función  $ndiffs(ndiffs(dataho) = 1)$ .

El modelo inicial se plantea en base al análisis de las funciones ACF y PACF de la serie diferenciada de orden 1 (Figura 4.4.3.3).

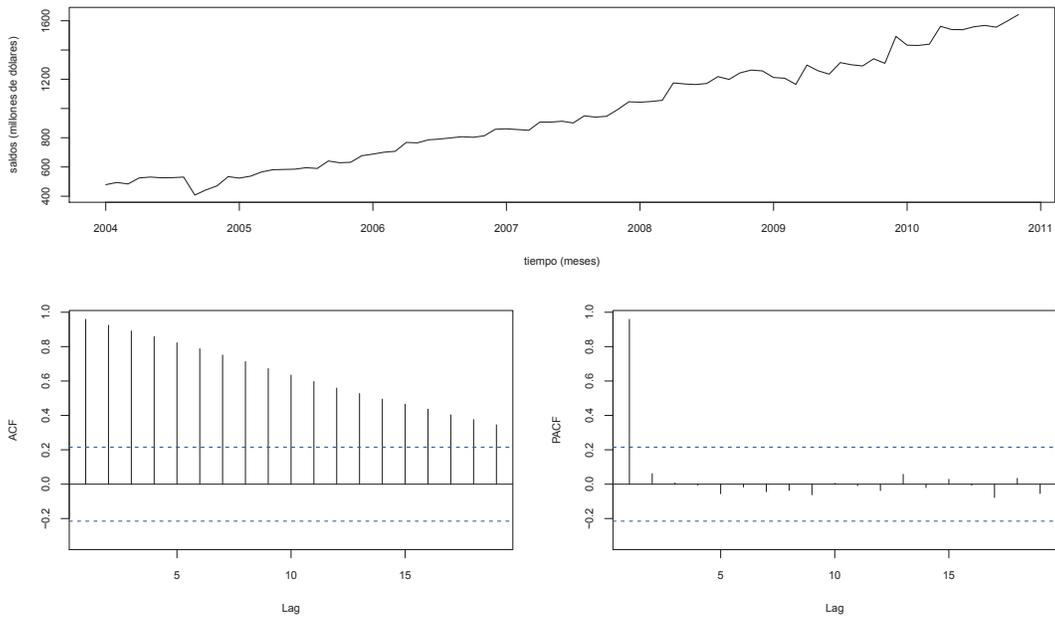


Figura 4.4.3.1: serie de ahorros y funciones ACF y PACF

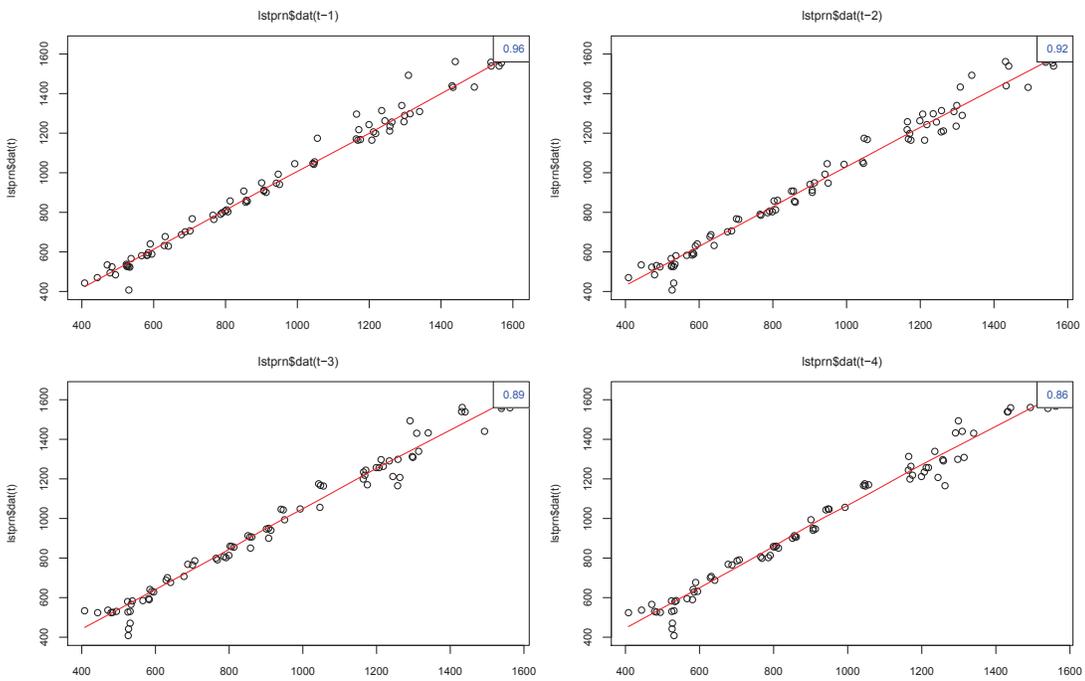


Figura 4.4.3.2: retardos del orden 1 al 4

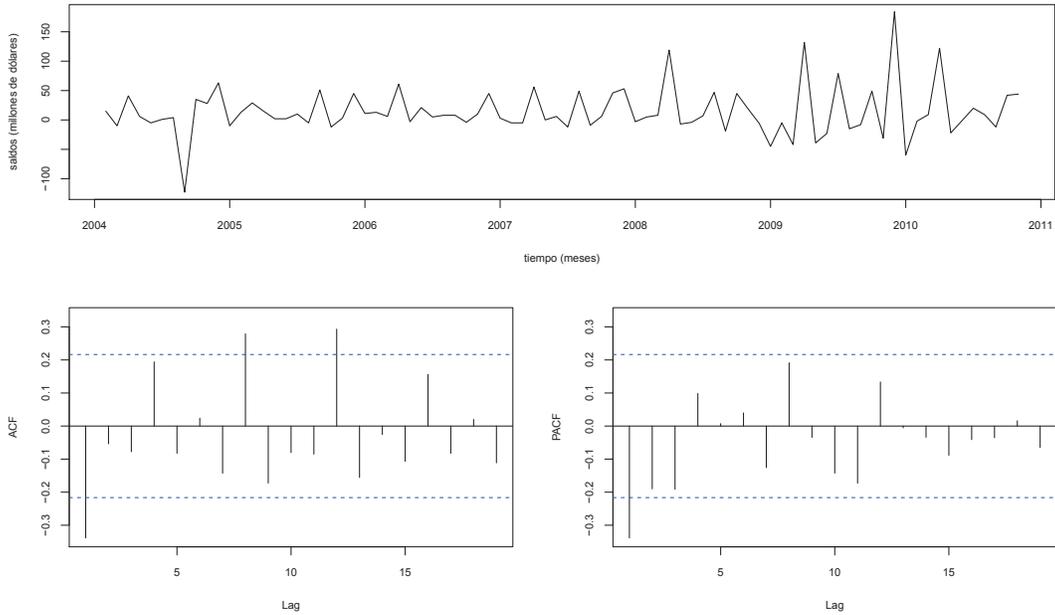


Figura 4.4.3.3: serie ahorros diferenciada de orden 1

La búsqueda de la minimización del AIC es realizada, y se llega al siguiente modelo:

```
> mdlaho
Series: dataho
ARIMA(2,1,0)(2,0,0)[12] with drift
Call: arima(x = dataho)
Coefficients:
          ar1      ar2      sar1      sar2      drift
      -0.4452  -0.1849   0.1741   0.4769  13.2853
s.e.    0.1133   0.1105   0.1167   0.1354   4.5863
sigma^2 estimated as 1076:  log likelihood = -406.47
AIC = 824.94   AICc = 826.06   BIC = 839.38
```

Las gráficas de las funciones ACF y PACF de los residuos (Figura 4.4.3.4) validan el modelo por no presentar correlación en los retardos.

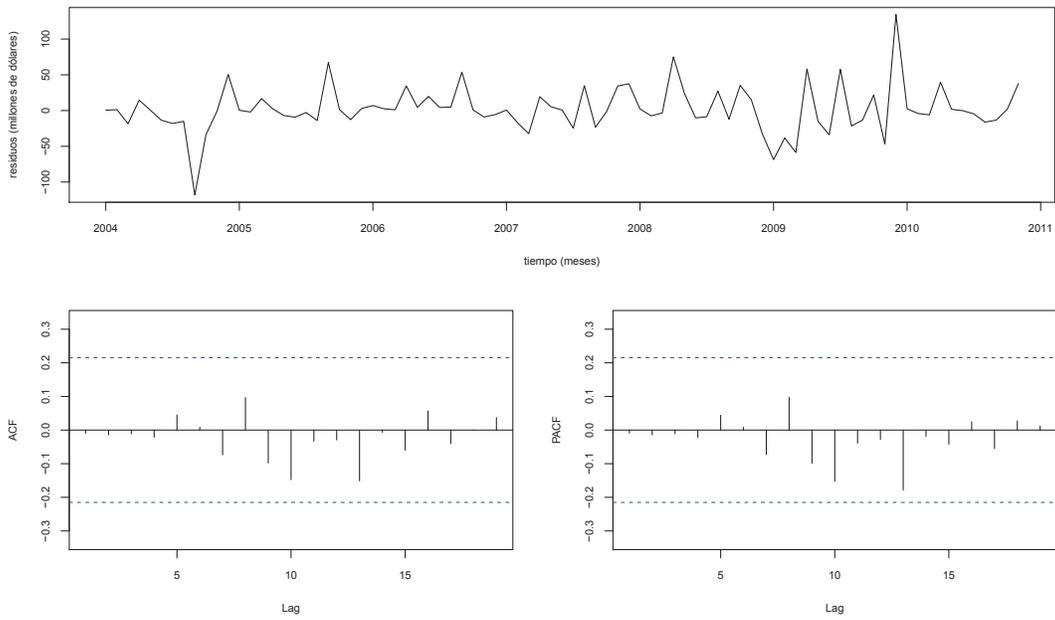


Figura 4.4.3.4: residuos y funciones ACF y PACF (ahorros)

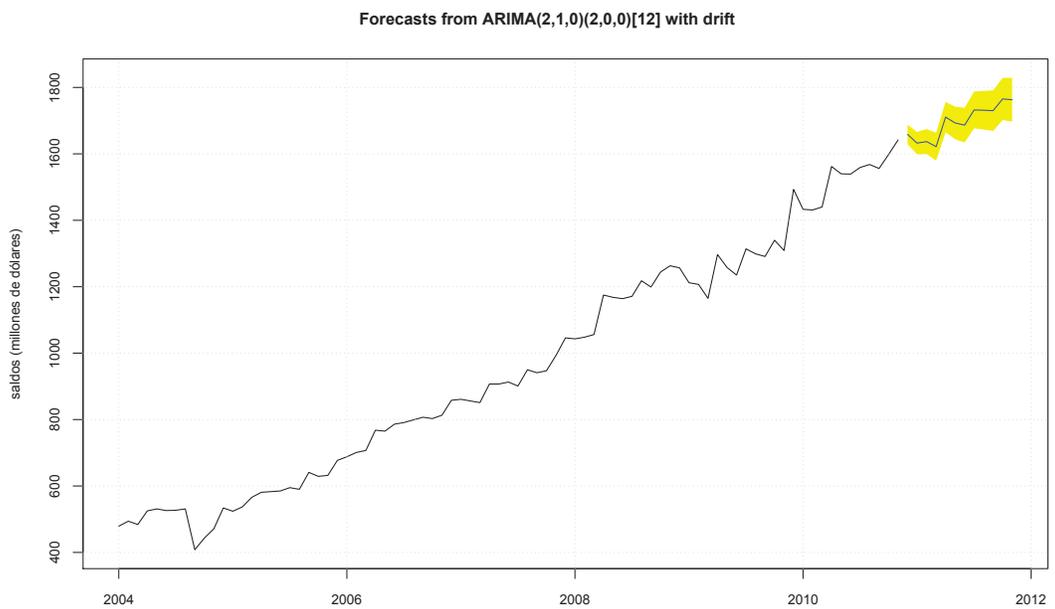


Figura 4.4.3.5: pronósticos de la serie de ahorros

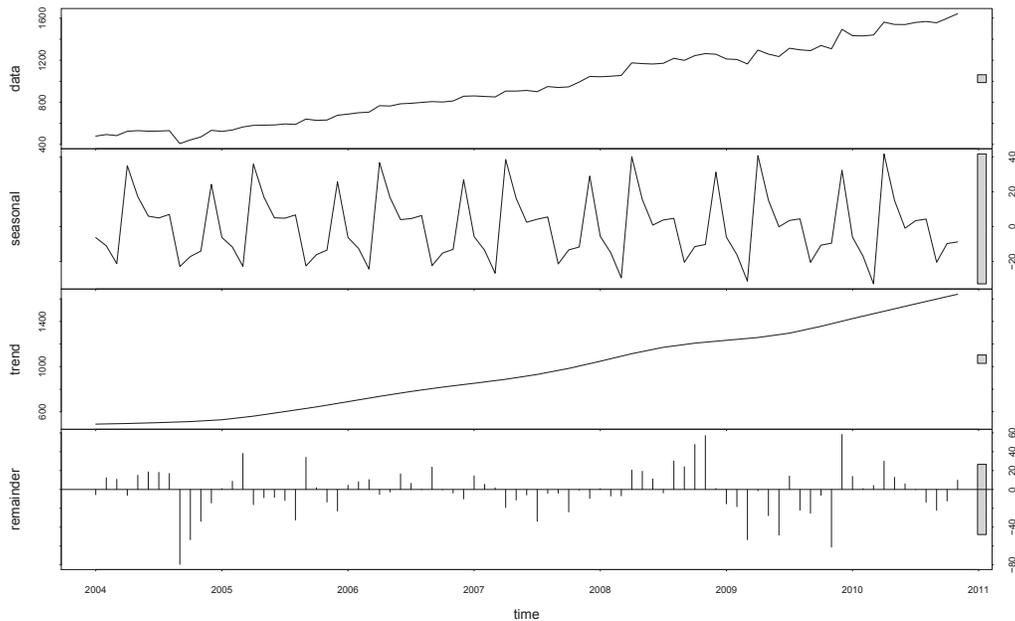


Figura 4.4.3.6: descomposición de la serie (ahorros)

Los pronósticos son generados con el modelo final, los que presentan una tendencia creciente, acompañados de las fluctuaciones estacionales (Figura 4.4.3.5).

