



ÍNDICE DE COMOVIMIENTO ENTRE SERIES TEMPORALES: UNA APLICACIÓN

Por
Juan Carlos Herrera Órdenes.

REQUIRIMIENTO PARA EL TÍTULO DE
INGENIERÍA EN ESTADÍSTICA
EN
UNIVERSIDAD DE VALPARAÍSO
VALPARAÍSO, CHILE
JUEVES 30 DE ABRIL 2009

© Escrito por Juan Carlos Herrera Órdenes. 2009

UNIVERSIDAD DE VALPARAÍSO
DEPARTAMENTO DE ESTADÍSTICA

Por medio de esto certifico que han leído y recomiendan a la Facultad de de Ciencias para la aceptación de mi Proyecto de Título en el cumplimiento parcial **“Índice de comovimiento entre series temporales: Una Aplicación”** por **Juan Carlos Herrera Órdenes**. los requisitos para el Título de **Ingeniería en Estadística**.

Fecha: Jueves 30 de Abril 2009

Supervisor del Proyecto de Título: _____
Dr. Ronny Vallejos

Comité Examinador: _____
Dr. Ronny Vallejos

Dr. Manuel Galea

Dr. Felipe Osorio

UNIVERSIDAD DE VALPARAÍSO

Fecha: **Jueves 30 de Abril 2009**

Autor: **Juan Carlos Herrera Órdenes.**

Título: **Índice de comovimiento entre series temporales: Una
Aplicación**

Departamento: **de Estadística**

Grado: **Ingeniería en Estadística** Convocación: **Jueves 30 de Abril** Año: **2009**

Universidad de Valparaíso

Firma del Autor

EL AUTOR SE RESERVA EL DERECHO DE AUTOR

Este trabajo esta dedicado a mis Padres Juan Carlos Herrera y Sara Elizabeth Órdenes por su ayuda en sabiduría e inteligencia durante mi período de estudio. Por iluminarme en los momentos difíciles y su apoyo incondicional.

Tabla de Contenidos

Tabla de Contenidos	v
Resumen	vii
Agradecimiento	viii
Introducción	1
1. Teoría y Fundamentos	4
1.1. Definición de Procesos Estocásticos	4
1.2. Ruido blanco	5
1.3. Procesos Autoregresivos	5
1.4. Procesos de Media Móvil	6
1.5. Procesos ARMA	6
1.6. Coeficientes de Codispersión para series temporales	7
1.7. Algunos ejemplos	9
1.7.1. Comovimiento	9
1.7.2. Anti comovimiento	10
1.7.3. Comovimiento nulo	11
1.8. Propiedades	11
1.9. Limitaciones teóricas	15
1.10. Métodos no paramétricos	15
2. Agrupamiento de series temporales	16
2.1. Introducción	16
2.2. Métodos de Agrupación	16
2.3. Índice de Disimilaridad Adaptativo	17
2.3.1. Distancia Euclidiana	18
2.3.2. Distancia de Minkowski	18
2.3.3. Distancia de Frechét	18
2.4. Alineamiento de tiempo distorsionado (Dynamic time warping DTW)	19
2.4.1. Formulación del problema	19
2.4.2. Hacia el Alineamiento de tiempo distorsionado	20

2.4.3.	Definición DTW	21
2.4.4.	Ejemplo	22
2.5.	Índice de Disimilaridad Adaptativo para medidas de proximidad entre series de tiempo.	23
2.5.1.	Correlación temporal para medidas de proximidad	24
2.5.2.	Índice de Disimilaridad Adaptativo entre series de tiempo	25
3.	Simulación	27
3.1.	Introducción	27
3.2.	Simulación entre series correlacionadas	27
3.2.1.	Gráficos Camino Distorsionado	30
3.2.2.	Índice de Disimilaridad Adaptativo con Distancia Euclidiana	31
3.2.3.	Índice de Disimilaridad Adaptativo con Distancia DTW	31
3.2.4.	Índice de Disimilaridad Adaptativo con Distancia de Frechét	32
3.3.	Conclusión	34
4.	Aplicación a datos de AFP	35
4.1.	Introducción	35
4.2.	Sistema de AFP	35
4.3.	Rentabilidad de los Fondos de Pensiones	36
4.4.	Contribución al Desarrollo Económico	36
4.5.	Antiguo Sistema Previsional	37
4.6.	Efecto Manada	38
4.6.1.	Contexto y consecuencias de la competencia en rentabilidad vía ranking . . .	38
4.7.	Acciones favoritas de las AFP	39
4.7.1.	El método de Encosult	39
4.8.	Análisis Descriptivo	40
4.9.	Aplicación Índice de Comovimiento o Codispersión	41
4.10.	Modelamiento	43
4.11.	Estimación de parámetros MA(1) y análisis de residuos	45
4.12.	Análisis de Cluster	48
4.12.1.	Camino Distorsionado entre AFP	49
4.12.2.	Aplicación Índice de Disimilaridad con Distancia Euclidiana	50
4.12.3.	Aplicación Índice de Disimilaridad con Distancia Frechét	51
4.12.4.	Aplicación Índice de Disimilaridad con Distancia DTW	51
4.13.	Conclusión	52
5.	Conclusiones	53
	Bibliografía	55
	Anexo	56
	Apéndice	71

Resumen

Para dos series temporales el Índice de Comovimiento es una medida de similaridad que contiene información temporal entre ellas. Este índice es también llamado Coeficiente de Codispersión, el cual, es una adecuada normalización de suma de productos internos para dos secuencias temporales.

De acuerdo a su definición, dos series comueven (o se mueven conjuntamente), si sus conjuntos respectivos de pendientes son proporcionales entre sí. Este índice tiene como característica estar acotado entre 1 y -1 . Si el coeficiente o índice es cercano a 1, se dice que las dos series se mueven juntas (comueven), en cualquier intervalo de tiempo $[t_i, t_{i+h}]$, donde h es el retardo del índice en el proceso. Un coeficiente cercano a -1 , se interpreta como un anti comovimiento. Ahora cuando este índice es cercano 0 se puede decir que no hay comovimiento entre las series, es decir, no hay relación entre las pendientes de las series en instantes sucesivos.

Este trabajo se enfoca en dos puntos esenciales. Primero, representar varias situaciones con modelos paramétricos asociados a esta medida. Segundo, aplicar un algoritmo de clasificación para series temporales basado en una medida de asociación que contiene el Índice de Comovimiento, llamado Índice de Disimilaridad Adaptativo.

El Índice de Disimilaridad Adaptativo, es un producto entre dos funciones, que contiene una función de afinación de balance entre el comportamiento respecto del comovimiento entre las series temporales y la cercanía de los valores basados en distancias convencionales. De esta manera se introduce una medida alternativa para la clasificación de series temporales utilizando los algoritmos clásicos de clasificación como son, por ejemplo, el método de agrupación jerárquico.

Ahora bien, estas medidas se aplicarán a 7 AFP del sistema de pensiones Chileno que ha sido exportado a otros países, cuya funcionalidad es velar por el ahorro de los trabajadores para tener una futura pensión al momento de su jubilación. La característica principal de las AFP es su rentabilidad, existen empresas especializadas para lograr la rentabilidad de los ahorros de los chilenos. No obstante, debido a la facilidad de información de los mercados bursátiles, las AFP buscan oportunidades en ella. Esto ha creado que la competencia de las empresas de AFP, no presente grandes variabilidades respecto a las otras AFP, lo que comúnmente se llama fenómeno manada.

Los resultados que se presentan en este trabajo, son la aplicación de estas medidas a estas 7 AFP. Para así, agruparlas considerando su información de comovimiento y comportamiento respecto a su cercanía simultáneamente.

Agradecimiento

Al apoyo incondicional de mis Padres, Familia y novia.
A mi profesor guía Dr. Ronny Vallejos por su apoyo y dedicación más allá de su labor académica.
A todos quienes colaboraron con este trabajo, profesores que me brindaron parte de su tiempo y paciencia.

Muchas gracias.

Introducción

Ha existido gran interés por estudiar el comportamiento de una variable aleatoria a través del tiempo como es el caso en series de tiempo, en que una función del pasado es predicha en el futuro.

Las aplicaciones de series de tiempo o series temporales es muy amplia, esto se puede apreciar en la Economía, Medicina, Meteorología, etc. Pero es natural estudiar, como se comportan dos o más series de tiempo.

En la actualidad existen modelos multivariados para respresentar esta situación, por ejemplo: Peña (2001) realiza simulaciones de Monte Carlo, donde compara modelos ARIMA y VARMA, y muestra que la dependencia entre las componentes de un vector de series de tiempo, hace crecer la precisión de los pronósticos multivariados respecto a los univariados. También, existe el coeficiente de correlación espúrea, Karl Pearson (1897) que dice que un alto coeficiente de correlación entre dos variables es espúreo si este se explica por la presencia de un tercer factor y no debido a la existencia de una relación con sentido entre las variables analizadas. En este caso, la correlación estadísticamente significativa entre las variables es una correlación espúrea o sin sentido, por nombrar algunos.

La motivación de este Proyecto de Titulación, es estudiar el comportamiento de las series temporales (Procesos estocásticos) $\{X_t\}$ e $\{Y_t\}$ asociado al Coeficiente de Codispersión o Índice comovimiento introducido por (Rukhin y Vallejos, 2008). En esencia este coeficiente se implementó para procesos espaciales autoregresivos y de media móvil intrínsecamente estacionarios. En este Proyecto de Titulación se particularizará a modelos y/o procesos unilaterales autoregresivos, de media móvil y algunos modelos ARMA. El Índice de Comovimiento tiene la ventaja de captar el comportamiento de dos series temporales y es una versión corregida del clásico coeficiente de correlación. Este Índice compara proporcionalmente las pendientes en común de pares de puntos, a través, del tiempo.

A modo de contraejemplo, consideremos la covarianza muestral de dos variables en estudio

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Donde este es un estimador crudo, que depende de la suma de los productos cruzados. La covarianza muestral permite identificar la dirección o sentido de la relación lineal entre variables, a

través de su signo y esto nos permite establecer en que cuadrante se encuentran los datos. Esta es la única información relevante que proporciona la covarianza muestral.

En la literatura se pueden encontrar otras medidas de asociación, un ejemplo es el Coeficiente de correlación de Spearman, como una versión no paramétrica. Sin embargo, este coeficiente no contiene información sobre el comportamiento temporal entre las series, más bien, está orientada a proveer la independencia de dos series (Ver Yong y Schreckengost, 1981).

El proyecto de título se trabajo se desarrollará de la siguiente forma, en el Capítulo I se harán definiciones formales y fundamentos del Índice de Comovimiento, se hará una introducción a esta medida de similaridad, se profundizará más sobre este Coeficiente de Codispersión, se mostraran propiedades, resultados importantes y limitaciones teóricas del mismo . En el Capítulo II, Se hará una reseña de los métodos de agrupación para series temporales introduciendo un Índice de Disimilaridad Adaptativo estudiado por Chuoakria y Nagabhushan (2007). Este índice es una función de balance, que contiene información del comovimiento entre las series temporales y el comportamiento respecto a la distancia, introduciendo así una nueva medida para la clasificación de las series temporales. En el Capítulo III se realizarán simulaciones, para entender y gráficar de manera más clara, las características y uso de este Índice de Disimilaridad Adaptativo.