El análisis de descomposición en la Figura 4.4.3.6, muestra que el componente estacional está caracterizado por dos crecientes: una en el mes de abril debido al pago de utilidades, y la otra en diciembre por el pago del décimo tercer sueldo y bonos, que son acreditados en su mayoría a las cuentas de ahorros. Los residuos no tienen un comportamiento aleatorio, esto es un indicio de que la estacionalidad es débil.

#### **4.4.3.2. Modelación Sistema Híbrido Difuso Ahorros**

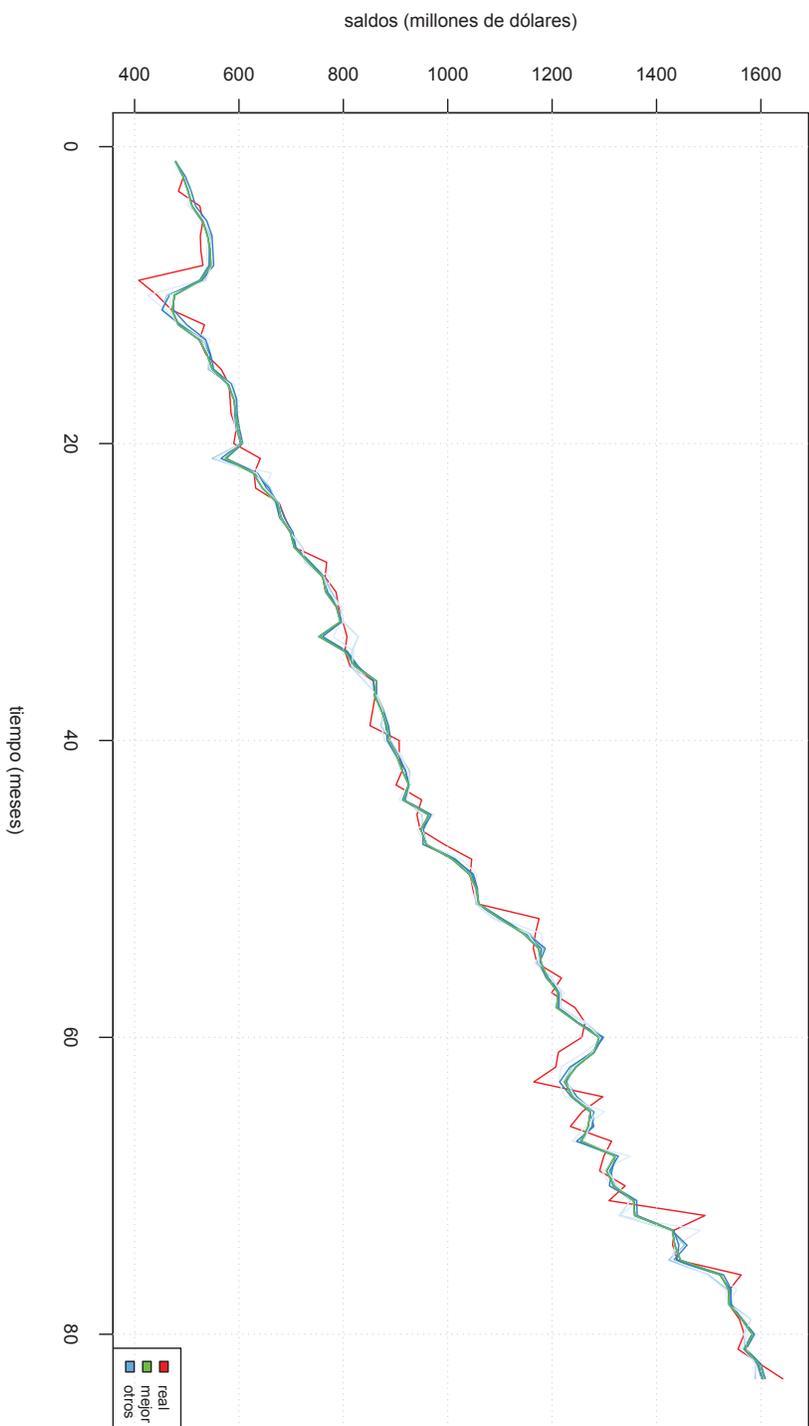
El sistema genera inicialmente doce modelos, de los cuales siete terminan siendo relevantes. La Figura 4.4.3.7 muestra los ajustes de estos modelos.

Las gráficas de los residuos (Figuras 4.4.3.8 y 4.4.3.9) muestran un patrón similar en todos los modelos. La corrida de depósitos genera un residuo de alrededor de los -100 millones que representa un  $-30\%$ , considerando al resto de datos como atípicos si su valor supera los  $\pm 50$  millones en promedio y que en porcentaje relativo es de  $\pm 5\%$ .

Al igual que en la modelación de monetarios, la Figura 4.4.3.10 muestra las ventajas de la modelación con el sistema híbrido difuso. La brecha entre los modelos extremos es variable, desde los 30 millones hasta los 100 millones. El pronóstico medio se aproxima hacia los pronósticos de modelos de bajo AIC, pero también considera al resto de modelos, pero a menor grado, también recoge el comportamiento estacional y la tendencia creciente.

En la comparación de los pronósticos del SHD y del mejor modelo SARIMA, con respecto a tres meses reales (Figura 4.4.3.11) el dato del mes de diciembre 2010 es inusual debido a una campaña realizada por el banco para la captación de clientes en este producto, el impacto de esta campaña es alrededor de 30 millones, a lo que se agrega, el efecto que provoca la combinación de productos que ofrece el banco. En la Tabla 4.4.3.1 se muestra las diferencias de los pronósticos con respecto a los datos reales, al igual que en las otras series, los resultados son mejores con el SHD.

Figura 4.4.3.7: ajustes de los modelos (ahorros)



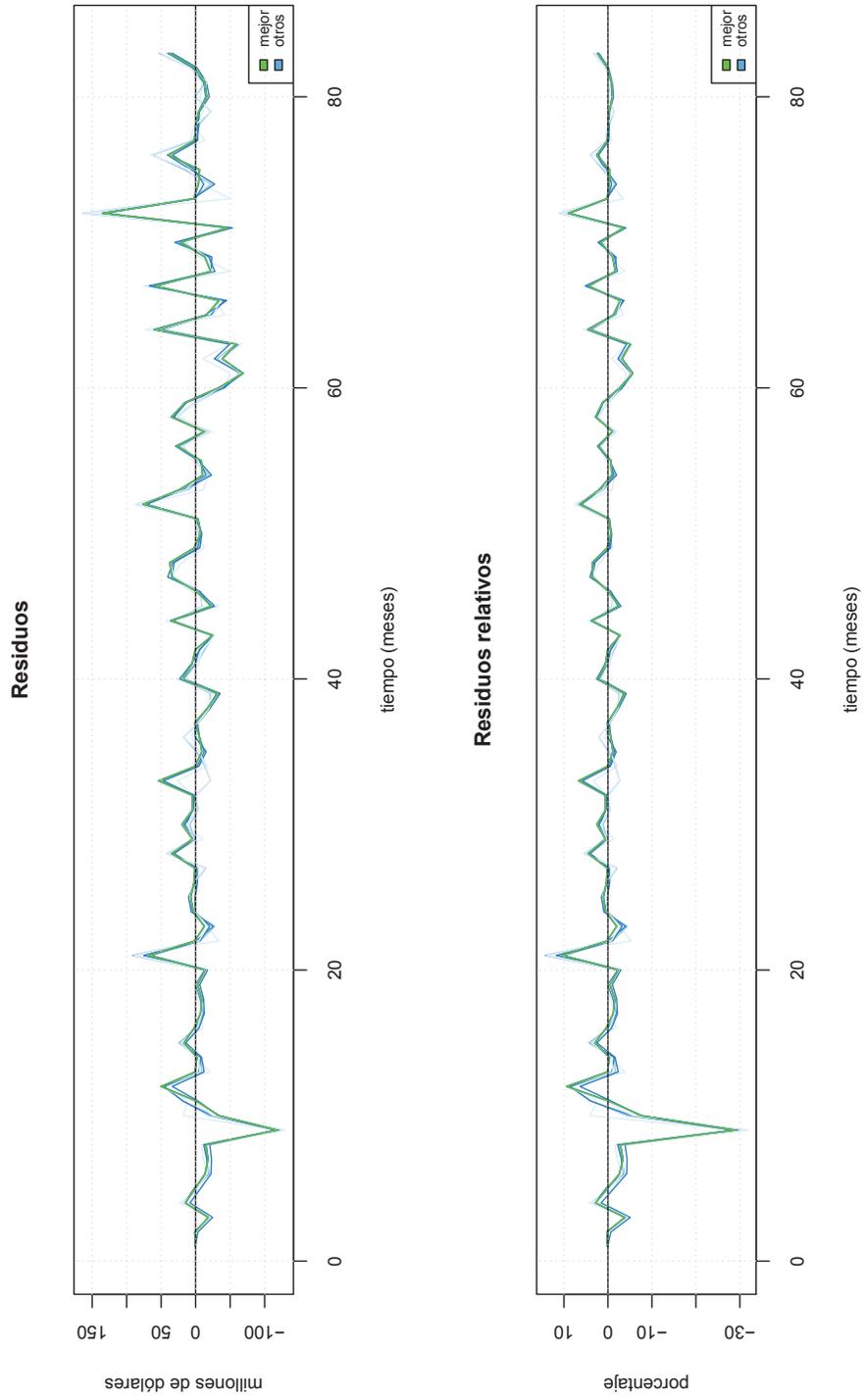


Figura 4.4.3.8: residuos de los modelos (ahorros)

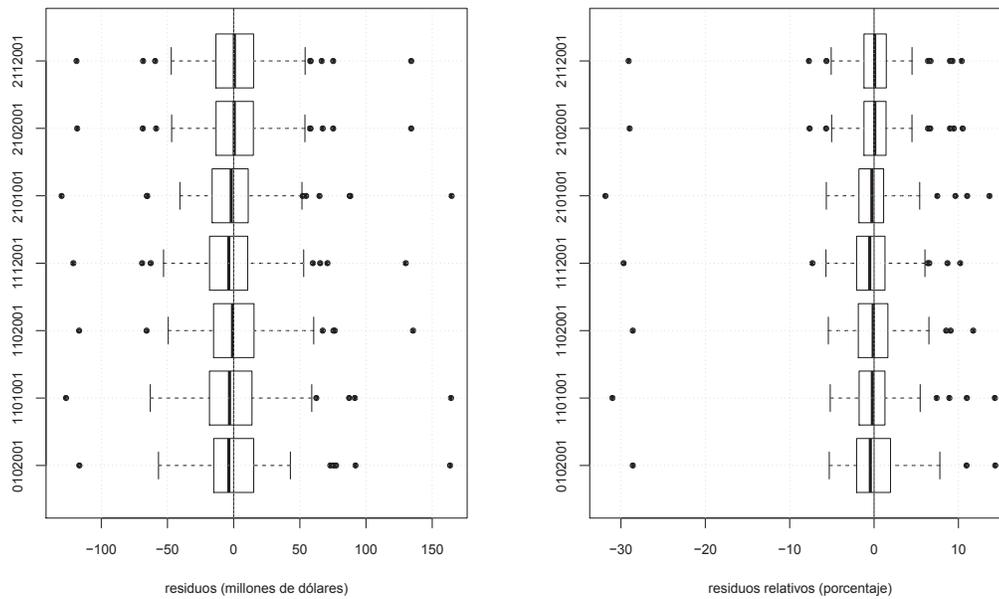


Figura 4.4.3.9: dispersión de residuos y residuos relativos por modelo (ahorros)

mes	AHORROS			RMSE		MAE		MAPE	
	real	sarima	shd	sarima.e2	shd.e2	sarima.abs	shd.abs	sarima.relabs	shd.relabs
2010-12	1804	1659	1672	21025	17424	145	132	8.04	7.32
2011-01	1702	1632	1648	4900	2916	70	54	4.11	3.17
2011-02	1726	1637	1652	7921	5476	89	74	5.16	4.29
<b>ESTADÍSTICOS</b>				<b>106.22</b>	<b>92.76</b>	<b>101.33</b>	<b>86.67</b>	<b>5.77</b>	<b>4.93</b>

Tabla 4.4.3.1: resultados de pruebas en los pronósticos (ahorros)

La gráfica que muestra: los datos reales, los ajustes y los resultados de los pronósticos en forma conjunta (Figura 4.4.3.12) es presentada como parte final.

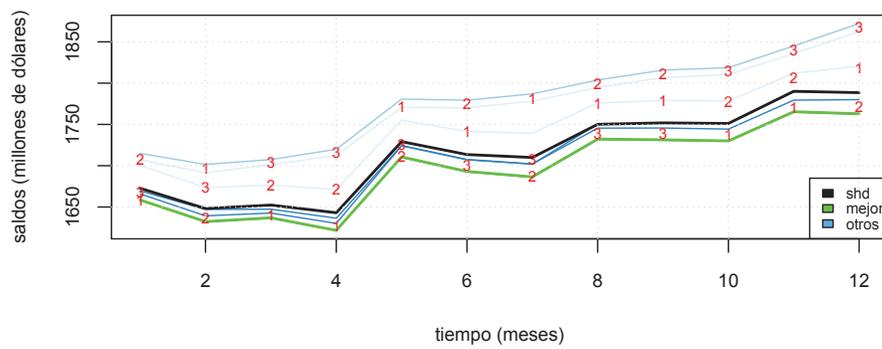
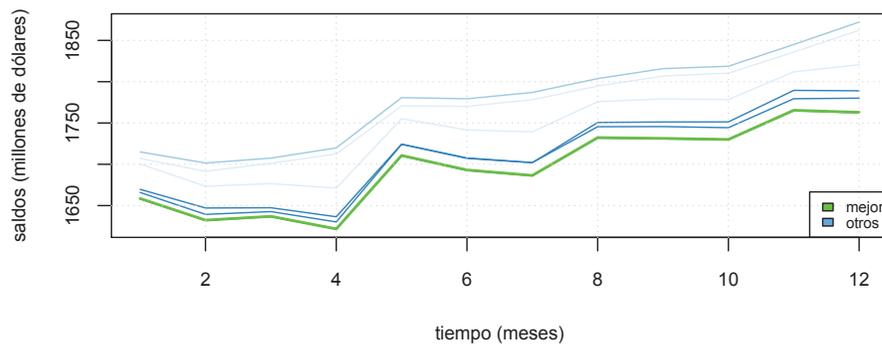


Figura 4.4.3.10: inferencia en los pronósticos (ahorros)

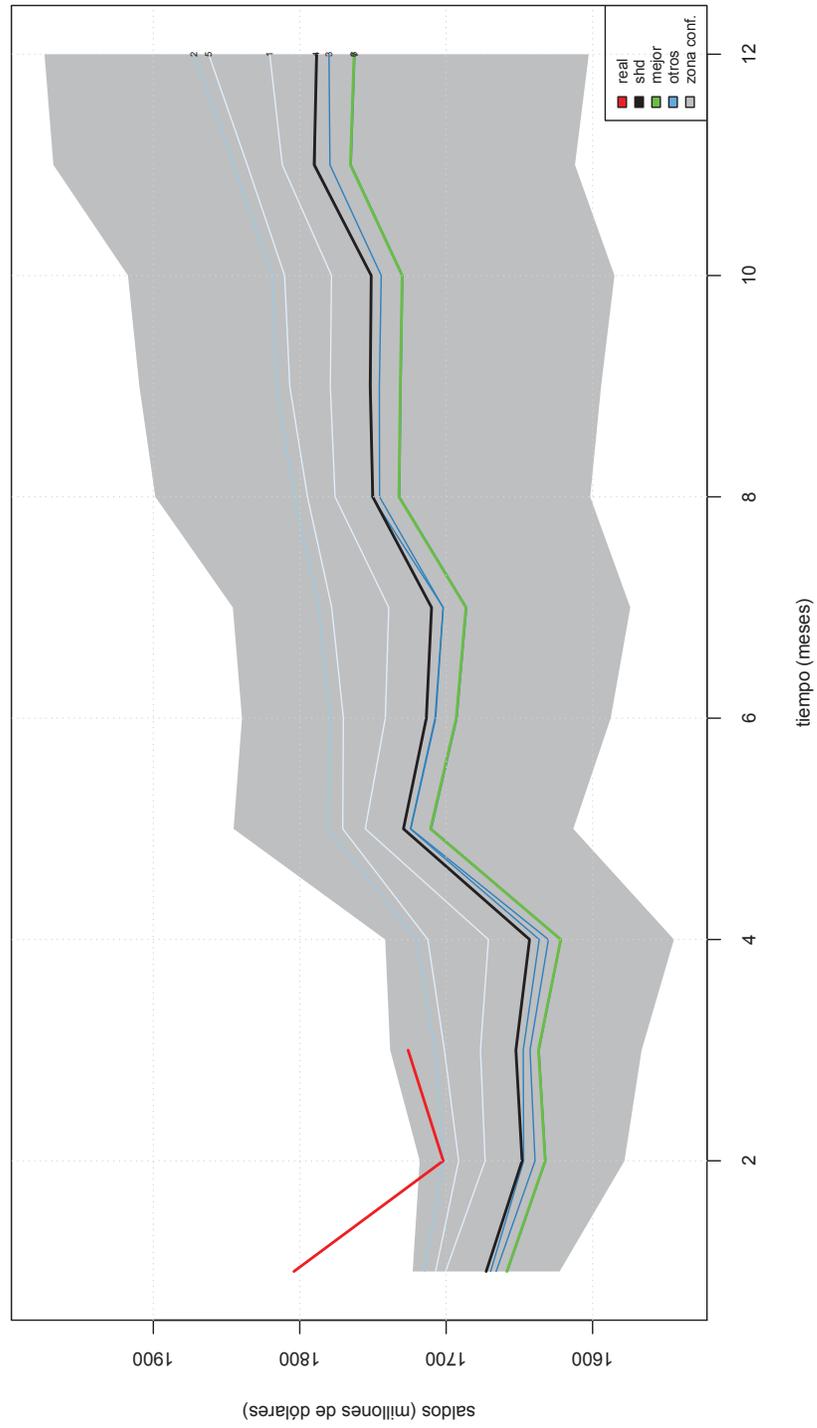


Figura 4.4.3.11: pruebas de pronósticos con datos reales (ahorros)

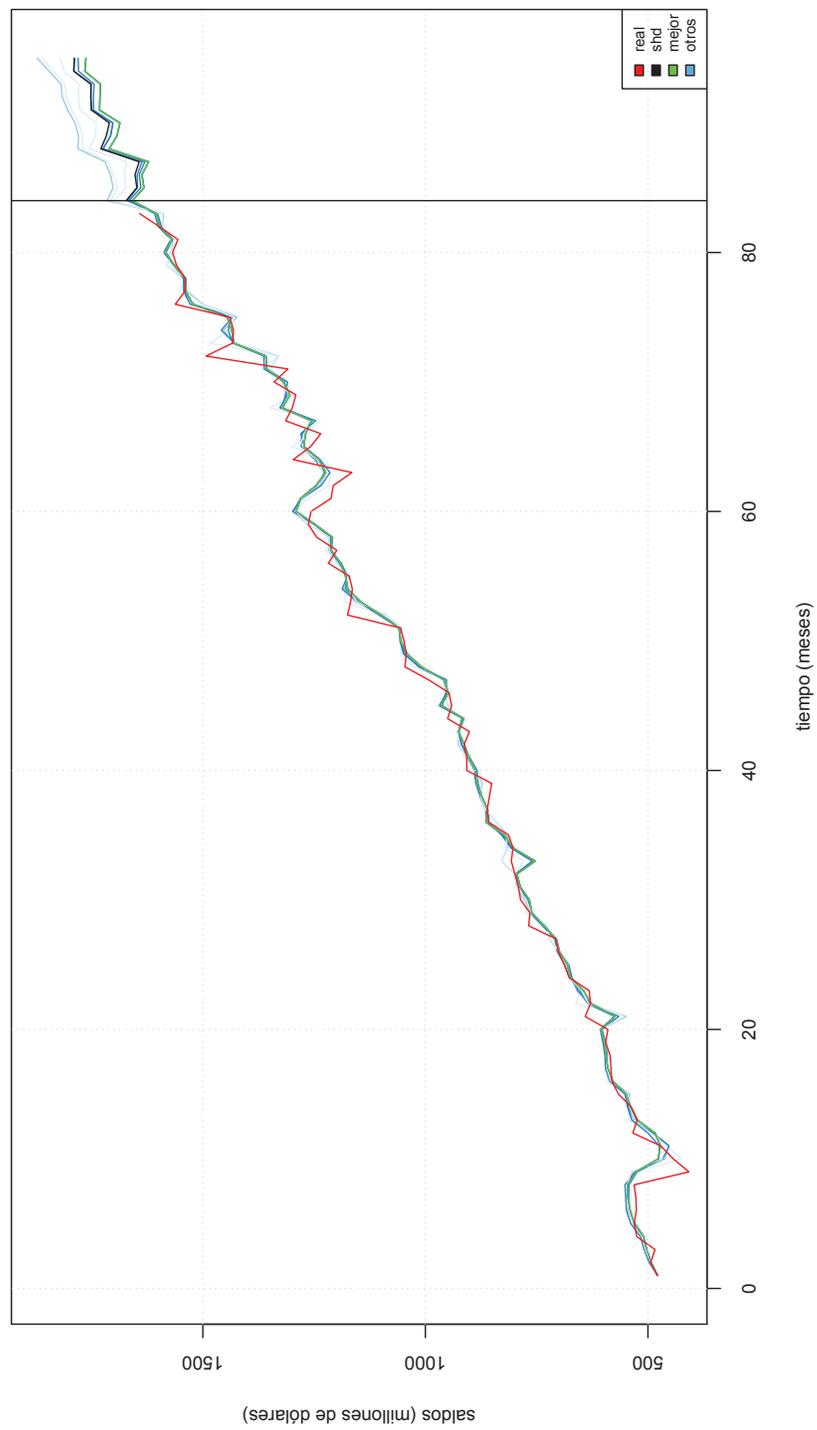


Figura 4.4.3.12: visualización conjunta de resultados (ahorros)

## 4.4.4 SERIE PLAZO

### 4.4.4.1. Modelación Tradicional Plazo

La evolución de la serie y sus funciones ACF y PACF son presentadas en una primera instancia. En la Figura 4.4.4.1 se pueden notar los siguientes aspectos:

- La serie tiene una tendencia cambiante, pero en gran parte creciente. El decaimiento en el 2004 es debido a la corrida de depósitos, mientras que la caída del 2010, es debido a la política económica del gobierno con respecto a la mantención de fondos de empresas públicas en bancos privados.
- La serie no presenta fluctuaciones que podrían ser de estacionalidad, por lo que no se realiza el correspondiente análisis.

El decaimiento de la función ACF en la Figura 4.4.4.1, sugiere una diferenciación estacionaria. Esto es ratificado por la gráfica de los retardos (Figura 4.4.4.2), donde la correlación es la más alta en la diferenciación de primer orden y por el resultado de la función  $ndiffs(ndiffs(datplz = 1))$ .