Por último, en el Capítulo IV se realizará una aplicación a datos reales, del Sistema de Pensiones Chileno ubicados en (www.svs.cl) del año 1990 al 2004. En esta parte se hará una breve introducción al Sistema Chileno de AFP, se hablará de la génesis de este sistema y algo sobre la nueva reforma de Previsión Social, se mencionarán las ventajas, por ejemplo como una forma de ahorro a futuro y desventajas de este sistema de Pensión como el **Efecto Manada**, la unidad de análisis de esta base de datos es la rentabilidad mensual de 7 AFP en estudio y finalmente se aplicará toda la metodología mencionada con sus respectivas conclusiones.

Objetivos del Proyecto

Objetivos Generales

El Coeficiente de Codispersión fue introducido por Matheron en el ao 1965, como una extensión del semivariograma para procesos espaciales intrínsecamente estacionarios.

Los avances de este coeficiente se pueden encontrar en la minería, procesamiento de imágenes y geoestadística entre otras.

En este trabajo se particularizará la teoría a modelos autoregresivos, de media móvil y ARMA, basados en fundamentos matemáticos de probabilidades, Inferencia, Series temporales y Métodos Multivariados, que ayudarán a sustentar este proyecto, para esto es necesario tener medidas o índices que resuman toda esta información en un solo número.

Ahora bien, dependiendo de la perspectiva que se plantee, en la literatura se puede encontrar muchas formas de clasificar y medir la similitud, por ejemplo la distancia Euclidiana. Por otra parte,

Warren Liao (2005), hacen una reseña de varias medidas de asociación y medidas de similaridad para secuencias temporales y algoritmos, para aplicar en Cluster.

Por otra parte, Chouakria y Nagabhushan (2007) proponen un Índice de Disimilaridad Adaptativo para medidas de proximidad en series temporales, la cual se llama *automatic adaptive tuning function*.

Rukhin y Vallejos (2008) introducen un coeficiente de similaridad para secuencias Espaciales y Temporales, donde este coeficiente, es una normalización de suma de incrementos para secuencias de tiempo o espacio.

También, revisaremos algunas medidas más usadas de asociación y similaridad. De la misma forma, se verán algunas definiciones básicas de procesos estocásticos, con algunas hipótesis que sustentan este Proyecto de Titulación, como es la estacionalidad de las series temporales. De la misma forma se planteará la lógica del Coeficiente de Codispersión o Índice de Comovimiento seguido de sus interpretaciones.

Seguidamente, se hará una conjunción entre el Índice de Disimilaridad Adaptativo y el Coeficiente de Codispersión, para así aplicar este Índice de Disimilaridad en algunos métodos de clasificación, con el cual se trabajara para la clasificación de series temporales, o Cluster el cual se aplicará al sistema chileno de AFP.

Objetivos Específicos

1. Estudiar el Índice de Comovimiento ó Coeficiente de Codispersión.
2. Aplicar el coeficiente de codispersión al sistema de AFP Chileno.
3. Implementar un algoritmo de clasificación para un conjunto de series de tiempo simulados y datos reales del sistema de AFP chileno.

Capítulo 1

Teoría y Fundamentos

Para dos o más secuencias temporales, un aspecto importante a considerar, nace de la siguiente pregunta. ¿Cómo resumir esta información de manera simple y compacta? Para esto es necesario tener medidas o índices que resuman toda esta información en un sólo número.

Dependiendo de la perspectiva que se plantee, en la literatura se puede encontrar muchas formas, por ejemplo la distancia Euclidiana. Por otra parte, Warren Liao (2005), hace una reseña de varias medidas de asociación y medidas de similaridad para secuencias temporales y algoritmos, para aplicar en el análisis de Cluster.

Por otra parte, Chouakria y Nagabhushan (2007) proponen un índice de disimilaridad adaptativo para medidas de proximidad en series temporales, la cual se llama *automatic adaptive tuning function*. Rukhin y Vallejos (2008) introducen un coeficiente de similaridad para secuencias Espaciales y Temporales, donde este coeficiente, es una normalización de una suma de incrementos para secuencias de tiempo o espacio.

En este Capítulo revisaremos algunas medidas más usadas de asociación y similaridad. También, se verán algunas definiciones básicas de procesos estocásticos, con algunas hipótesis que sustentan este Proyecto de Titulación, como es la estacionalidad de las series temporales. Planteando también la lógica del Coeficiente de Codispersión y su interpretación.

1.1. Definición de Procesos Estocásticos

El Coeficiente de Codispersión o de Comovimiento tiene la capacidad de captar el comportamiento respecto a si dos series se mueven juntas en el tiempo, en varias ocasiones esto se puede modelar a través de modelos paramétricos que están asociados a esta medida de Codispersión. Primero es necesario definir algunos objetivos y estructuras que nos ayudarán a describir diferentes escenarios. Comenzaremos por la definición de Procesos Estocásticos.

Definición 1.1.1. Sea X_t una función,

$$\begin{aligned} X : \Omega \times T &\longrightarrow \mathbb{R} \\ (\omega, t) &\mapsto X(\omega, t) \end{aligned}$$

tal que para cada $t \in T$, $X_t(\omega)$ es una variable aleatoria. Llamamos al conjunto $\{X_t, t \in T\}$ Proceso Estocástico sobre el espacio Ω .

Una secuencia $\{X_t, t \in T\}$ es fuertemente o estrictamente estacionaria si $\{X_{t_1}, \dots, X_{t_n}\} = \{X_{t_1+h}, \dots, X_{t_n+h}\}$ en distribución, para toda colección t_1, \dots, t_n y para todo $h \in T$. Es decir, el proceso es invariante bajo traslación.

Por otra parte, una secuencia es débilmente, o de segundo orden estacionario si:

1. $\mathbb{E}(X_t) = \mu$.
2. $\mathbb{V}(X_t) = \sigma^2$.
3. $\mathbb{Cov}(X_t, X_{t+k}) = \gamma_k$. La función de autocovarianza es una función que depende únicamente de la distancia entre t y $t+k$.

Seguidamente se define:

$\rho_k = \gamma_k / \gamma_0$ y $\{\rho_k, k \in T\}$ es llamada función de autocorrelación (ACF) del proceso $\{X_t, t \in T\}$.

Como caso particular cuando $T = \mathbb{Z}$, se está en presencia de una serie temporal.

Entonces las series de tiempo son una realización de los procesos estocásticos. Ahora, en esta sección se darán algunas definiciones de modelos y/o procesos autoregresivos, de media móvil y procesos ARMA.

1.2. Ruido blanco

La secuencia ϵ_t , consiste en variables aleatorias independientes con media 0 y varianza σ^2 . ϵ_t es llamada ruido blanco. Esta serie es estacionaria de segundo orden con $\gamma_0 = \sigma^2$ y $\gamma_k = 0, k \neq 0$.

1.3. Procesos Autoregresivos

Sea $\{X_t, t \in \mathbb{Z}\}$ un Proceso Estocástico. Se dice que $\{X_t\}$ es un modelo autoregresivo de orden p denotado por $\text{AR}(p)$, si puede ser escrito por la ecuación:

$$X_t = \sum_{r=1}^p \phi_r X_{t-r} + \epsilon_t, t \in T \quad (1.3.1)$$

donde ϕ_r son los parámetros del modelo y ϵ_t es un ruido blanco independiente de media 0 y varianza σ^2 . Equivalentemente, X_t puede ser denotado por

$$\Phi(B)X_t = \epsilon_t \quad (1.3.2)$$

donde

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \text{ con } B^k X_t = X_{t-k}. \quad (1.3.3)$$

Un proceso AR(p) es estacionario si y solo si, las raíces de $\Phi(B)$, están fuera del disco unitario. En tal caso, el proceso X_t puede ser representado por un modelo de media móvil infinito (MA(∞)).

$$X_t = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j}. \quad (1.3.4)$$

Ejemplo: El proceso AR(1) está definido:

$$X_t = \phi_1 X_{t-1} + \epsilon_t,$$

donde $|\phi| < 1$ garantiza la estacionariedad del proceso.

1.4. Procesos de Media Móvil

Sea $\{X_t, t \in Z\}$ un Proceso Estocástico. Se dice que X_t es un modelo de media móvil de orden q denotado por MA(q) si puede ser descrito por la ecuación:

$$X_t = \sum_{s=0}^q \theta_s \epsilon_{t-s}, \quad (1.4.1)$$

donde los θ_s son los parámetros del modelo y ϵ_t es un ruido blanco independiente con media 0 y varianza σ^2 . Equivalentemente, X_t puede ser denotado por

$$\Theta(B)\epsilon_t = X_t, \quad (1.4.2)$$

donde,

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \text{ con } B^k \epsilon_t = \epsilon_{t-k}. \quad (1.4.3)$$

Si las raíces de $\Theta(B)$, están fuera del disco unitario, entonces el proceso es invertible. En tal caso:

$$\epsilon_t = \sum_{j=0}^{\infty} \Psi'_j X_{t-j}, \text{ es un proceso AR}(\infty). \quad (1.4.4)$$

1.5. Procesos ARMA

Los procesos Autoregresivos de Media Móvil, ARMA(p, q) están definidos por:

$$\sum_{r=0}^p \phi_r X_{t-r} = \sum_{s=0}^q \theta_s \epsilon_{t-s} \Leftrightarrow \Phi X_t = \Theta \epsilon_t. \quad (1.5.1)$$

donde $\phi_0 = 1$, ϵ_t es un ruido blanco. El proceso X_t es estacionario para apropiados valores de ϕ y θ .

Teorema 1.5.1. Sea $\{X_t, t \in Z\}$ un modelo ARMA. $\{X_t, t \in Z\}$ es estacionario si y solo si las raíces del polinomio $\Phi(B)$ están fuera del disco unitario. En tal caso.

$$X_t = \Phi(B)^{-1} \Theta(B) \epsilon_t.$$

Representación causal de $\{X_t\}$ (No depende del futuro). Representación $MA(\infty)$ asociada al proceso $\{X_t\}$.

Obs: $\epsilon_t = \Theta(B)^{-1} \Phi(B)$ invertible con las raíces del polinomio $\Theta(B)$ están fuera del disco unitario.

Entonces un proceso ARMA(p, q) es un proceso más general.

En la literatura existen técnicas como el procedimiento de Box-Jenkins; identificación, estimación y predicción. En la identificación se analiza la estacionariedad de las series, se aplicarán transformaciones para estabilizar la varianza de series, la diferenciación de la serie, ayudará a transformar de pasar de una serie no estacionaria, a una series estacionaria, transformaciones de $\log(\cdot)$ y $\sqrt{(\cdot)}$ para estabilizar la varianza. También, el cálculo de la función de autocorrelación y autocorrelación parcial, nos ayudan a ver la memoria del proceso cuantos pasos atrás y que con cierto nivel de confianza $\sqrt{(n)}$ vemos cuantos pasos atrás o retardos (lag) está representado.

En la estimación, para procesos AR(p), se estimará a través de las ecuaciones de **Yule-Walker**. Esta estimación puede ser resuelta por la recursión de **Levinson-Durbin** (parecido al usar la función de autocorrelación parcial). La forma más característica para escoger el numero de parámetros es minimizar el coeficiente de Akaike AIC (criterio de información de Akaike) que tiene la siguiente expresión $AIC = -2 \log L + 2k$, donde k es el número de parámetros estimados.

La estimación para procesos ARMA, generalmente es a través de la máxima verosimilitud o estimación por mínimos cuadrados.

La tercera etapa del procedimiento de Box-Jenkins es la verificación de los supuestos de independencia de los residuos.

1.6. Coeficientes de Codispersión para series temporales

El coeficiente de codispersión fue introducido por Matheron en el año 1965, como una extensión del semivariograma para procesos espaciales intrínsecamente estacionarios.

Sean $\{X_t\}$ e $\{Y_t\}$ dos procesos débilmente estacionarios. El variograma cruzado está definido por:

$$\gamma(h) = \mathbb{E}[X_{s+h} - X_s][Y_{s+h} - Y_s], \quad (1.6.1)$$

tal que s y $s+h \in Z$. Entonces el coeficiente de codispersión es

$$\rho_{X_t, Y_t}(h) = \gamma(h) / \sqrt{\mathbb{V}_x(h) \mathbb{V}_y(h)}, \quad (1.6.2)$$

donde

$$\mathbb{V}_x(h) = \mathbb{E}[X_{s+h} - X_s]^2.$$

Se puede apreciar que (1.6.2) es una normalización del variograma cruzado y como consecuencia este coeficiente está acotado en valor absoluto por 1, es decir, $|\rho_{X_t Y_t}(h)| < 1$. La demostración consiste en usar la desigualdad de Cauchy-Shwartz,

$$\mathbb{E}[XY] \leq [\mathbb{E}[X^2]]^{1/2} [\mathbb{E}[Y^2]]^{1/2}. \quad (1.6.3)$$

Un estimador de momentos natural para $\rho_{X_t, Y_t}(h)$ es el coeficiente de codispersión muestral para dos procesos como en (1.6.2).

$$\hat{\rho}_{X_t, Y_t}(h) = \frac{\sum_{i=1}^{n-h} [X_{i+h} - X_i][Y_{i+h} - Y_i]}{\sqrt{\sum_{i=1}^{n-h} [X_{i+h} - X_i]^2 \sum_{i=1}^{n-h} [Y_{i+h} - Y_i]^2}}. \quad (1.6.4)$$

Este coeficiente tiene la ventaja de captar el comportamiento de dos series temporales, en donde se compara proporcionalmente las pendientes en común de pares de puntos.

La codispersión definida en (1.6.2), corresponde a productos internos normalizados de primeras diferencias para las secuencias $\{X_t\}$ e $\{Y_t\}$.

$$cm_{X_t, Y_t} = \rho(1, 0) = \frac{\sum \Delta X \Delta Y}{\sqrt{[\sum \Delta X]^2 [\sum \Delta Y]^2}}. \quad (1.6.5)$$

El coeficiente cm_{X_t, Y_t} , es un coeficiente geoméricamente natural de comovimiento en que se compara proporcionalmente pendientes de pares de puntos en común.

En esta sección se examinará las bondades y características de las series de tiempo que comueven, intuitivamente se puede decir que dos curvas comueven si sus conjuntos de pendientes son proporcionales o casi proporcionales.

Para dos procesos diferenciables estacionarios X_t e Y_t el coeficiente de comovimiento esta definido por:

$$cm_{X_t, Y_t} = \frac{\mathbb{E}(X'_t) \mathbb{E}(Y'_t)}{\sqrt{\mathbb{V}(X'_t) \mathbb{V}(Y'_t)}}, \quad (1.6.6)$$

donde se asume que $\mathbb{E}(X'_t) < \infty$ y $\mathbb{E}(Y'_t) < \infty$.

Para dos secuencias X_t y Y_t , el coeficiente muestral es el siguiente,

$$\widehat{cm}_{X_t, Y_t} = \frac{\sum_{t=1}^{n-1} [X_{t+1} - X_t][Y_{t+1} - Y_t]}{\sqrt{\sum_{t=1}^{n-1} [X_{t+1} - X_t]^2 \sum_{t=1}^{n-1} [Y_{t+1} - Y_t]^2}}. \quad (1.6.7)$$

En la forma integral para curvas suavizadas X_t y Y_t , el estadístico muestral es definido como:

$$\widehat{cm}_{X_t, Y_t} = \frac{\int X'_t dt \int Y'_t dt}{\sqrt{[\int X'_t dt]^2 [\int Y'_t dt]^2}}. \quad (1.6.8)$$

Como está definido el coeficiente de comovimiento, no es el coeficiente de correlación entre las primeras derivadas. Debido a la naturaleza, local del comovimiento en el numerador y denominador

del coeficiente, la $\mathbb{E}(X_t')$ y $\mathbb{E}(Y_t')$ no son restados. De hecho sería indeseable la resta de estos valores esperados (Como en el caso de la covarianza y correlación). Considere el ejemplo de dos líneas rectas con pendientes positivas, donde el valor del coeficiente es igual a 1, pero donde la recta de las medias lleva una expresión indeterminada de la forma $0/0$. Una manera alternativa es pensar que la primera diferenciación, ya ha logrado la restas de las medias.

Aunque el coeficiente de codispersión no es el coeficiente de correlación, comparten varias propiedades. Es claro mostrar que el coeficiente de comovimiento y sus formas muestrales son invariante bajo traslación, positivamente homogéneo, simétrico en sus argumentos, definido positivo para una secuencia y versiones con log para si mismo, e interpretable como el coseno del ángulo entre vectores formados por la primera diferencia de las series muestrales mostradas.

Note que, no hay nada en la definición del estadístico de comovimiento que requiera que las secuencias sean muestreadas equiespaciadamente, como es usualmente en el caso de series de tiempo. La primera diferencia se puede lograr sin la uniformidad o continuidad de espacio entre observaciones. Lo que es requerido por la definición, es que todas las observaciones de las secuencias sean muestreadas en iguales intervalos de tiempo. Es decir, los datos muestreados en los procesos tienen que estar medidos en los instantes t_1, \dots, t_n .

1.7. Algunos ejemplos

Conforme con las definiciones anteriores, se presentará gráficamente como se comportan dos series temporales, esto a través del cálculo del Coeficiente de Codispersión el cual tiene la capacidad de captar su grado de comovimiento.

Esta sección ayudará a interpretar y entender los distintos casos de comovimiento, basados en gráficos y en el cálculo del Índice de Comovimiento, a través del estimador de este.

Seguidamente, se presentaran 3 valores puntuales, los cuales ayudarán a ver la importancia de estos valores para dos series temporales. Los valores del Índice de Comovimiento serán de importancia dependiendo del investigador y las consecuencias que el fenómeno conjunto de las dos series temporales represente. El grado de comovimiento que analizaremos en esta sección serán los casos teóricos extremos 1, 0 y -1 , con $h = 1$.

Para los tres ejemplos las series que se presentarán son modelos simulados AR(1), con estos datos se calculará el grado de comovimiento con el estimador muestral, el tamaño de las series simuladas son de 90 datos.

1.7.1. Comovimiento

Estas series fueron simuladas para que el comovimiento entre las dos series este cercano a 1.

En esta figura se puede observar como las dos series presentan un comportamiento similar. En la práctica cuando uno observa este tipo comportamiento entre dos series, es natural pensar que

las series se mueven juntas, el problema es cuantificar el grado de comovimiento. Si utilizamos el estimador muestral definido anteriormente se puede observar $\hat{\rho}_{X_t, Y_t}(h = 1) = 0,999$.

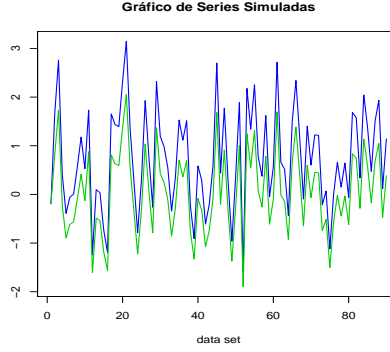


Figura 1.1: Valores de $\hat{\rho}_{X_t, Y_t} \approx 1$.

Cuando se habla del comportamiento de las series de tiempo, se está en presencia de estructuras dinámicas que no están fijas, como en el caso de las variables aleatorias, ya que se ha calculado que grado de comovimiento tienen, entonces de acuerdo a su definición un $\hat{\rho}_{X_t, Y_t}(h = 1)$ cercano a 1, significa que las series se mueven juntas en cualquier período de tiempo $[t_i, t_{i+1}]$.

1.7.2. Anti comovimiento

Estas series fueron simuladas para que el comovimiento entre las dos series este cercano a -1 .

En el siguiente gráfico se puede apreciar un anti comovimiento, es decir, que sus conjuntos de pares de pendientes son inversamente proporcionales. Se puede ver como las series tienen un

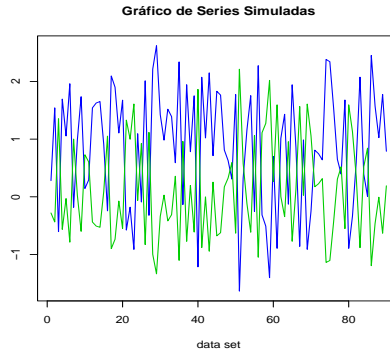


Figura 1.2: Valores de $\hat{\rho}_{X_t, Y_t} \approx -1$.

comportamiento inverso respecto del otro. Si calculamos su grado de comovimiento con el estimador muestral este es $-0,999$. Es decir, estas series comueven de manera inversa en cualquier periodo de tiempo $[t_i, t_{i+1}]$.

1.7.3. Comovimiento nulo

De igual forma, estas series fueron simuladas para que el comovimiento entre las dos series este cercano a 0.

Finalmente, en este gráfico se puede observar que el conjunto de pares de pendientes no tiene nada en común. Es decir, la series en cualquier período de tiempo no comueven. El Índice estimado es 0,0001.

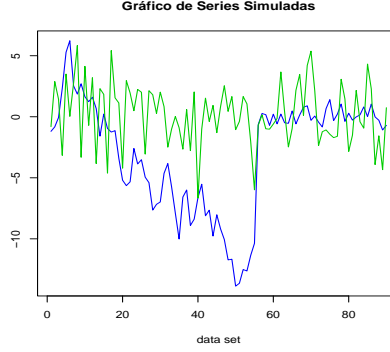


Figura 1.3: Valores de $\hat{\rho}_{X_t, Y_t} \approx 0$.