La búsqueda empieza con el modelo tentativo planteado con el análisis de las funciones ACF y PACF de la serie diferenciada de orden 1 (Figura 4.4.4.3). La búsqueda del modelo que minimice el AIC se realiza, y se obtiene el siguiente modelo:

```
> mdlplz
Series: datplz
ARIMA(1,1,0) with drift
Call: arima(x = datplz)
Coefficients:
      ar1  drift
      0.3079  5.1789
s.e.    0.1058  3.1105
sigma^2 estimated as 383.9:  log likelihood = -360.37
AIC = 726.75  AICc = 727.05  BIC = 733.97
```

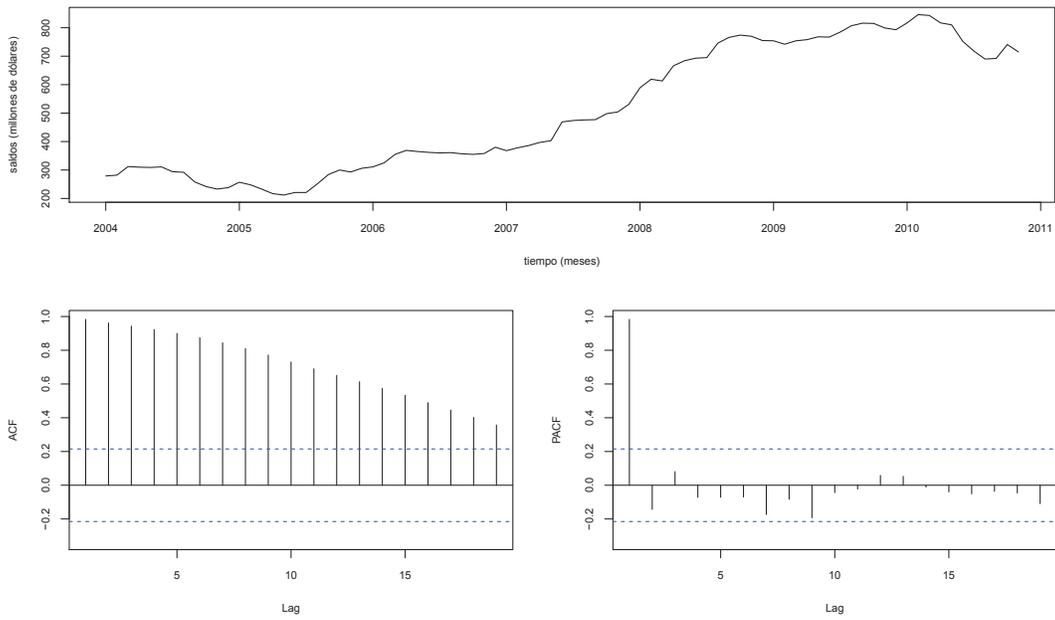


Figura 4.4.4.1: serie de plazo y funciones ACF y PACF

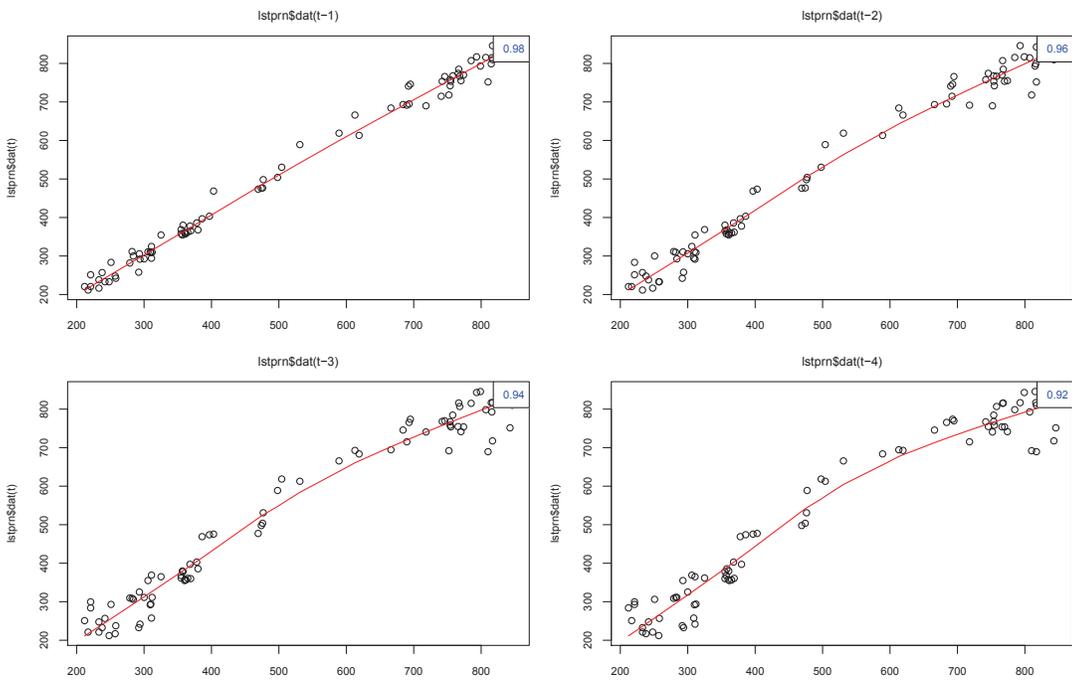


Figura 4.4.4.2: retardos del orden 1 al 4

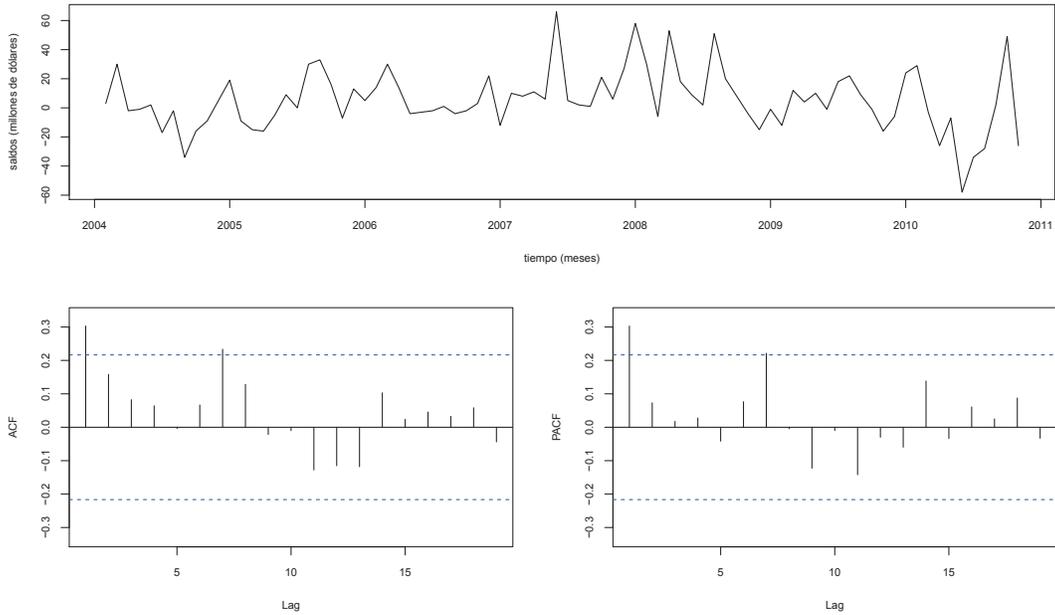


Figura 4.4.4.3: serie plazo diferenciada de orden 1

Las gráficas de las funciones ACF y PACF de los residuos (Figura 4.4.4.4) muestran un comportamiento aleatorio, por lo que el modelo es validado.

Los pronósticos son generados con el modelo final, y tienen una tendencia creciente y lineal (Figura 4.4.4.5). Este comportamiento no se ha presentado históricamente, por lo que en este caso el mejor modelo SARIMA no es admisible. Este es otro de los propósitos de la modelación mediante un sistema híbrido difuso, dar una alternativa de solución a la problemática que se presenta cuando el mejor modelo en base a la minimización del CI no es aceptado por los expertos.

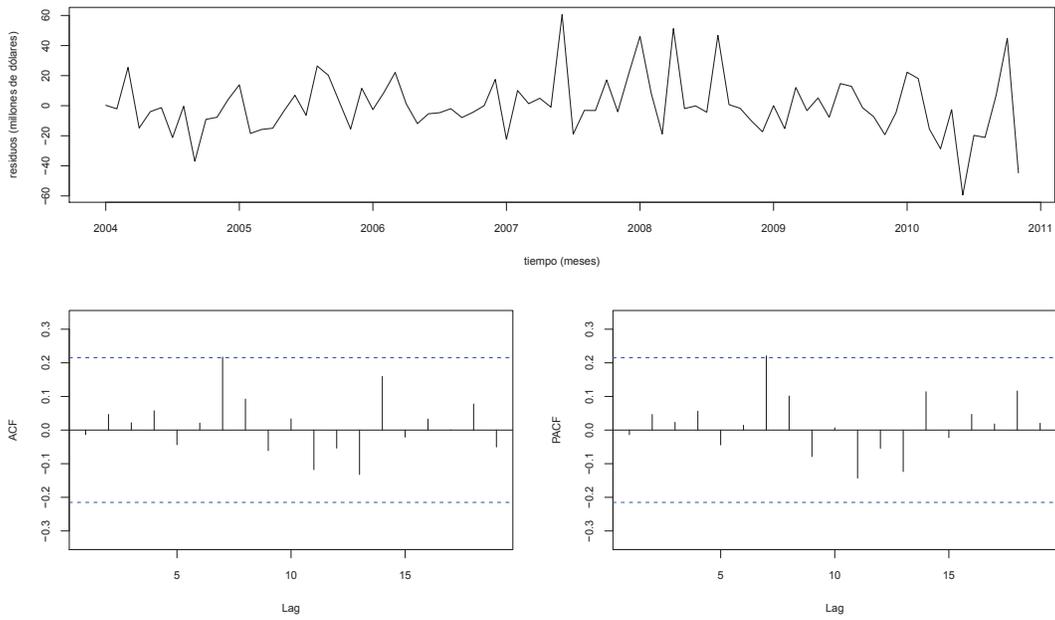


Figura 4.4.4.4: residuos y funciones ACF y PACF (plazo)

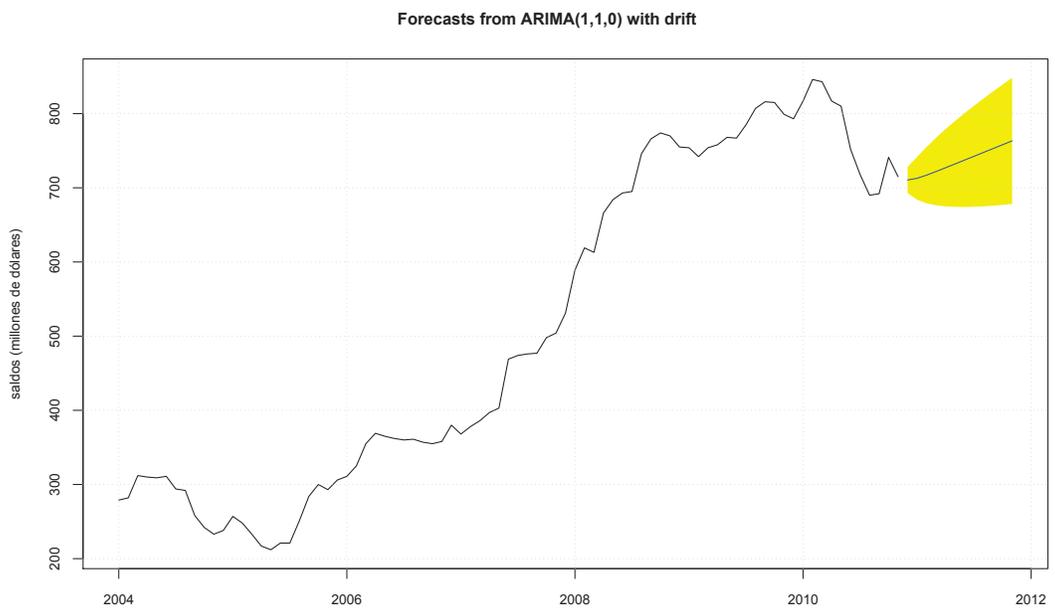


Figura 4.4.4.5: pronósticos de la serie de plazo

#### 4.4.4.2. Modelación Sistema Híbrido Difuso Plazo

El sistema genera inicialmente once modelos, que luego del término del proceso de modelación, resultan siete los relevantes, en los cuales no consta el mejor modelo SARIMA por presentar pronósticos no admisibles. La Figura 4.4.4.6 muestra los ajustes de estos modelos.

Las gráficas de los residuos (Figuras 4.4.4.7 y 4.4.4.8) muestran un patrón similar en todos los modelos. La corrida de depósitos genera un residuo de alrededor de los -40 millones que representa un  $-15\%$  con respecto al total del saldo de ese mes. Esta es la modelación que mayor número de datos atípicos presenta, debido a los cambios de comportamiento de la serie.

La Figura 4.4.4.9 muestra la opción que brinda la modelación con el sistema híbrido difuso. El pronóstico medio se aproxima hacia los pronósticos de modelos de bajo AIC, pero también considera al resto de modelos. Los modelos considerados no presentan un comportamiento lineal exacto para los pronósticos como lo hace el mejor modelo SARIMA.

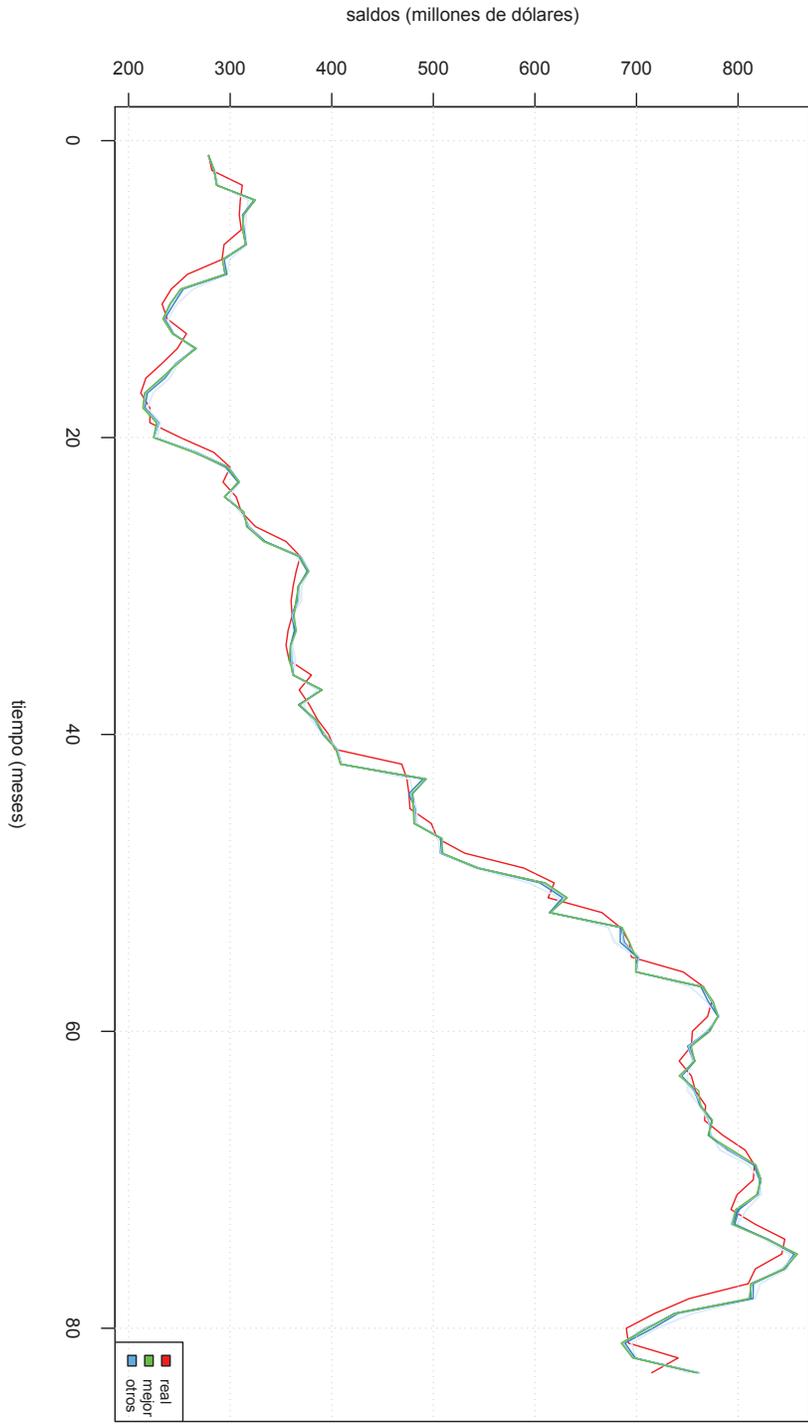
La comparación de los pronósticos del SHD y del mejor modelo SARIMA (a pesar de que no es admisible) con respecto a tres meses reales (Figura 4.4.4.10) se presenta. El dato del mes de febrero 2011 contiene el efecto de depósitos del sector público de alrededor de 30 millones.

En la Tabla 4.4.4.1 se muestra las diferencias de los pronósticos con respecto a los datos reales.

mes	PLAZO			RMSE		MAE		MAPE	
	real	sarima	shd	sarima_e2	shd_e2	sarima_abs	shd_abs	sarima_relabs	shd_relabs
2010-12	705	711	714	36	81	6	9	0.85	1.28
2011-01	712	713	715	1	9	1	3	0.14	0.42
2011-02	747	717	718	900	841	30	29	4.02	3.88
<b>ESTADÍSTICOS</b>				<b>17.67</b>	<b>17.62</b>	<b>12.33</b>	<b>13.67</b>	<b>1.67</b>	<b>1.86</b>

Tabla 4.4.4.1: resultados de pruebas en los pronósticos (plazo)

Figura 4.4.4.6: ajustes de los modelos (plazo)



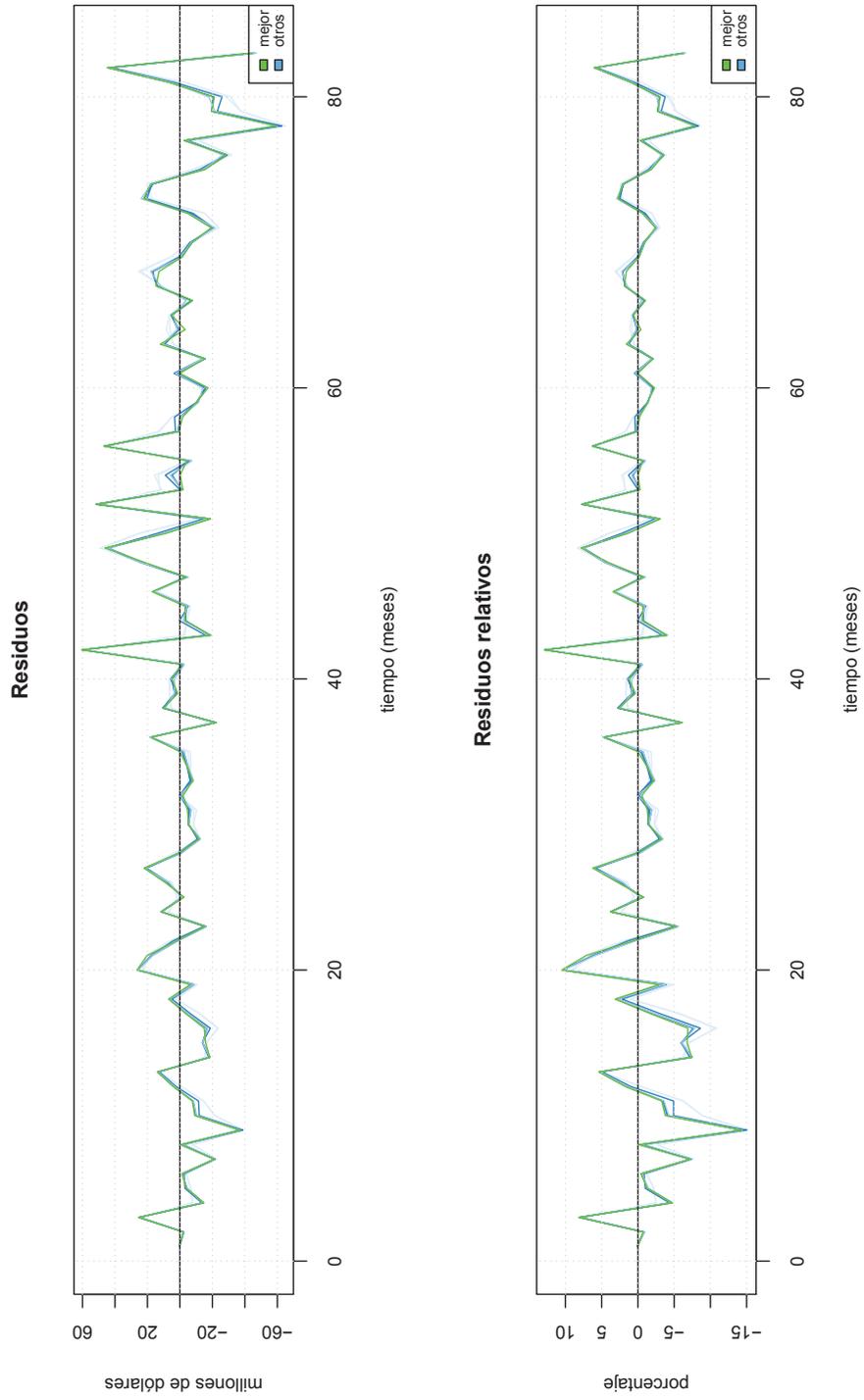


Figura 4.4.4.7: residuos de los modelos (plazo)

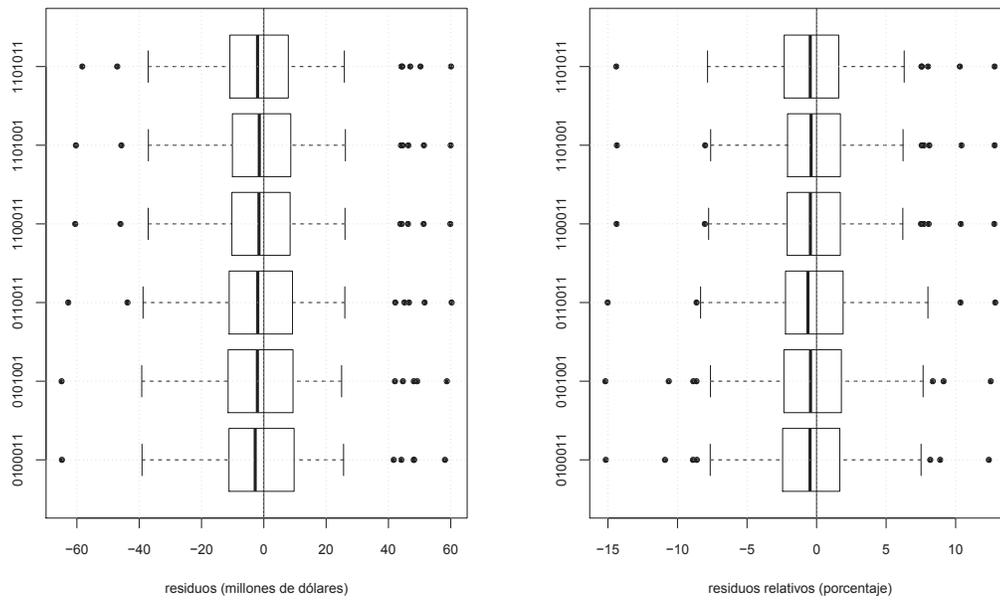


Figura 4.4.4.8: dispersión de residuos y residuos relativos por modelo (plazo)

Para culminar con la modelación, se presenta una gráfica que muestra: los datos reales, los ajustes y los resultados de los pronósticos en forma conjunta (Figura 4.4.4.11).

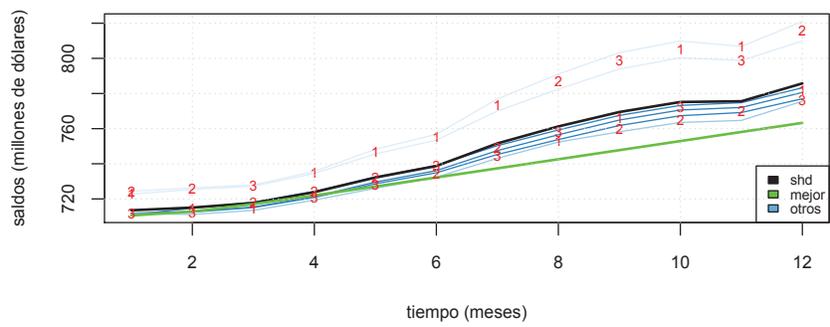
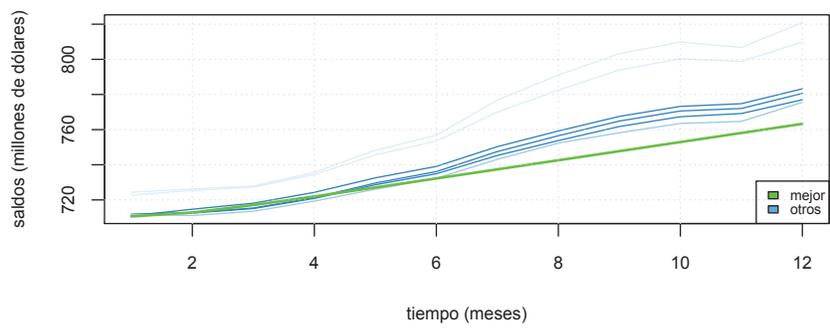


Figura 4.4.4.9: inferencia en los pronósticos (plazo)