1.8. Propiedades

Como ya se a dicho, las series de tiempo pueden ser modelados por procesos estocásticos. Esta sección se realizaron los siguientes cálculos para dos procesos lineales generales.

Ahora bien, para analizar la similaridad entre dos series de tiempo, se presentarán algunos cálculos que serán asociados al Índice de Comovimiento. La importancia de esta sección es presentar una medida de similaridad explícita, expresada a través de dos modelos lineales generales. También, se discutirá la normalidad asintótica del Coeficiente de Codispersión muestral. Entonces, sean dos procesos lineales generales de la forma:

$$X_t = \sum_{j=0}^{\infty} \phi_j \epsilon_1(t-j), \quad (1.8.1)$$

$$Y_t = \sum_{k=0}^{\infty} \theta_k \epsilon_2(t-k), \quad (1.8.2)$$

tal que $\sum_{j=0}^{\infty} \phi_j < \infty$, $\sum_{k=0}^{\infty} \theta_k < \infty$ y ϵ_1 y ϵ_2 son ruidos blancos con media 0 y varianza σ^2 y τ^2 respectivamente. Además, consideremos la estructura de covarianza para los errores dada por

$$\text{Cov}(\epsilon_1(t), \epsilon_2(s)) = \begin{cases} \rho\tau\sigma & \text{si } t = s, \\ 0 & \text{en otro caso.} \end{cases} \quad (1.8.3)$$

Hay varios procesos que satisfacen la forma lineal general, por ejemplo para procesos autoregresivos de orden 1 y procesos de media móvil de orden 1. Sin embargo, para procesos ARMA de orden superior no hay una forma explícita (análitica para $\rho_{X_t, Y_t}(h)$).

Ahora se desarrollara los siguientes cálculos. Sea:

$$\mathbb{E}[X_{t+h} - X_t][Y_{t+h} - Y_t] = \mathbb{E}[X_{t+h}Y_{t+h}] - \mathbb{E}[X_{t+h}Y_t] - \mathbb{E}[X_tY_{t+h}] + \mathbb{E}[X_tY_t],$$

$$\begin{aligned} \mathbb{E}[X_{t+h}Y_{t+h}] &= \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \phi_j \epsilon_1(t+h-j) \right) \left(\sum_{k=0}^{\infty} \theta_k \epsilon_2(t+h-k) \right) \right], \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_j \theta_k \mathbb{E}[\epsilon_1(t+h-j) \epsilon_2(t+h-k)], \end{aligned}$$

Si $t+h-j = t+h-k$, entonces $j = k$. Por lo tanto

$$\begin{aligned} \mathbb{E}[X_{t+h}Y_{t+h}] &= \sum_{j=0}^{\infty} \phi_j \theta_j \rho\tau\sigma. \\ \mathbb{E}[X_{t+h}Y_t] &= \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \phi_j \epsilon_1(t+h-j) \right) \left(\sum_{k=0}^{\infty} \theta_k \epsilon_2(t-k) \right) \right] \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_j \theta_k \mathbb{E}[\epsilon_1(t+h-j) \epsilon_2(t-k)] \end{aligned}$$

Si $t+h-j = t-k$, entonces $j = k+h$. Luego

$$\begin{aligned} \mathbb{E}[X_{t+h}Y_t] &= \sum_{j=0}^{\infty} \phi_{k+h} \theta_k \rho\tau\sigma. \\ \text{Además, } \mathbb{E}[X_{t+h}Y_t] &= \sum_{j=0}^{\infty} \phi_{j+h} \theta_j \rho\tau\sigma. \text{ Similarmente,} \\ \mathbb{E}[X_tY_{t+h}] &= \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \phi_j \epsilon_1(t-j) \right) \left(\sum_{k=0}^{\infty} \theta_k \epsilon_2(t+h-k) \right) \right] \end{aligned}$$

$$= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_j \theta_k \mathbb{E}[\epsilon_1(t-j) \epsilon_2(t+h-k)]$$

Si $t-j = t+h-k$, entonces $k = j+h$

$$\mathbb{E}[X_t Y_{t+h}] = \sum_{j=0}^{\infty} \phi_j \theta_{j+h} \rho \tau \sigma. \text{ Finalmente,}$$

$$\begin{aligned} \mathbb{E}[X_t Y_t] &= \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \phi_j \epsilon_1(t-j) \right) \left(\sum_{k=0}^{\infty} \theta_k \epsilon_2(t-k) \right) \right] \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_j \theta_k \mathbb{E}[\epsilon_1(t-j) \epsilon_2(t-k)] \end{aligned}$$

Si $t-j = t-k$, entonces $k = j$. Por lo tanto,

$$\mathbb{E}[X_t Y_t] = \sum_{j=0}^{\infty} \phi_j \theta_j \rho \tau \sigma$$

Note que $\mathbb{E}[X_{t+h} Y_{t+h}] = \mathbb{E}[X_t Y_t]$

Para el cálculo del denominador del coeficiente se tiene que

$$\begin{aligned} \mathbb{E}[X_{t+h} - X_t]^2 &= \mathbb{V}ar[X_{t+h} - X_t] = \mathbb{V}ar[X_{t+h}] + \mathbb{V}ar[X_t] - 2\mathbb{C}ov[X_{t+h} X_t]. \\ \mathbb{V}ar[X_{t+h}] &= \mathbb{V}ar \left[\sum_{j=0}^{\infty} \phi_j \epsilon_1(t+h-j) \right] = \sum_{j=0}^{\infty} \phi_j^2 \mathbb{V}ar[\epsilon_1(t+h-j)] = \sum_{j=0}^{\infty} \phi_j^2 \sigma^2. \\ \mathbb{V}ar[X_t] &= \mathbb{V}ar \left[\sum_{j=0}^{\infty} \phi_j \epsilon_1(t-j) \right] = \sum_{j=0}^{\infty} \phi_j^2 \mathbb{E}[\epsilon_1(t-j)] = \sum_{j=0}^{\infty} \phi_j^2 \sigma^2. \\ \mathbb{C}ov[X_{t+h} X_t] &= \mathbb{E}[X_{t+h} X_t] - \mathbb{E}[X_{t+h}] \mathbb{E}[X_t] = \mathbb{E}[X_{t+h} X_t] \\ \mathbb{E}[X_{t+h} X_t] &= \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \phi_j \epsilon_1(t+h-j) \right) \left(\sum_{k=0}^{\infty} \phi_k \epsilon_1(t-k) \right) \right] \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_j \phi_k \mathbb{E}[\epsilon_1(t+h-j) \epsilon_1(t-k)] \end{aligned}$$

Si $t+h-j = t-k$, entonces, $k = j+h$

$$= \sum_{j=0}^{\infty} \phi_j \phi_{j+h} \sigma^2$$

Similarmente, se tiene:

$$\begin{aligned} \mathbb{E}[Y_{t+h} - Y_t]^2 &= \mathbb{V}ar[Y_{t+h} - Y_t] = \mathbb{V}ar[Y_{t+h}] + \mathbb{V}ar[Y_t] - 2\mathbb{C}ov[Y_{t+h} Y_t] \\ \mathbb{V}ar[Y_{t+h}] &= \mathbb{V}ar \left[\sum_{k=0}^{\infty} \theta_k \epsilon_2(t+h-k) \right] = \sum_{k=0}^{\infty} \theta_k^2 \mathbb{V}ar[\epsilon_2(t+h-k)] = \sum_{k=0}^{\infty} \theta_k^2 \tau^2. \end{aligned}$$

$$\begin{aligned}
\text{Var}[Y_t] &= \text{Var} \left[\sum_{k=0}^{\infty} \theta_k \epsilon_2(t-k) \right] = \sum_{j=0}^{\infty} \theta_j^2 \text{Var}[\epsilon_2(t-j)] = \sum_{k=0}^{\infty} \theta_k^2 \tau^2. \\
\text{Cov}[Y_{t+h} Y_t] &= \mathbb{E}[Y_{t+h} Y_t] - \mathbb{E}[Y_{t+h}] \mathbb{E}[Y_t] = \mathbb{E}[Y_{t+h} Y_t] \\
\mathbb{E}[Y_{t+h} Y_t] &= \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \theta_j \epsilon_2(t+h-j) \right) \left(\sum_{k=0}^{\infty} \theta_k \epsilon_2(t-k) \right) \right] \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_j \theta_k \mathbb{E}[\epsilon_1(t+h-j) \epsilon_1(t-k)]
\end{aligned}$$

Si $t+h-j = t-k$, entonces, $k = j+h$

$$\mathbb{E}[Y_{t+h} Y_t] = \sum_{j=0}^{\infty} \theta_j \theta_{j+h} \tau^2$$

Ahora si juntamos los términos y factorizamos nos queda.

$$\rho_{X_t, Y_t}(h) = \rho \cdot \frac{\sum_{j=0}^{\infty} (2\phi_j \theta_j - \phi_{j+h} \theta_j - \phi_j \theta_{j+h})}{2\sqrt{\sum_{j=0}^{\infty} (\phi_j^2 - \phi_j \phi_{j+h}) \sum_{j=0}^{\infty} (\theta_j^2 - \theta_j \theta_{j+h})}}.$$

Como caso particular:

Cuando $X_t \sim \text{AR}(1)$ y $Y_t \sim \text{AR}(1)$, los parámetros ϕ y θ satisfacen $\phi_j = \phi^j$, $|\phi| < 1$ y $\theta_j = \theta^j$, $|\theta| < 1, |\theta\phi| < 1$

Usando las siguientes identidades.

$$\begin{aligned}
\sum_{k=0}^{\infty} \theta^k \phi^k &= \sum_{k=0}^{\infty} (\theta\phi)^k = \frac{1}{1-\theta\phi}, \\
\sum_{k=0}^{\infty} \theta^k \phi^{k+h} &= \phi^h \sum_{k=0}^{\infty} (\theta\phi)^k = \frac{\phi^h}{1-\theta\phi}, \\
\sum_{k=0}^{\infty} \theta^k \phi^{k+h} &= \phi^h \sum_{k=0}^{\infty} (\theta\phi)^k = \frac{\phi^h}{1-\theta\phi}, \\
\sum_{k=0}^{\infty} \phi^{2k} &= \frac{1}{1-\phi^2}, \\
\sum_{k=0}^{\infty} \theta^{2k} &= \frac{1}{1-\theta^2}, \\
\sum_{k=0}^{\infty} \phi^k \phi^{k+h} &= \phi^h \sum_{k=0}^{\infty} (\phi)^{2k} = \frac{\phi^h}{1-\phi^2}, \\
\sum_{k=0}^{\infty} \theta^k \theta^{k+h} &= \theta^h \sum_{k=0}^{\infty} (\theta)^{2k} = \frac{\theta^h}{1-\theta^2},
\end{aligned}$$

Se tiene

$$\rho_{X_t, Y_t}(h=1) = \rho \cdot \frac{(2 - \phi - \theta)\sqrt{(1 + \phi)(1 + \theta)}}{1 - \phi\theta}. \quad (1.8.4)$$

Desde aquí se puede apreciar una forma explícita del Coeficiente de Codispersión $\rho_{X_t, Y_t}(h=1)$. El Coeficiente de Codispersión es una versión corregida del coeficiente de correlación clásico. Este resultado fue obtenido por Rukhin y Vallejos (2008). Las propiedades asintóticas de $\hat{\rho}(h)$ fueron establecidas para el proceso $\mathbb{Z}_s = (\mathbb{X}_s, \mathbb{Y}_s)^t$ admitiendo la siguiente estructura,

$$\mathbb{Z}_{s+h} - \mathbb{Z}_s = \sum_l \mathbb{A}_l \epsilon_{s-l}, \quad (1.8.5)$$

Donde $\mathbb{A}_l = \mathbb{A}_h(l)$ son matrices 2×2 definida para todo l tal que, $\sum \|\mathbb{A}(l)\|^2 < \infty$, donde $\|\cdot\|$ denota cualquier norma matricial y ϵ_t son vectores independientes con media cero y matriz de covarianza Σ .

Teorema 1.8.1. *Si el valor observado admite la representación del proceso $Z_t = (X_t, Y_t)^t$ (1.8.5) con matrices $A(k, l) = \text{diag}(\alpha_{kl}, \beta_{kl})$, la distribución límite de $M[\rho_{X,Y}(\hat{h}) - \rho]$ es normalmente distribuída con media 0 y varianza*

$$\nu^2 = \left(1 - \frac{\rho^2 \sum_{s=0}^{\infty} (\alpha_s \beta_s)^2}{\sum_{s=0}^{\infty} \alpha_s^2 \sum_{s=0}^{\infty} \beta_s^2}\right). \quad (1.8.6)$$

Con este resultado, se han realizados trabajos con Dóclimas de hipótesis, intervalos de Confianza y Bandas de confianzas para el Coeficiente de Codispersión (Ver Rukhin y Vallejos, 2008).

La ventaja de encontrar una forma explícita para la varianza del coeficiente facilita los cálculos para dóclimar y encontrar intervalos de Confianza ya que en caso contrario, habría que calcular la varianza, a través, de métodos de re-muestreo.

1.9. Limitaciones teóricas

Dentro de las limitaciones teóricas se puede mencionar que para modelos con muchos parámetros no existe una forma analítica para el coeficiente, en este caso es necesario implementar técnicas computacionales. Cabe, mencionar que para series de tiempo no estacionarias el índice de codispersión no ha sido estudiado. Cuando una de las series presenta outliers, el coeficiente de codispersión es muy sensible. La definición de una version robusta del coeficiente de Codispersión aún es un problema abierto.

1.10. Métodos no paramétricos

En la literatura existe una medida de asociación para procesos espaciales. Tjøstheim (1978), propone una medida de asociación para variables espaciales, esta medida está basada en los rangos de las observaciones y la localización de las coordenadas de los puntos medios.

Capítulo 2

Agrupamiento de series temporales

2.1. Introducción

En este Capítulo se dará una definición del Índice de Disimilaridad, para así implementar una medida con el cual se trabajará para la clasificación de series temporales.

La motivación de analizar el comportamiento de un par de series temporales, es mayor aún cuando éste se extiende para un conjunto de series temporales.

Ahora bien, como símil al análisis de cluster, es necesario extender algunos métodos ya conocidos de agrupación, para un conjunto de variables aleatorias a través del tiempo. La variación de esta técnica, es considerar la información del comovimiento a través del tiempo de los distintos pares de series temporales, donde el análisis multivariante clásico no considera.

El análisis por agrupación (Cluster analysis) es una técnica muy utilizada en problemas multivariados donde se quiere agrupar unidades experimentales con ciertas características o generalizar grupo de variables de interés. El objetivo de este Capítulo es definir algunas técnicas de agrupación. Las técnicas más usadas en este contexto son:

1. Métodos de agrupación no jerárquicos.
2. Métodos de agrupación jerárquicos.

Estas definiciones ayudarán de base para definir una nueva medida de proximidad entre series temporales, asociados a las técnicas clásicas de agrupación.

2.2. Métodos de Agrupación

Dependiendo del método de agrupación que utilicemos los resultados finales variarán. Entonces, cada método tiene una característica en particular que es necesario considerar, en esta sección se estudiarán los siguientes métodos de agrupación:

1. **Métodos de agrupación no jerárquicos** Se usan para agrupar objetos en un conjunto de k-cluster predeterminados, se parte de un conjunto inicial de clusters elegidos al azar luego se van cambiando de modo iterativo, habitualmente se usa el método de las k-medias.

Método de las k-medias

Es un método que permite asignar a cada observación el cluster que se encuentra más próximo en términos el centroide (Valor medio de las observaciones de las variables en el valor del conglomerado).

- Se eligen al azar k-clusters iniciales.
- Para el conjunto de observaciones se vuelve a calcular la distancia al centroide de los clusters y se reasignan a los que están mas próximos. Se vuelven a calcular los centroides de los k-cluster después de las reasignaciones de los elementos.
- Se repiten los pasos anteriores hasta que no se produzca ninguna reasignación.

2. **Métodos de agrupación jerárquicos**

Métodos Divisivos

Se comienza con un gran conglomerado que contiene todas las observaciones u objetos en los pasos sucesivos en las observaciones que son más diferentes, se dividen y se construyen conglomerados más pequeños. Este proceso continua hasta que cada observación es un mismo conglomerado.

2.1 Método de encadenamiento simple (Vecino más cercano)

Se basa en la distancia mínima, encuentra los dos objetos separados por la distancia más pequeña y las coloca en el primer conglomerado, a continuación se encuentra la distancia más pequeña o bien un tercer objeto. Se une a los dos primeros para formar un conglomerado o se forma un nuevo conglomerado de dos miembros. El proceso continua hasta que todos los objetos se encuentren en un conglomerado.

2.2 Método de encadenamiento completo (Vecino más lejano)

Este método es similar al anterior excepto que el criterio de aglomeramiento se basa en la distancia máxima.

También, el uso de dendogramas es una herramienta útil cuando se quiere graficar grupos.

El método que se utilizará para clasificación es el de agrupación jerárquico, ya que, se asumirá que no se conocen la cantidad de grupos iniciales. En nuestro análisis usaremos el software R.

2.3. Índice de Disimilaridad Adaptativo

Para la clasificación y agrupación de series temporales es necesario tener en cuenta la estructura de correlación temporal, tanto como su comportamiento respecto de su comovimiento a través

del tiempo, como el comportamiento respecto de su cercanía entre las series. En esta sección se presentará una propuesta para abarcar dichos elementos en la agrupación de series temporales.

Existen autores que se han interesado en la agrupación de series temporales, por ejemplo Liao (2005) hace una reseña de técnicas, medidas y algoritmos para la clasificación de series temporales. Por otra parte, Chouakria y Nagabbbhusham (2007) estudiaron un método para detectar la interdependencia entre las series, dado que con medidas convencionales, no logran captar su dependencia temporal y proponen una medida alternativa que cubre estos dos aspectos del análisis de series temporales. A continuación se hará una pequeña descripción de éstos métodos, introduciendo definiciones clásicas de medidas de desemejanzas, entre ellas se pueden destacar.

2.3.1. Distancia Euclidiana

Sean $X_t = (x_1, \dots, x_n)$ e $Y_t = (y_1, \dots, y_n)$ dos series de tiempo de n valores observados en el tiempo en los instantes t_1, \dots, t_n . La distancia Euclidiana δ_E entre X_t e Y_t está definida como:

$$\delta_E = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (2.3.1)$$

2.3.2. Distancia de Minkowski

Para dos secuencias X_t e Y_t la distancia de Minkowski δ_M de orden p entre X_t e Y_t está definida como:

$$\delta_M = \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p}. \quad (2.3.2)$$

2.3.3. Distancia de Frechét

Se define una rejilla $r \in M$ entre dos series de tiempo $X_t = (x_1, x_2, \dots, x_n)$ e $Y_t = (y_1, y_2, \dots, y_k)$ como la secuencia de pares preservando el orden de la información.

$$r = ((x_{a_1}, y_{b_1}), (x_{a_2}, y_{b_2}), \dots, (x_{a_n}, y_{b_m}))$$

con $a_i, i \in \{1, \dots, n\}, b_j, j \in \{1, \dots, m\}$ y satisfaciendo para $i \in \{1, \dots, m-1\}$ la siguiente condición: $a_1 = 1, b_m = k$ $a_i = (a_i \text{ ó } a_{i+1})$ y $b_j = (b_j \text{ ó } b_{j+1})$.

Se define $|r| = \max_{i=1, \dots, m} |x_{a_i} - y_{b_i}|$, la rejilla $|r|$ está representada por la máxima distancia entre dos observaciones pareadas.

La distancia $\delta_F(X_t, Y_t)$ se define como:

$$\delta_F(X_t, Y_t) = \min_{r \in M} |r| = \min_{r \in M} \left(\max_{i=1, \dots, m} |x_{a_i} - y_{b_i}| \right) \quad (2.3.3)$$

2.4. Alineamiento de tiempo distorsionado (Dynamic time warping DTW)

El alineamiento de tiempo distorsionado (DTW) es una técnica que encuentra la alineación óptima entre dos series temporales. El alineamiento de tiempo distorsionado es a menudo usado en el reconocimiento del lenguaje hablado para determinar si dos formas de onda representan la de la misma forma de locución hablada.

El alineamiento de tiempo distorsionado ha sido útil dentro de muchas otras disciplinas, la robótica, la manufactura, y la medicina. Un ejemplo de cómo una serie temporal está distorsionada es mostrado en la siguiente figura.

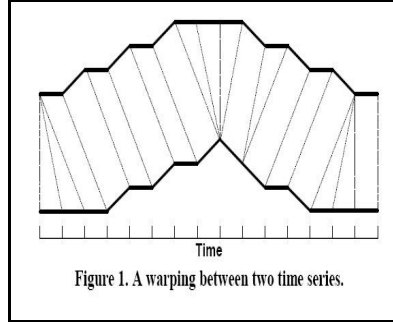


Figura 2.1: Alineamiento de tiempo distorsionado

En la Figura 2.1, cada línea vertical conecta un punto en una serie temporal con su punto correspondiente, similar en la otra serie temporal. Las líneas realmente tienen valores similares en el eje vertical pero han sido separados, así es que en las líneas verticales entre ellos pueden ser miradas más fácilmente.

Si ambas series temporales en la figura fueran idénticas, todas las líneas serían líneas directamente verticales porque todos los puntos serían similares entre ellos.