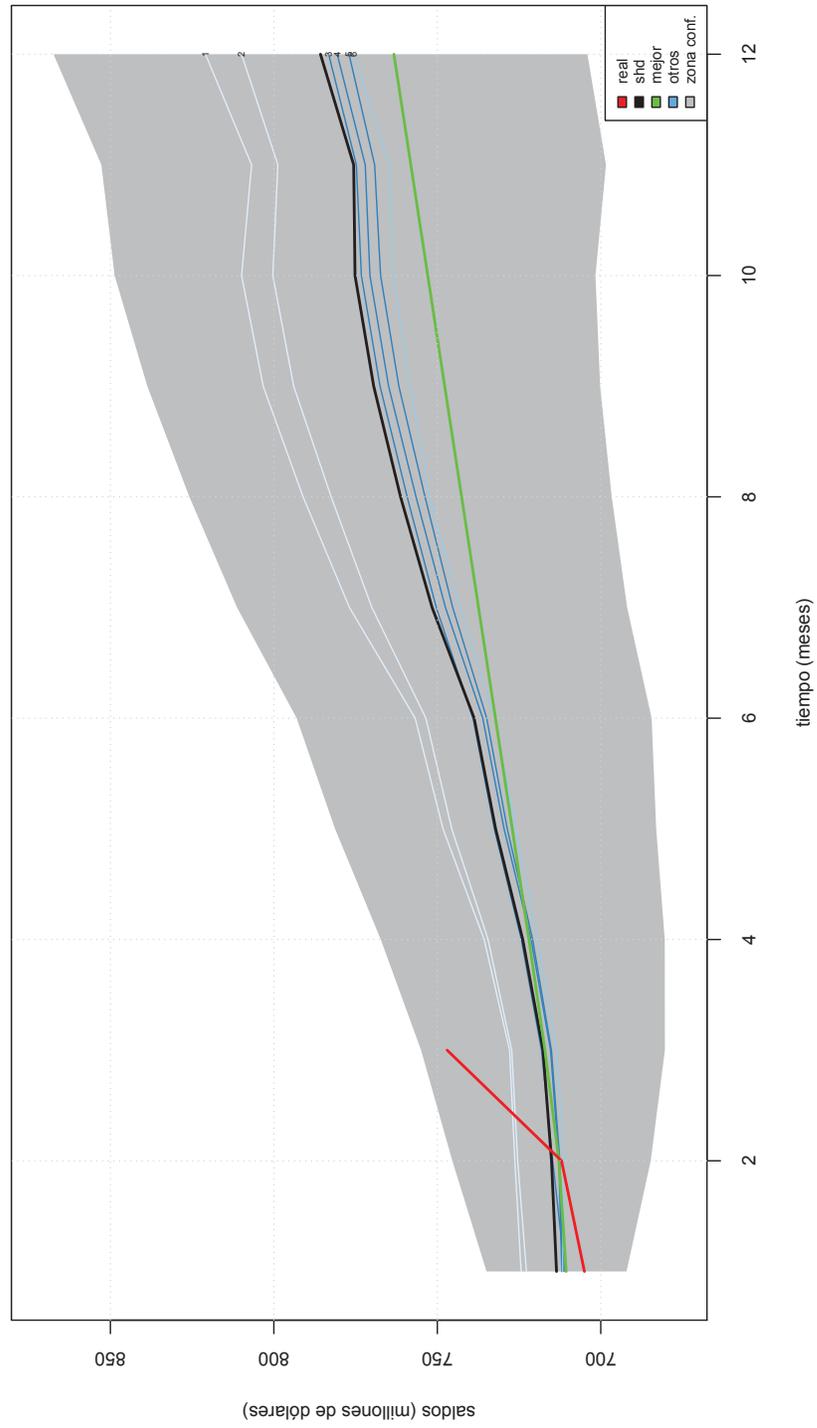


Figura 4.4.4.10: pruebas de pronósticos con datos reales (plazo)

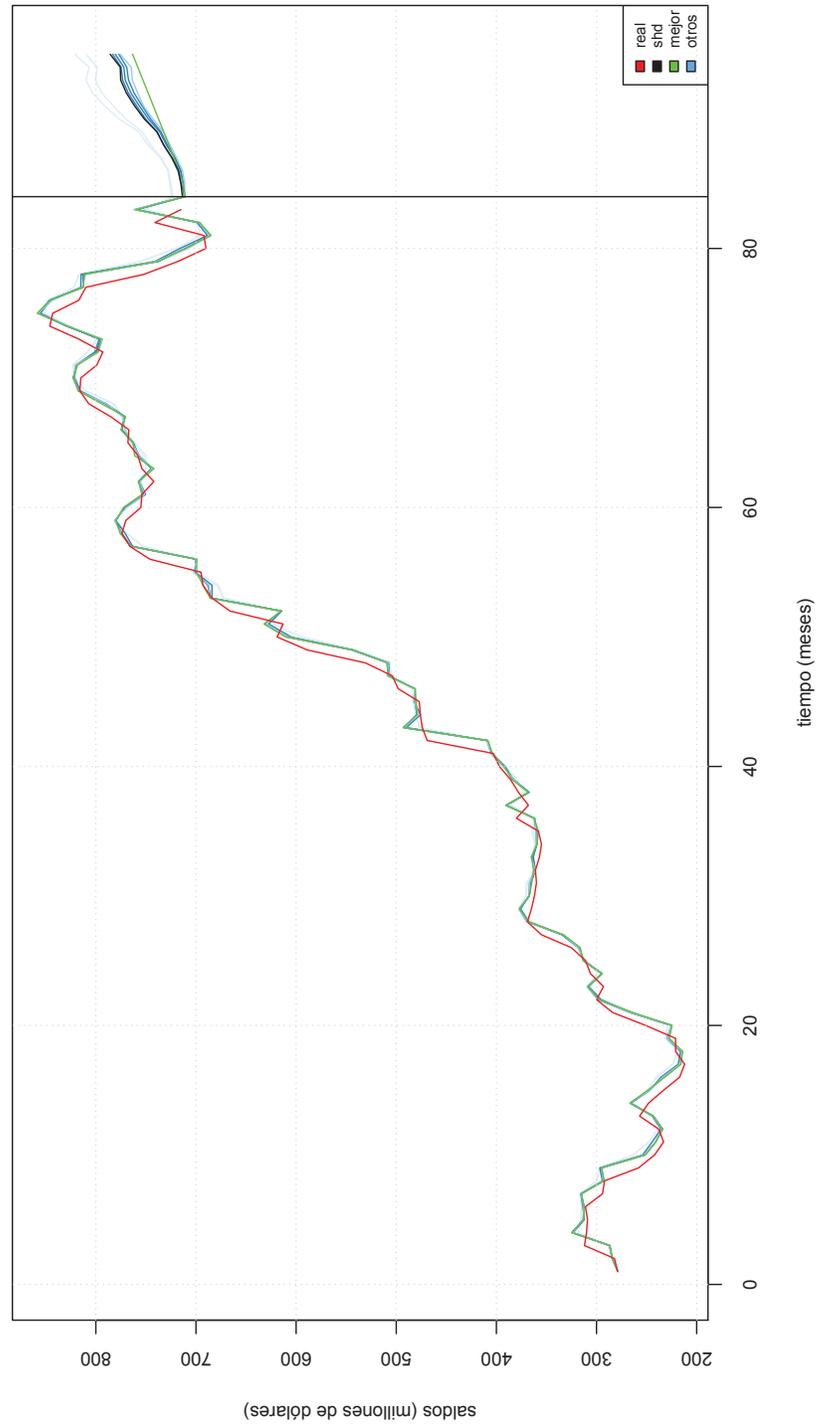


Figura 4.4.4.11: visualización conjunta de resultados (plazo)

## 5 CONCLUSIONES Y RECOMENDACIONES

El trabajo desarrollado muestra una perspectiva alentadora con respecto al uso de la inteligencia computacional en el tratamiento de datos financieros. Particularmente, los principios de lógica difusa como técnica de agrupamiento y su incorporación en el sistema híbrido permiten modelar fenómenos financieros bajo incertidumbre adecuadamente. Cabe señalar que la metodología utilizada no se restringe a aplicaciones en Finanzas, sino que puede utilizarse en diversas áreas tales como: meteorología, bioinformática, control de procesos, etc., y su estudio es un problema abierto.

La técnica de modelación en base a sistemas híbridos difusos puede entenderse como un método local, el cual realiza una partición difusa del dominio en un número finito de regiones sobre las cuales se aplican inferencias para identificar patrones, en general, tendencias. Una vez agrupados los patrones se pueden ponderar los mismos y obtener un valor representativo de todo el dominio, con lo que se logra entender el comportamiento del fenómeno en una forma conjunta.

El método propuesto es computacionalmente eficiente respecto al tiempo de proceso y requerimientos de memoria, esto permite modelar sistemas a gran escala (cientos de series con historial mayor a 10 años con datos mensuales).

El sistema es validado con cuatro series financieras, en las cuales los resultados son mejores que las técnicas tradicionales, en este caso, modelos SARIMA.

La utilización de pronósticos automáticos permite encontrar modelos que brindan ajustes estadísticamente aceptables, es decir, con un nivel de confianza adecuado. El desarrollo y modelamiento más específico del fenómeno, con técnicas más complejas, tales como, redes neuronales, algoritmos evolutivos, etc., para las cuales, el resultado obtenido por el sistema híbrido difuso, pueden considerarse como un punto más de comparación. Es así, que la utilización de técnicas para modelar fenómenos es una

puerta abierta para cada aplicación específica.

En el ámbito de los negocios es importante el uso de varios escenarios, los cuales permiten estimar el comportamiento del fenómeno. Este análisis específico determina la decisión más acertada. En este contexto no se debe dejar de lado el criterio de los expertos, ya que su aporte es imprescindible para la comprensión de fenómenos, y en este sentido, la inclusión de los criterios de expertos constituye una gran ventaja del método propuesto.

Para un mejor desempeño del método implementado es importante entender el comportamiento de las series y de esta manera poder explicar la presencia de datos inusuales generados por los residuos; los cuales afectan directamente a los criterios de ajuste, y permiten elegir el mejor modelo.

El uso de R como herramienta de implementación brinda una gran ventaja en lo que respecta: al manejo de datos, la visualización de resultados y la implementación en sí. Puesto que R dispone de funciones que incorporan análisis específicos, y que son utilizados por la comunidad científica, los cuales facilitan el proceso.

El uso de *scrips* deja abierta la oportunidad de seguir manejando el método en muchas maneras, como por ejemplo: la incorporación de la modelación de la varianza, o la posibilidad de definir nuevos criterios de inferencia, los cuales permitirán obtener estimaciones más robustas.

## BIBLIOGRAFÍA

- [1] Rob J. Hyndman and Yeasmin Khandakar (2008). *Automatic Time Series Forecasting: The forecast Package for R*. Journal of Statistical Software, Volume 27, Issue 3.
- [2] E. J. Hannan and J. Rissanen (1982). *Recursive estimation of mixed autoregressive-moving average order*. Biometrika 69. Great Britain.
- [3] M. Theodosiou and Muraliv Swamy (2006). *A Hybrid forecasting approach: Structural Decomposition, generalized regression neural networks and Theta method*.
- [4] Ajoy K. Palit and Dobrivoje Popovic (2005). *Computational intelligence in time series forecasting*. Springer-Verlag, London.
- [5] Timothy J. Ross (2004). *Fuzzy logic with engineering applications*, Second edition. Wiley, England.
- [6] Willi-Hans Steeb (2005). *The nonlinear workbook*, Third edition. World Scientific Publishing, USA.
- [7] Kevin M. Passino (1998). *Fuzzy control*. Addison Wesley Longman, USA.
- [8] Andries P. Engelbrecht (2002). *Computational intelligence*. Wiley, England.
- [9] Hung T. Nguyen and Berlin Wu (2006). *Fundamentals of statistics with fuzzy data*. Springer-Verlag, Berlin Heidelberg.
- [10] Frank Höpper, Frank Klawonn, Rudolf Kruse and Thomas Runkler (2000). *Fuzzy cluster analysis*. Wiley, England.
- [11] Boris Kovalerchuk and Evgenii Vityaev (2002). *Data mining in finance*. Kluwer Academic Publishers, USA.