2.4.1. Formulación del problema

El problema del alineamiento de tiempo distorsionado es definido como sigue:

Dadas dos series de tiempo $X_t = (x_1, \dots, x_i, \dots, x_m)$ e $Y_t = (y_1, \dots, y_j, \dots, y_n)$ de largo m y n respectivamente. Se construye el camino distorsionado que llamaremos W_k , esta es una secuencia de pares de puntos ordenados, basado en los índices de las series X_t e Y_t , donde:

$|r| = W_k = \{w_1, \dots, w_K\}$, $\max\{m, n\} \leq K \leq n \cdot m$. Además, $K = m \times n$ es la longitud máxima del camino distorsionado. El camino distorsionado W_k se define de la siguiente forma:

$$w_k = (i, j), \text{ con } k = 1, \dots, K \quad (2.4.1)$$

donde i es el índice de la Serie X_t y j es el índice de la serie Y_t .

El camino distorsionado debe empezar desde $w_1 = (1, 1)$ y el final de ambas en $w_K = (m, n)$. Esto asegura que cada índice de ambas serie temporales es usado en el camino distorsionado. Hay una restricción en el camino distorsionado que obliga a i y j a ser monotonamente creciente. Se define mas formalmente:

$$r = ((x_{a_1}, y_{b_1}), (x_{a_2}, y_{b_2}), \dots, (x_{a_m}, y_{b_n})) = (w_1, w_2, \dots, w_K),$$

con $a_i, i \in \{1, \dots, n\}, b_j, j \in \{1, \dots, n\}$ y satisfaciendo para $i \in \{1, \dots, m-1\}$ la siguiente condición: $a_1 = 1, b_n = k$ $a_i = (a_i \text{ ó } a_{i+1})$ y $b_j = (b_j \text{ ó } b_{j+1})$.

Para detreminar DTW para dos secuencias de tiempo, construimos una matriz de $m \cdot n$ donde (i^{th}, j^{th}) son elementos de la matriz contiene la distancia $d(x_{a_i}, y_{b_i})$ entre los dos puntos x_{a_i} e y_{b_i} , es decir, $d(x_{a_i}, y_{b_i}) = |x_{a_i} - y_{b_i}|$. Cada elemento matricial (i, j) corresponde a la alineación entre el punto de x_{a_i} e y_{b_i} .

Este camino puede ser encontrado usando programación dinámica para evaluar la siguiente recurrencia, que defina la distancia acumulativa $\gamma(i, j)$ como la distancia $d(i, j)$ encontrada en la celda actual y el mínimo de las distancias acumulativas de los elementos adyacentes,

$$\gamma(i, j) = d(i, j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\}. \quad (2.4.2)$$

2.4.2. Hacia el Alineamiento de tiempo distorsionado

El Alineamiento de tiempo distorsionado es un algoritmo para medir la similitud entre dos secuencias que pueden variar en el tiempo o en la velocidad. DTW se ha aplicado en video, audio y gráficos, de hecho en cualquier dato que pueda ser convertido en una representación lineal se puede analizar con DTW. Una conocida ha sido la aplicación automática de reconocimiento de voz, para hacer frente a diferentes velocidades haciendo uso de la palabra.

En general, DTW es un método que permite a un ordenador encontrar una óptima adecuación entre dos secuencias dadas (por ejemplo, series de tiempo), con ciertas restricciones.

Un ejemplo de las restricciones impuestas a la adecuación de las secuencias se encuentra en la monotocidad de la cartografía en la dimensión temporal. La continuidad es menos importante en DTW que en otros algoritmos de Coincidencia de patrones; DTW es un algoritmo especialmente adecuado para las secuencias de concordancia con la información que falta. El proceso de optimización se realiza utilizando programacin dinámica, de ahí el nombre. Ejemplo de una de las muchas formas del algoritmo.

```
#####
#### RUTINA PARA DETERMINAR EL CAMINO DISTROSIONADO #####
int DTWDistance (char s [1 .. n], char t [1 .. m], int d [1 .. n, 1 .. m]) (
    DTW declarar int [0 .. n, 0 .. m]
    declarar int i, j, el costo
```

```

for i: = 1 to m
    DTW [0, i]: = infinito
for i: = 1 hasta n
    DTW [i, 0]: = infinito
DTW [0,0]: = 0

for i: = 1 hasta n
    for j: = 1 to m
        coste: = d [s [i], t [j]]
        DTW [i, j]: = + costo mnimo (DTW [i-1, j], / / insercin
                                   DTW [i, j-1], / / supresin
                                   DTW [i-1, j-1]) / / partido

    DTW retorno [n, m]
)
#####

```

Este algoritmo calcula la distancia DTW para dos secuencias, donde los vectores pueden ser de distinto tamaño. En la siguiente figura se puede apreciar que el algoritmo encuentra el camino mas corto entre dos series, que dependen de la similitud entre las mismas.

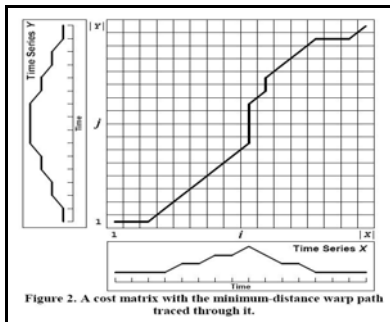


Figura 2.2: Camino distorsionado

Ya mencionadas las características se definirá mas formalmente la Distancia DTW.

2.4.3. Definición DTW

Considere una nueva definición de la rejilla propuesta en (2.3.3) como la suma de las distancias de todas las observaciones apareadas.

$$|r| = \sum_{i=1}^m |x_{a_i} - y_{b_i}|. \quad (2.4.3)$$

Entonces la distancia δ_{DTW} se define de la siguiente manera:

$$\delta_{DTW}(X_t, Y_t) = \min |r| = \min_{r \in M} \left(\sum_{i=1}^m |x_{a_i} - y_{b_i}| \right). \quad (2.4.4)$$

2.4.4. Ejemplo

La finalidad de este ejemplo es resumir esta sección para dejar más claro cuales son las bondades de DTW. Entonces, considere dos secuencias de tiempo. Sea $X_t = (1, 0.34, 0.65, 2, 3, 3.4, 1, 1.2, 0.88, 5.9, 7)$ e $Y_t = (2, 5.5, 5.67, 3.45, 7, 3.3, 2.43, 1.34)$. Calcular $\delta_{DTW}(X_t, Y_t)$. Considere la siguiente rutina.

```
#####

library(dtw)
x<-c(1,0.34,0.65,2,3,3.4,1,1.2,0.88,5.9,7)
y<-c(2,5.5,5.67,3.45,7,3.3,2.43,1.34)
a<-dtw(x,y)
b<-a$distance
plot(a,xlab="indice de x",ylab="indice de y",main="Camino Distorcionado")
> length(x)
[1] 11
> length(y)
[1] 8
> b
[1] 26.45

#####
```

En la salida de esta rutina se puede apreciar que la longitud del vector X_t es de 11 elementos y la longitud del vector Y_t es 8 elementos. Por otra parte la distancia es 26.45.

La opción `plot(a)` en R muestra el siguiente gráfico.

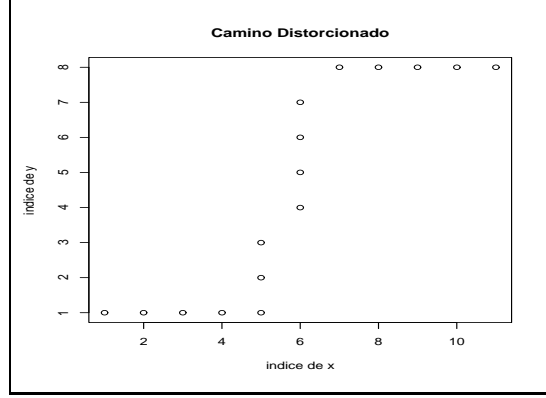


Figura 2.3: Camino distorsionado de $\delta_{DTW}(X_t, Y_t)$.

Una de las características de DTW, es la representación gráfica de la similitud entre dos series temporales y en este caso se puede apreciar que las series no presentan un gran grado de similitud, ya que estos puntos no definen una línea recta.

Desde aquí, se puede apreciar y entender más fácilmente la definición de camino distorsionado.

En esta sección se han presentado las distancias δ_E , δ_M , δ_F , y δ_{DTW} pero ellas ignoran la estructura temporal de los valores como proximidad, ya que, están basadas sobre las diferencias entre los valores $|x_{a_i} - y_{b_i}|$.

En la siguiente sección se propondrá un Índice que considere la estructura de comovimiento y la cercanía entre las series en el contexto de clasificación de series temporales.

2.5. Índice de Disimilaridad Adaptativo para medidas de proximidad entre series de tiempo.

Las medidas convencionales de proximidad en series temporales, están basadas sobre la cercanía de los valores observados de las series de tiempo.

La idea de esta sección es presentar un Índice de Disimilaridad, que contenga la información de la codispersión o comovimiento y la cercanía de las series temporales. En otras palabras se define la disimilaridad entre dos series de tiempo, considerando estas dos características, uno el comportamiento respecto a su comovimiento y la otra sobre su cercanía. Generalmente, se asume que esta medida entre las secuencias X_t e Y_t toman valores positivos.

Uno puede cuantificar el concepto de similaridad considerando el clásico coeficiente de correlación de Pearson. Sin embargo, esta correlación guía una sobre estimación de la Codispersión. De igual forma, el coeficiente de correlación de Spearman es también usado como una medida de similaridad. No obstante, se ha visto que la correlación está principalmente basada sobre los rangos y no sobre los valores observados. Dejando de lado la información respecto a su comovimiento.

2.5.1. Correlación temporal para medidas de proximidad

Ya se ha mencionado las desventajas de las medidas de distancias convencionales y los coeficientes de similaridad que no ayudan a establecer un índice que pueda conjugar estas dos características. Una estructura que considere la interdependencias entre dos serie puede deducirse considerando los siguientes elementos:

Sea $X_t = (x_1, \dots, x_n)$ una colección de n números reales independientes. La varianza clásica de X_t puede ser escrita en dos vias equivalentes,

$$Var(X_t) = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 = \frac{1}{n-1} \sum_{i,i'} (x_i - m)^2, \quad (2.5.1)$$

donde m es la media de los valores de X_t . Similarmente el coeficiente clásico de correlación entre $X_t = (x_1, \dots, x_n)$ e $Y_t = (y_1, \dots, y_n)$ puede ser escrito de dos formas:

$$Corr_{X_t, Y_t} = \frac{\sum_{i=1}^n (x_i - m_1)(y_i - m_2)}{\sqrt{Var(X_t)Var(Y_t)}} = \frac{\sum_{i,i'} (x_i - x_{i'})(y_i - y_{i'})}{\sqrt{\sum_{i,i'} (x_i - x_{i'})^2 \sum_{i,i'} (y_i - y_{i'})^2}}. \quad (2.5.2)$$

En el caso general de medidas independientes varianza/covarianza y coeficiente de correlación son medidas basadas sobre la contribución de todos los pares de medición. A la inversa en el caso de las interdependencias se define una relación entre vecino más cercano.