- [12] Pavel Čížek, Wolfgang Härdle and Rafal Weron (2005). *Statistical tools for finance and insurance*. Springer-Verlag, Germany.
- [13] Saman K. Halgamuge and Lipo Wang (2005). *Computational Intelligence for Modelling and Prediction*. Springer-Verlag, Netherlands.
- [14] . Zimmermann H.-J. (1993). *Fuzzy Set Theory - and Its Applications*, Second, Revised Edition. Kluwer Academic Publishers, USA.
- [15] Robert H. Shumway and David S. Stoffer (2006). *Time series analysis and its applications with R examples*, Second edition. Springer-Verlag, USA.
- [16] Jonathan D. Cryer and Kung-Sik Chan (2008). *Time series analysis with applications in R*, Second edition. Springer-Verlag, USA.
- [17] Ruey Tsay (2002). *Analysis of financial time series*. Wiley, USA.
- [18] Pedro A. Morettin (2002). *Séries temporais em finanças*. Universidad de São Paulo, Brasil. Texto para um Curso na Universidad Católica de Lima, Peru.
- [19] Alfonso Novales (2000). *Econometría*, Segunda edición. McGraw Hill, España.
- [20] Chris Chatfield (2005). *The analysis of time-series: an introduction*, Sixth edition. CH/CRC, USA.
- [21] Christophe Planas (1997). *Applied Time Series Analysis: Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filter*. Eurostat, Luxembourg.
- [22] Genshiro Kitagawa (2010). *Introduction to Time Series Modeling*. CH/CRC, USA.
- [23] Douglas C. Montgomery, Cheryl L. Jennings and Murat Kulahci. *Introduction to Time Series Analysis and Forecasting*. Wiley, Canada.
- [24] Michael J. Crawley (2007). *The R book*. Wiley, England.

- [25] René Carmona (2004). *Statistical analysis of financial data in S-Plus*, Springer-Verlag, New York.
- [26] Simon J. Sheather (2009). *A modern approach to regression with R*. Springer-Verlag, USA.
- [27] R News. The Newsletter of the R Project Volume 2/2, June 2002.
- [28] <http://www.r-project.org/>, web R Project.
- [29] J.P.Morgan-Reuters (1996). *RiskMetrics-Technical Document*, Fourth Edition, J.P. Morgan and Reuters, New York.
- [30] Patricio Fuenmayor y Hermann Mena (2009). *Inteligencia computacional en la modelación de series financieras: enfoque de la lógica difusa*, memorias I Congreso científico internacional en economía y finanzas.
- [31] Patricio Fuenmayor y Hermann Mena (2011). *Fuzzy Hybrid System for Forecasting Time Series*, Preprint.
- [32] Patricio Fuenmayor y Hermann Mena, Pronósticos de series financieras mediante un sistema híbrido difuso - SCRIPT shd para R, <http://www.math.epn.edu.ec/hmena/>, sección software.

# ANEXO 1

SISTEMA HÍBRIDO DIFUSO (shd)

PATRICIO FUENMAYOR VITERI

FECHA INICIO 2010-08-01

El script esta disponible como código abierto en [32]

Script shd para R

El script shd para R contiene la metodología y herramientas para la obtención de pronósticos de series financieras mediante un sistema híbrido difuso. La modelación de las series se basa en modelos lineales SARIMA, se aplica agrupamiento difuso en los pronósticos generados para cada modelo obteniendo de esta forma los conjuntos en los cuales se realizan inferencias para la obtención del pronóstico óptimo medio.

```
# -----
# FUNCION: shd_mdl
# genera lista de objetos resultantes de la modelación automática: matriz de modelos y mejores modelos en
# base a los criterios de información (CI)
# PARAMETROS:
# dat: datos o serie financiera
# mdlexc: modelos excluidos (archivo dat\mdlexcxxx.txt)
# quaprb: probabilidad para seleccionar los modelos con ic menor al quantile(probs=quaprb) 0=min,
# i=max (default)
# trace: para imprimir en consola los modelos testeados. Por defecto FALSE
# fil: nombre del archivo para guardar resultados (por defecto rstmdl.Rdata)
# SALIDA (lstmdl):
# mtzmdl: matriz de modelos
# mdl_aic: mejor modelo en base a aic
# mdl_bic: mejor modelo en base a bic
# mdl_aicc: mejor modelo en base a aicc

shd_mdl <- function(dat,mdlexc,quaprb=1,trace=FALSE,fil="rstmdl.Rdata")
{
  setwd(dirwrk)
  # GENERAR MODELOS
  mdl_aic <- auto.arima2(dat,ic="aic",trace=trace,file="rst_aic.Rdata")
  mdl_bic <- auto.arima2(dat,ic="bic",trace=trace,file="rst_bic.Rdata")
  mdl_aicc <- auto.arima2(dat,ic="aicc",trace=trace,file="rst_aicc.Rdata")
}
```

```

# CARGAR RESULTADOS
load("rst_aic.Rdata")
rst_aic <- with(rst,data.frame(rst,icn="aic",mdl=paste(p,d,q,P,D,Q,c,sep="")))
rst_aic <- rst_aic[rst_aic$ic<1e20,]
load("rst_bic.Rdata")
rst_bic <- with(rst,data.frame(rst,icn="bic",mdl=paste(p,d,q,P,D,Q,c,sep="")))
rst_bic <- rst_bic[rst_bic$ic<1e20,]
load("rst_aicc.Rdata")
rst_aicc <- with(rst,data.frame(rst,icn="aicc",mdl=paste(p,d,q,P,D,Q,c,sep="")))
rst_aicc <- rst_aicc[rst_aicc$ic<1e20,]
rst_aic <- rst_aic[rst_aic$ic <= quantile(rst_aic$ic,probs=quaprb),]
rst_bic <- rst_bic[rst_bic$ic <= quantile(rst_bic$ic,probs=quaprb),]
rst_aicc <- rst_aicc[rst_aicc$ic <= quantile(rst_aicc$ic,probs=quaprb),]
mtzmdl <- rbind(rst_aic,rst_bic,rst_aicc)
mtzmdl <- mtzmdl[,c(-8,-9)]
mtzmdl <- sqldf("select distinct * from mtzmdl")
mtzmdl <- sqldf("select * from mtzmdl where mdl not in (select t01.mdl from mdlexc t01)")
names(mtzmdl) <- c("p","d","q","P","D","Q","c","mdl")
# GUARDAR MATRIZ DE MODELOS
save(mtzmdl,file=fil)
return(list("mtzmdl"=mtzmdl,"mdl_aic"=mdl_aic,"mdl_bic"=mdl_bic,"mdl_aicc"=mdl_aicc))
}
#-----
# FUNCION: shd_prn
# genera lista con matrices de: ajustes, residuos y pronósticos. A más de esto, guarda los resultados
# del mejor modelo obtenido en base al AIC
# PARAMETROS:
# lstmdl: lista de modelación
# hrzprn: horizonte de pronóstico (h en la función forecast)
# nivtv: nivel de confianza de los intervalos en la función forecast
# nivsel: nivel de selección de los modelos (valor del aic max)
# fil: nombre del archivo para guardar resultados. Por defecto (rstprn.Rdata)
# SALIDA (lstprn)
# mtzaju: matriz de ajustes
# mtzrsd: matriz de residuos
# mtzrel: matriz de residuos relativos
# mtzprn: matriz de pronósticos
# mtzaic: matriz de valores de AIC
# mtzajumej: matriz de ajuste del mejor modelo
# mtzrsdmej: matriz de residuos del mejor modelo
# mtzrelmej: matriz de residuos relativos del mejor modelo
# mtzprnmej: matriz de pronósticos del mejor modelo

```

```

# dat: datos o serie financiera
# mtzmdl: matriz de modelos

shd_prn <- function(lstmdl,hrzprn=12,nivitv=63,nivsel=1e20,fil="rstprn.Rdata")
{
  setwd(dirwrk)
  mtzmdl <- lstmdl$mtzmdl
  dat <- lstmdl$mdl_aic$x
  # DIMENSIONES DE LAS MATRICES
  nummdl <- dim(mtzmdl)[1]
  numdat <- length(dat)
  prd <- frequency(dat)
  lstprn <- list(NULL)

  mtzaju <- matrix(NA,nrow=numdat,ncol=nummdl) # matriz ajustes
  mtzrsd <- mtzaju # matriz residuos
  mtzrel <- mtzaju # matriz residuos relativos
  mtzprn <- matrix(NA,nrow=hrzprn,ncol=nummdl) # matriz pronósticos
  mtzaic <- matrix(NA,nrow=1,ncol=nummdl) # matriz criterio de información AIC

  for(i in 1:nummdl){
    # CARGAR PARAMETROS DEL MODELO
    mdlpar <- mtzmdl[i,]
    mdl <- with(mdlpar,Arima(dat,order=c(p,d,q),seasonal=list(order=c(P,D,Q),period=prd),include.drift= c=1))
    lstprn[[i]] <- forecast(mdl,h=hrzprn,level=nivitv)
    if(lstprn[[i]]$model$aic <= nivsel)
    {
      mtzaju[,i] <- lstprn[[i]]$fit
      mtzrsd[,i] <- lstprn[[i]]$residuals
      mtzrel[,i] <- lstprn[[i]]$residuals/lstprn[[i]]$x
      mtzprn[,i] <- lstprn[[i]]$mean
      mtzaic[,i] <- lstprn[[i]]$model$aic
    }
  }

  # ELIMINA COLUMNAS CON VALORES NA
  mtzaju <- clean.na(mtzaju,margin=2)
  mtzrsd <- clean.na(mtzrsd,margin=2)
  mtzrel <- clean.na(mtzrel,margin=2)
  mtzprn <- clean.na(mtzprn,margin=2)
  mtzaic <- clean.na(mtzaic,margin=2)

  # IDENTIFICACIÓN POR AIC
  col01 <- sort(brewer.pal(3,"Blues"),decreasing=F) # display.brewer.pal(3, "Blues")
  minaic <- min(mtzaic)

```

```

maxaic <- max(mtzaic)
itvaic <- (maxaic-minaic)/3
mtzaic<- t(mtzaic)
mtzaic <- data.frame("aic"=mtzaic,"col"=cut(mtzaic,include.lowest=T,
      breaks=seq(from=minaic,to=maxaic,by=itvaic), right=F,labels=c(col01)))
# CARGAR PARAMETROS DEL MEJOR MODELO SEGUN AIC
prnmej <- forecast(lstmdl$mdl_aic,h=hrzprn,level=nivitv)
mtzajumej <- prnmej$fit
mtzrsdmej <- prnmej$residuals
mtzrelmej <- prnmej$residuals/prnmej$x
mtzprnmej <- prnmej$mean
# GUARDAR LISTA DE PRONOSTICOS
save(lstprn,file=fil)
return(list("mtzaju"=mtzaju,"mtzrsd"=mtzrsd,"mtzrel"=mtzrel,"mtzprn"=mtzprn,"mtzaic"=mtzaic,
      "prnmej"=prnmej,"mtzajumej"=mtzajumej,"mtzrsdmej"=mtzrsdmej,"mtzrelmej"=mtzrelmej,
      "mtzprnmej"=mtzprnmej,"dat"=dat,"mtzmdl"=mtzmdl))
}
#-----
# FUNCION: shd_cdp
# genera agrupamientos difusos de los pronósticos
# PARAMETROS:
# lstprn: lista de ajustes y pronósticos
# tipifr: tipo de inferencia (max = maximo, ptc = participación)
# fil: nombre del archivo para guardar resultados. Por defecto (rstcdp.Rdata)
# SALIDA (lstcdp):
# lstclu: lista de resultados del algoritmo c-medias
# mtzclu: matriz de clusters de los pronósticos
# mtzcen: matriz de centros
# mtzele: matriz de número de elementos por grupo
# mtzifr: matriz resultante de la inferencia

shd_cdp <- function(lstprn,tipifr="ptc",fil="rstcdp.Rdata"){
  setwd(dirwrk)
  mtzprn <- lstprn$mtzprn
  numprn <- dim(mtzprn)[1]
  nummdl <- dim(mtzprn)[2]
  lstclu <- list(NULL)
  mtzclu <- matrix(NA,nrow=nummdl,ncol=numprn) # matriz clusterns
  mtzcen <- matrix(NA,nrow=3,ncol=numprn) # matriz centros
  mtzele <- mtzcen # matriz elementos
  mtzifr01 <- matrix(NA,nrow=1,ncol=numprn) # matriz inferencia

```

```

for(i in 1:numprn){
# CREAR CLUSTERS PARA CADA VALOR DE PRONOSTICO, SI NO SE PUEDE REALIZAR EL CLUSTER, GRABA EL VALOR PROMEDIO
if (sd(mtzprn[i,]) <= 5)
lstclu[[i]] <- list("cluster"=rep(c(1,2,3),length.out=nummdl),"centers"=rep(mean(mtzprn[i,]),3))
else
lstclu[[i]] <- cmeans(mtzprn[i,],3,1000,verbose=F,method="cmeans")
mtzclu[,i] <- lstclu[[i]]$cluster
mtzcen[,i] <- lstclu[[i]]$centers
mtzele[,i] <- table(mtzclu[,i])
}
# TIPOS DE INFERENCIA (máximo o participación)
# CRITERIO DE PARTICIPACIÓN
if(tipifr=="max")
{
posmax <- max.col(t(mtzele))
for(i in 1:numprn){
mtzifr01[i] <- mtzcen[posmax[i],i]
}
}
else
{
mtzptc01 <- prop.table(mtzele,margin=2)
mtzifr01 <- apply(mtzcen*mtzptc01,2,sum)
}
# PONDERA POR EL NIVEL DE AJUSTE DEL MODELO
dumcol <- data.frame(dummy(lstprn$mtzaic$col))
numcol <- dim(dumcol)[2] #variable reutilizada en el for
names(dumcol) <- paste("C",1:numcol,sep="")
mtzcol <- data.frame(mtzclu)
mtzcol <- data.frame(mtzcol,dumcol)
mtzacu <- matrix(NA,nrow=3,ncol=numprn)
nomvar <- as.vector(names(mtzcol))
for(i in 1:numprn){
var <- nomvar[i]
numcol <- fn$sqldf("select $var, sum(C1) as 'C' from mtzcol group by $var")
mtzacu[,i] <- numcol[,2]
}
mtzptc02 <- prop.table(mtzacu,margin=2)
mtzifr02 <- apply(mtzcen*mtzptc02,2,sum)
mtzifr <- rbind(mtzifr01,mtzifr02)
mtzifr <- apply(mtzifr,2,mean)
# GUARDAR RESULTADOS DE CLUSTERS DIFUSOS DE PRONOSTICOS