La expresión de la varianza puede ser descompuesta como sigue:

$$Var(X_t) = \frac{1}{2n-1} \sum_{i,i'} (x_i - x_{i'})^2. \quad (2.5.3)$$

De igual forma,

$$Var(S) = \frac{1}{2n-1} \sum_{i \text{ es vecino de } i'} (x_i - x_{i'})^2 + \frac{1}{2n-1} \sum_{i \text{ no es vecino de } i'} (x_i - x_{i'})^2. \quad (2.5.4)$$

La idea principal para incluir la información sobre la dependencia de las series, es una restricción a la expresión de la varianza/covarianza de los pares de valores dependientes (Vecinos).

$$VarT(X_t) = \frac{1}{2n-1} \sum_{i \text{ no es vecino de } i'} (x_i - x_{i'})^2, \quad (2.5.5)$$

Para la similaridad, se considerara una relación temporal de vecinos de orden h , medidas en el período $[t_i, t_{i+h}]$. Este coeficiente de correlación temporal es definido como:

$$\hat{\rho}_{X_t, Y_t}(h) = \frac{\sum_{i=1}^{n-1} (x_{i+h} - x_i)(y_{i+h} - y_i)}{\sqrt{\sum_{i=1}^{n-1} (x_{i+h} - x_i)^2 \sum_{i=1}^{n-1} (y_{i+h} - y_i)^2}}. \quad (2.5.6)$$

donde $\hat{\rho}_{X_t, Y_t}(h)$ pertenece al intervalo $[-1, 1]$. La interpretación de los valores se discutió en la sección de ejemplos en el capítulo I.

2.5.2. Índice de Disimilaridad Adaptativo entre series de tiempo

El propósito de este capítulo es apuntar a un nuevo Índice con aspiración a cubrir ambas medidas convencionales de similaridad y cercanía para la proximidad de valores y la correlación temporal entre dos series temporales para medir el comportamiento conjunto.

El resultado de la medida de disimilaridad debe también permitir ajustar el peso de la contribución entre ambos cuantiles (Comovimiento y Cercanía). Para ajustar el peso de la contribución de los cuantiles es necesario tener una función $f(x)$ que module la distancia incrementando o disminuyendo las medidas convencionales antes definidas.

Conforme con las especificaciones anteriores, se propone un Índice de Disimilaridad $D(X_t, Y_t)$ basado sobre una función de afinación automática que module el comportamiento de las medidas convencionales.

Ahora bien, en el contexto de análisis de grupos (Cluster) Chouakria y Nagabhushan (2007) propusieron un Índice de Disimilaridad Adaptativo para medidas de proximidad en series temporales, como una nueva medida de distancia, la cual definieron de la siguiente forma,

$$D(X_t, Y_t) = f(\hat{\rho}_{X_t, Y_t}(h))\delta_C(X_t, Y_t), \quad (2.5.7)$$

donde $f(x)$ es una función de afinación exponencial adaptativa, es dado por

$$f(x) = \frac{2}{1 + \exp(kx)}, \quad k \geq 0. \quad (2.5.8)$$

y $\delta_C(X_t, Y_t)$ es cualquier distancia convencional entre X_t y Y_t . Note que la función $f(x)$ depende del estimador muestral del Coeficiente de Codispersión. En la Figura 2.4, se puede apreciar la función de afinamiento adaptativo para distintos valores de k .

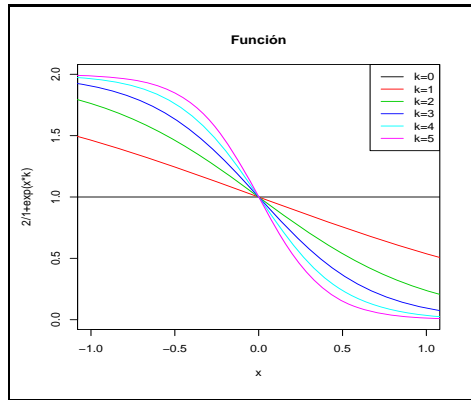


Figura 2.4: Función exponencial adaptativo

La función de k es modular con más o menos fuerza la contribución de los cuantiles sobre las medidas convencionales. Esto se puede apreciar en la tabla.

Contribución en % de k con respecto a		
k	$f(x)$	$D(X_t, Y_t)$
0	0	100
1	46.2	53.7
2	76.2	23.8
3	90.5	9.4
≥ 5	≈ 100	≈ 0

Por ejemplo, cuando $k = 0$ y $f(0) = 1$, el Índice de Disimilaridad está representado sólo por la distancia convencional y estamos en presencia de una medida de semejanza clásica, para los algoritmos de clasificación en Cluster.

Con este nuevo Índice de Disimilaridad se espera captar mejor la correlación temporal que las medidas convencionales de distancias para clasificación de series temporales.

Este Índice de disimilaridad es una medida alternativa para clasificar series temporales a los métodos clásicos que se conocen. Un método de clasificación que depende de este índice de disimilaridad será presentada en el Capítulo III.

Capítulo 3

Simulación

3.1. Introducción

Para entender mejor la funcionalidad del Índice de Disimilaridad, es necesario realizar algunas simulaciones donde se examine como funciona teóricamente por simulación? y desde ahí, explicar mejor esta nueva medida para clasificar series temporales. Como se mencionó anteriormente el Índice de Disimilaridad considera la distancia entre las series ponderada por una función adaptable de afinación que balancea la proximidad con relación a los valores y la proximidad con relación al comportamiento. La contribución del comportamiento y componentes de valores para el Índice de Disimilaridad es comparada con las medidas convencionales.

La idea es sensibilizar esta medida y compararla con las distancias convencionales que fueron definidas anteriormente y ver cuales son sus diferencias.

El objetivo de este capítulo es implementar un algoritmo de clasificación para series temporales. Se espera que el Índice de Disimilaridad al considerar la información del comovimiento agrupe las series, que las medidas convencionales no agrupan, es decir, se quiere demostrar que este Índice agrupa series temporales con una justificación mayor que las distancias convencionales.

3.2. Simulación entre series correlacionadas

Antes de representar y graficar la funcionalidad del Índice de Disimilaridad es necesario explicar, cuales son las condiciones de esta simulación, detallar los modelos que serán analizados con sus respectivas características.

Los elementos necesarios para esta simulación es la estructura de correlación de los errores. Por ejemplo, considere la siguiente estructura de correlación:

$$\mathbb{C}ov(\epsilon_c(t), \epsilon_b(s)) = \begin{cases} \rho\tau\sigma & \text{si } t = s \\ 0 & \text{eoc} \end{cases} . \quad (3.2.1)$$

Donde $\epsilon_c(t)$ y $\epsilon_b(t)$, son errores correlacionados con media 0 y varianza τ^2 , σ^2 , respectivamente. Seguidamente, esta estructura se aplicará para 7 modelos AR(1), cada uno con errores $\{\epsilon_i, i = 1, \dots, 7\}$ aleatorios independientes idénticamente distribuidos normal con media 0 y varianza σ^2 .

Para verificar las bondades del Índice de Disimilaridad antes mencionados, intencionalmente se simularan errores $\epsilon_2(t)$ y $\epsilon_3(t)$ que estarán correlacionados con $\epsilon_1(t)$. Entonces, dependiendo de que valor se le de a cada correlación ρ_1 y ρ_2 es cuan correlacionados estarán los errores. Se esperaría que el Índice de Disimilaridad agrupara primero las series con los errores que fueron intencionalmente correlacionados.

Ahora, considere la siguiente estructura de correlación:

$$\begin{aligned} \text{corr}(\epsilon_1(t), \epsilon_2(s)) &= \rho_1, \quad t = s, \\ \text{corr}(\epsilon_1(t), \epsilon_3(s)) &= \rho_2, \quad t = s. \end{aligned}$$

Los valores de ρ_1 y ρ_2 , ayudaran a diferenciar los otros 4 modelos en los cuales sus errores serán variables independientes normales con media 0 y varianza 1.

Otra forma de considerar estos errores correlacionados, es a través de una matriz de correlación para el vector $\mathbf{y} = (\epsilon_1, \epsilon_2, \epsilon_3)^t$, donde cada $\epsilon_i, i = 1, 2, 3$ tiene varianza σ^2 , τ^2 y η^2 , respectivamente. Entonces, el proceso es simulado con la estructura para la covarianza Σ .

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho_1 \tau \sigma & \rho_2 \sigma \eta \\ \rho_1 \tau \sigma & \tau^2 & \rho \tau \eta \\ \rho_2 \sigma \eta & \rho \tau \eta & \eta^2 \end{pmatrix},$$

y como consecuencia la matriz de correlación del vector de errores \mathbf{y} es:

$$\mathbb{Corr}(\mathbf{y}) = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho \\ \rho_2 & \rho & 1 \end{pmatrix}.$$

Luego, en estructura de correlación se incorpora otro elemento que ayudará entender mejor esta nueva medida. Se definen 7 modelos AR(1), los cuales 3 de ellos tendrán una característica en común, difiriendo en el grado de interdependencia.

Ahora, ya creados los errores con la estructura de correlación antes mencionada para ϵ_2 y ϵ_3 correlacionadas con ϵ_1 , se incorporan en los modelos Y_t y W_t .

Finalmente, se construyen los siguientes modelos AR(1):

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \epsilon_1(t) \\ Y_t &= \phi_2 Y_{t-1} + \epsilon_2(t) \\ W_t &= \phi_3 W_{t-1} + \epsilon_3(t) \\ V_t &= \phi_4 V_{t-1} + \epsilon_4(t) \\ T_t &= \phi_5 T_{t-1} + \epsilon_5(t) \\ U_t &= \phi_6 U_{t-1} + \epsilon_6(t) \\ Z_t &= \phi_7 Z_{t-1} + \epsilon_7(t) \end{aligned}$$

En todos los casos $|\phi_i| < 1$ para $i = 1, 2, \dots, 7$ para estar en presencia de series estacionarias. En las simulaciones los parámetros fueron $\phi_1 = -0,5$; $\phi_2 = 0,3$; $\phi_3 = -0,8$; $\phi_4 = 0,7$; $\phi_5 = 0,1$; $\phi_6 = -0,9$ y $\phi_7 = 0,2$, $\rho_1 = 0,9$ y $\rho_2 = 0,7$, $n = 200$. Donde $\{\epsilon_i\}$ con $i = 4, 5, 6, 7$ son variables aleatorias independientes normal con media 0 y varianza 1.

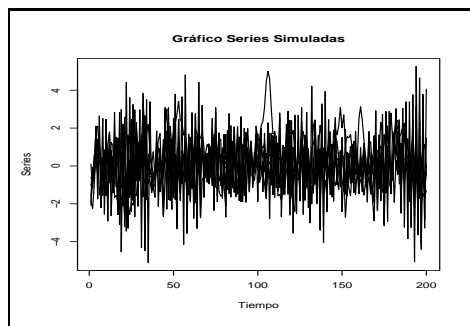


Figura 3.1: Series simuladas

En este gráfico se puede observar un conjunto de series estacionarias.