```

```

rstcdp <- list("lstclu"=lstclu,"mtzclu"=mtzclu,"mtzcen"=mtzcen,"mtzele"=mtzele,"mtzifr"=mtzifr)
save(rstcdp,file=fil)
return(rstcdp)
}
#-----
# FUNCION: shd_grf
# genera gráficas
# PARAMETROS:
# lstprn: salida de shd_prn
# lstcdp: salida de shd_cdp
# tst: datos para test
# pre: prefijo para nombre de gráficas
# uni: unidades (leyendas)
# SALIDAS:
# _dat.pdf: gráfica de la serie de datos y de las funciones ACF y PACF
# _datdif.pdf: gráfica de la serie de datos diferenciada
# _lag.pdf: gráfica de los retardos de la serie
# _prn.pdf: gráfica de los pronósticos del mejor modelo
# _rsd.pdf: gráfica de los residuos del mejor modelo
# _dcpstl.pdf: gráfica de descomposición stl (método loess)
# _dcpmul.pdf: gráfica de descomposición multiplicativa
# _dcpadd.pdf: gráfica de descomposición aditiva
# _dcpmul_s.pdf: gráfica del componente estacional multiplicativo
# _dcpadd_s.pdf: gráfica del componente estacional aditivo
# _dcpadd_sr.pdf: gráfica del componente estacional y del residuo (aditivos)
# _mtzaju.pdf: gráfica de la matriz de ajustes
# _mtzrsd.pdf: gráfica de la matriz de residuos
# _mtzifr.pdf: gráfica de la matriz de pronósticos e inferencia
# _mtztst.pdf: gráfica de la matriz de inferencia con pruebas de pronósticos e identificadores
de los modelos
# _mtzdat.pdf: gráfica de la matriz de datos conjuntos (dat + prn + inf)
# _boxrsd.pdf: gráfica en boxplot de los residuos
# _boxrel.pdf: gráfica en boxplot de los residuos relativos
# _disrsd.pdf: gráfica de dispersión de los residuos y residuos relativos por modelo

shd_grf <- function(lstprn,lstcdp,tst,pre="",uni="")
{
mtzdat <- rbind(lstprn$mtzaju,lstprn$mtzprn)
mtzdatmej <- rbind(as.matrix(lstprn$mtzajumej),as.matrix(lstprn$mtzprnmej))
numdat <- length(lstprn$dat)
numprn <- numdat + 1
hrzprn <- dim(lstprn$mtzprn)[1]

```

```

numtot <- length(mtzdatmej)
col01 <- as.character(lstprn$mtzaic$col)
nummdl <- dim(lstprn$mtzprn)[2]
ordmdl <- 1:nummdl
low <- lstcdp$mtzifr*(1-(0.03*sqrt(1:hrzprn)))
hgh <- lstcdp$mtzifr*(1+(0.03*sqrt(1:hrzprn)))
labx <- rep(hrzprn,nummdl)
laby <- lstprn$mtzprn[hrzprn,]
miny <- min(min(lstprn$mtzprn),min(tst),min(low)) - 5
maxy <- max(max(lstprn$mtzprn),max(tst),max(hgh)) + 5
# DESCOMPOSICIÓN
dcpadd <- decompose(lstprn$dat,type="additive")
dcpmul <- decompose(lstprn$dat,type="multiplicative")

# GRAFICACIÓN
setwd(diring)
win.graph(width=11,height=7,pointsize=10)

tsdisplay(lstprn$dat,points=F,main="",xlab="tiempo (meses)",ylab=paste("saldos (",uni,"dólares)",sep=""))
dev.copy2pdf(file=paste(pre,"_dat.pdf",sep=""))

tsdisplay(diff(lstprn$dat),points=F,main="",xlab="tiempo (meses)",ylab=paste("saldos (",uni,"dólares)",sep=""))
dev.copy2pdf(file=paste(pre,"_datdif.pdf",sep=""))

lag.plot(lstprn$dat,4)
dev.copy2pdf(file=paste(pre,"_lag.pdf",sep=""))

plot(lstprn$prnmej,ylab=paste("saldos (",uni,"dólares)",sep=""))
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
dev.copy2pdf(file=paste(pre,"_prn.pdf",sep=""))

tsdisplay(lstprn$prnmej$residuals,points=F,main="",xlab="tiempo (meses)",
  ylab=paste("residuos (",uni,"dólares)",sep=""))
dev.copy2pdf(file=paste(pre,"_rsd.pdf",sep=""))

plot(stl(lstprn$dat,s.window=12))
dev.copy2pdf(file=paste(pre,"_dcpstl.pdf",sep=""))

plot(dcpmul)
dev.copy2pdf(file=paste(pre,"_dcpmul.pdf",sep=""))

plot(dcpadd)

```

```

dev.copy2pdf(file=paste(pre, "_dcpadd.pdf", sep=""))

plot(dcpmul$figure,type="l",xlab="tiempo (meses)",ylab="estacionalidad (factores)")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
dev.copy2pdf(file=paste(pre, "_dcpmul_s.pdf", sep=""))

plot(dcpadd$figure,type="l",xlab="tiempo (meses)",ylab=paste("estacionalidad (" ,uni,"dólares)", sep=""))
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
dev.copy2pdf(file=paste(pre, "_dcpadd_s.pdf", sep=""))

plot(dcpadd$random,type="l",col="blue",xlab="tiempo (meses)",
      ylab=paste("estacionalidad y residuos (" ,uni,"dólares)", sep=""))
lines(dcpadd$seasonal,col="red")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("estacionalidad","residuos"),fill=c("red","blue"),xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre, "_dcpadd_sr.pdf", sep=""))

plot(1:numdat,lstprn$dat,type="l",lty=1,col="red",xlab="tiempo (meses)",
      ylab=paste("saldos (" ,uni,"dólares)", sep=""))
matpoints(lstprn$mtzaju,type="l",lty=1,col=col01)
points(1:numdat,lstprn$mtzajumej,type="l",lty=1,col="green")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("real","mejor","otros"),fill=c("red","green","steelblue2"),
      xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre, "_mtzaju.pdf", sep=""))
#-----
layout(matrix(c(1,2), byrow=T, ncol=1))
matplot(lstprn$mtzrsd,type="l",lty=1,col=col01,main="Residuos",xlab="tiempo (meses)",
        ylab=paste(uni,"dólares", sep=""))
points(1:numdat,lstprn$mtzrsdmej,type="l",lty=1,col="green")
abline(a=0,b=0,col="black")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("mejor","otros"),fill=c("green","steelblue2"),xjust=0.15,yjust=1,cex=0.8)

matplot(lstprn$mtzrel*100,type="l",lty=1,col=col01,main="Residuos relativos", xlab="tiempo (meses)",
        ylab="porcentaje")
points(1:numdat,lstprn$mtzrelmej*100,type="l",lty=1,col="green")
abline(a=0,b=0,col="black")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("mejor","otros"),fill=c("green","steelblue2"),xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre, "_mtzrsd.pdf", sep=""))
graphics.off()

```

```

#-----
win.graph(width=11, height=7,pointsize=10)
layout(matrix(c(1, 2), byrow=F, ncol=2))
matplot(lstprn$mtzprn,type="l",lty=1,col=col01,xlab="tiempo (meses)",
  ylab=paste("saldos (",uni,"dólares)",sep=""))
points(1:hrzprn,lstprn$mtzprnmej,type="l",lty=1,lwd=2,col="green")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("mejor","otros"),fill=c("green","steelblue2"),xjust=0.15,yjust=1,cex=0.8)

matplot(lstprn$mtzprn,type="l",lty=1,col=col01,xlab="tiempo (meses)",
  ylab=paste("saldos (",uni,"dólares)",sep=""))
points(1:hrzprn,lstcdp$mtzifzr,type="l",lty=1,lwd=2,col="black")
points(1:hrzprn,lstprn$mtzprnmej,type="l",lty=1,lwd=2,col="green")
matpoints(t(lstcdp$mtzcen),type="p",lty=1,col="red",cex=0.8)
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("shd","mejor","otros"),fill=c("black","green","steelblue2"),
  xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre,"_mtzifzr.pdf",sep=""))
graphics.off()
#-----
win.graph(width=11, height=7,pointsize=10)
matplot(lstprn$mtzprn,type="n",ylim=c(miny,maxy),xlab="tiempo (meses)",
  ylab=paste("saldos (",uni,"dólares)",sep=""))
polygon(c(1:hrzprn,rev(1:hrzprn)),c(low,rev(hgh)),col="grey95",border=F,fillOddEven=T)
matpoints(lstprn$mtzprn,type="l",lty=1,col=col01)
points(1:hrzprn,lstcdp$mtzifzr,type="l",lty=1,lwd=2,col="black")
points(1:hrzprn,lstprn$mtzprnmej,type="l",lty=1,lwd=2,col="green")
points(tst,type="l",lty=1,lwd=2,col="red")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
text(labx,laby,ordmdl,cex=0.5)
legend("bottomright",legend=c("real","shd","mejor","otros","zona conf."),
  fill=c("red","black","green","steelblue2","grey95"),xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre,"_mtztst.pdf",sep=""))
#-----
matplot(mtzdat,type="l",lty=1,col=col01,xlab="tiempo (meses)",
  ylab=paste("saldos (",uni,"dólares)",sep=""))
points(1:numtot,mtzdatmej,type="l",lty=1,col="green")
points(1:numdat,lstprn$dat,type="l",lty=1,col="red",xlab=NA)
points(numprn:numtot,lstcdp$mtzifzr,type="l",lty=1,col="black",xlab=NA)
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
abline(v=numprn,col="black")
legend("bottomright",legend=c("real","shd","mejor","otros"),fill=c("red","black","green","steelblue2"),

```

```

    xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre,"_mtzdat.pdf",sep=""))
#-----
boxplot(t(lstprn$mtzrsd),outpch=".",xlab="tiempo (meses)",
  ylab=paste("residuos (",uni,"dólares)",sep=""))
lines(1:numdat,lstprn$mtzrsdmej,type="l",lty=1,col="green")
lines(1:numdat,t(apply(lstprn$mtzrsd,1,median)),type="l",lty=1,col="orange")
abline(a=0,b=0,col="black")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("mejor","otros (mediana)"),fill=c("green","orange"),
  xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre,"_boxrsd.pdf",sep=""))
#-----
boxplot(t(lstprn$mtzrel*100),outpch=".",xlab="tiempo (meses)",ylab="residuos (porcentaje)")
lines(1:numdat,lstprn$mtzrelmej*100,type="l",lty=1,col="green")
lines(1:numdat,t(apply(lstprn$mtzrel*100,1,median)),type="l",lty=1,col="orange")
abline(a=0,b=0,col="black")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
legend("bottomright",legend=c("mejor","otros (mediana)"),fill=c("green","orange"),
  xjust=0.15,yjust=1,cex=0.8)
dev.copy2pdf(file=paste(pre,"_boxrel.pdf",sep=""))
#-----
layout(matrix(c(1, 2), byrow=F, ncol=2))
boxplot(lstprn$mtzrsd,horizontal = T,outpch=20,names=lstprn$mtzmdl$mdl,
  xlab=paste("residuos (",uni,"dólares)",sep=""),ylab=NA)
abline(a=0,b=0,col="black")
abline(v=0,col="black")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)

boxplot(lstprn$mtzrel*100,horizontal = T,outpch=20,names=lstprn$mtzmdl$mdl,
  xlab="residuos relativos (porcentaje)",ylab=NA)
abline(a=0,b=0,col="black")
abline(v=0,col="black")
grid(nx=NULL,ny=NULL,col="lightgray",lty="dotted",lwd=par("lwd"),equilogs=TRUE)
dev.copy2pdf(file=paste(pre,"_disrsd.pdf",sep=""))

graphics.off()
setwd(dirwrk)
}

```