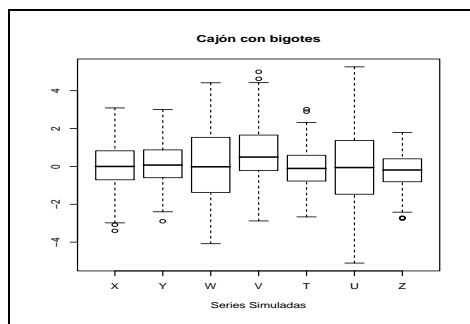


Figura 3.2: Series simuladas

Además, si aplicamos el algoritmo de clasificación con las distancias convencionales, se pueden obtener los siguientes resultados.

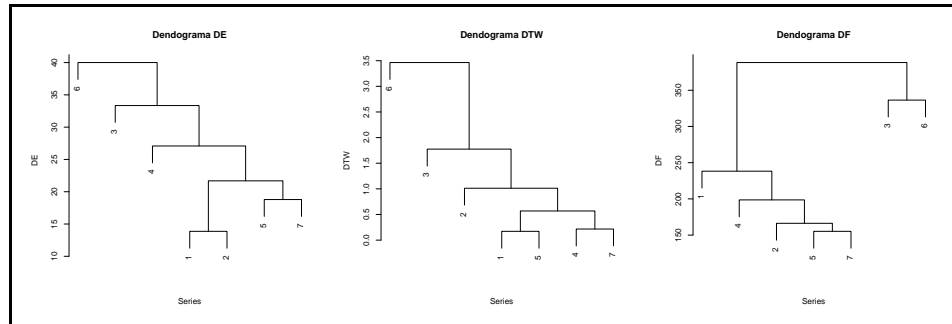


Figura 3.3: Dendogramas.

En base a la información de las distancias convencionales y con el algoritmo de clasificación utilizando el método jerárquico, del vecino más cercano, se puede decir que la distancia euclidiana agrupa 2 de las 3 series correlacionadas, en cambio las distancias de Frechét y DTW no agrupan ninguna de las series con errores correlacionados.

3.2.1. Gráficos Camino Distorsionado

Una característica que tiene la distancia DTW, es graficar la opción de camino distorsionado entre las series, a continuación se presentarán algunos gráficos con distintas combinaciones entre las distintas series y analizar la similitud entre las series.

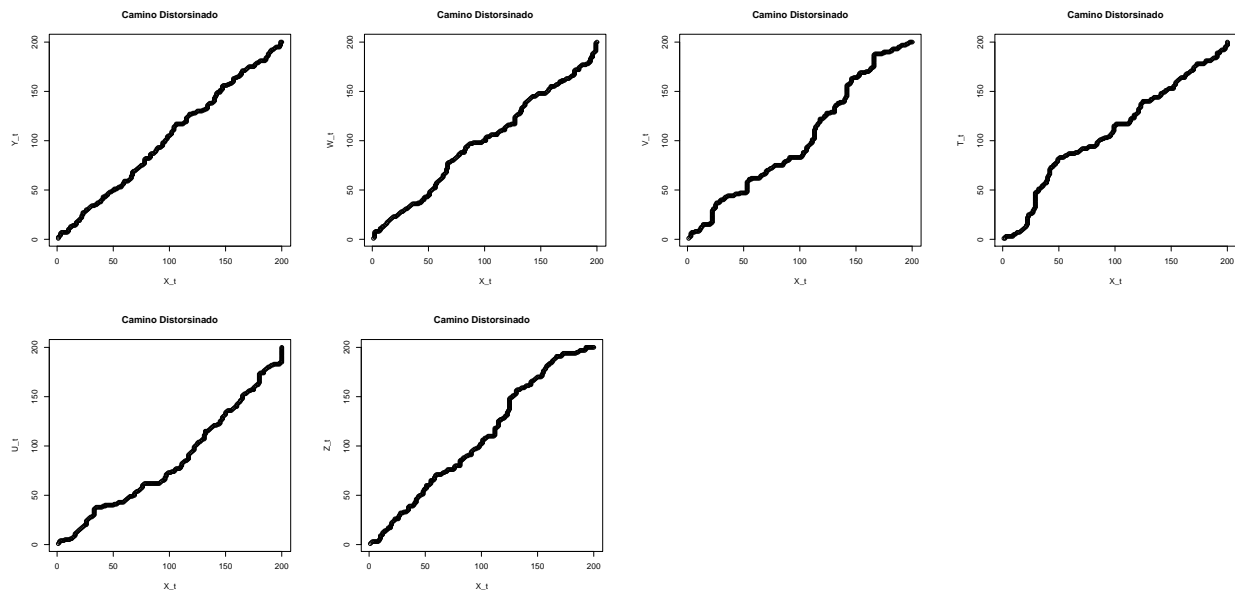


Figura 3.4: Camino Distorsionado.

Estos gráficos nos dan una idea de como interactúan las series, se puede apreciar un grado de similitud entre las series, ya que, los puntos del camino distorsionado presentan una recta.

Por otra parte, si se aplica el algoritmo de clasificación con el Índice de Disimilaridad Adaptativo con $h = 1, 2, 3$ y $k = 1, 2, 3, 4$ se puede observar los siguientes resultados.

3.2.2. Índice de Disimilaridad Adaptativo con Distancia Euclidiana

En base a la simulación y el cálculo de las matrices de distancia con esta nueva medida de clasificación se puede apreciar los siguientes resultados.

El Índice con $h = 1$ agrupa las series X_t , Y_t y W_t , como era de esperar, agrupa primero X_t con Y_t , ya que sus errores tienen un grado más alto de correlación, luego incorpora la series W_t . A continuación se presenta el algoritmo de clasificación para series temporales con $h = 2, 3$. Se puede observar que para $h = 2$, el Índice de Disimilaridad mantiene los grupos, que están correlacionados. Con $h = 3$, el Índice de Disimilaridad mantiene los grupos establecidos.

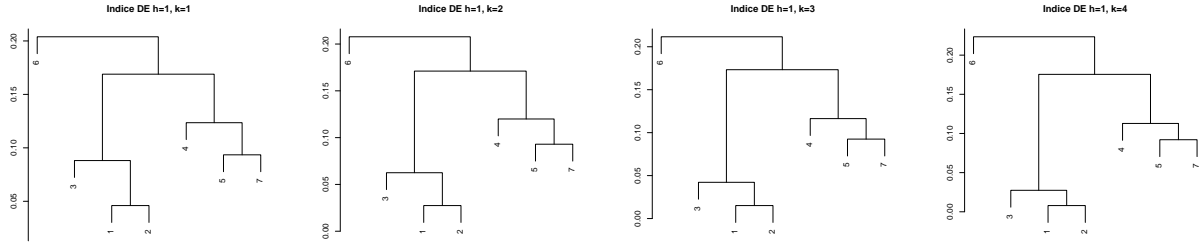


Figura 3.5: Dendogramas Índice con δ_E con $h = 1$, $k = 1, 2, 3, 4$.

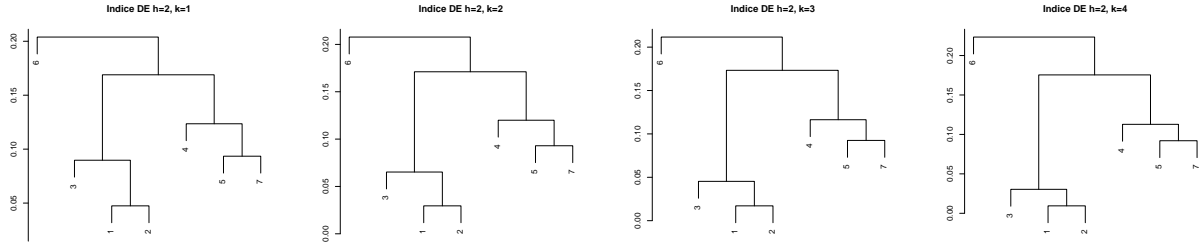


Figura 3.6: Dendogramas Índice con δ_E con $h = 2$, $k = 1, 2, 3, 4$.

3.2.3. Índice de Disimilaridad Adaptativo con Distancia DTW

De igual forma que anteriormente calculamos el Índice de Disimilaridad Adaptativo con la Distancia DTW con $h = 1, 2, 3$ y $k = 1, 2, 3, 4$, se observan los siguientes resultados.

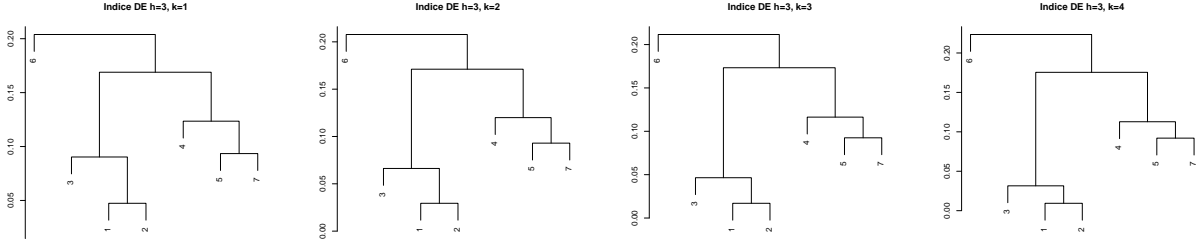


Figura 3.7: Dendogramas Índice con δ_E con $h = 3$, $k = 1, 2, 3, 4$.

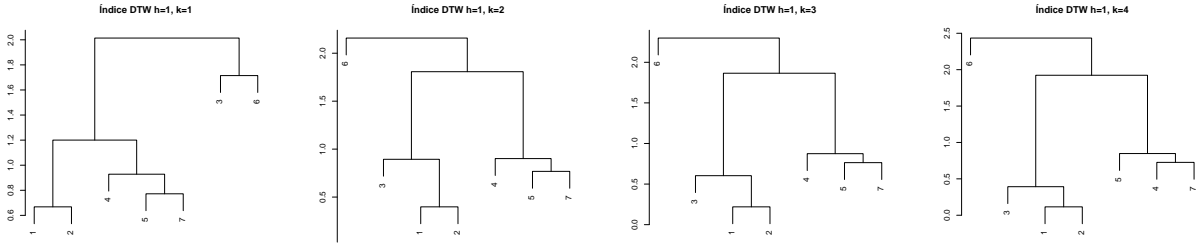


Figura 3.8: Dendogramas Índice con δ_{DTW} con $h = 1$, $k = 1, 2, 3, 4$.

En el este caso con $h = 1$ y $k = 1$, el Índice no agrupó las series esperadas. No obstante, con $k = 2, 3$, el Índice agrupa las series esperadas.

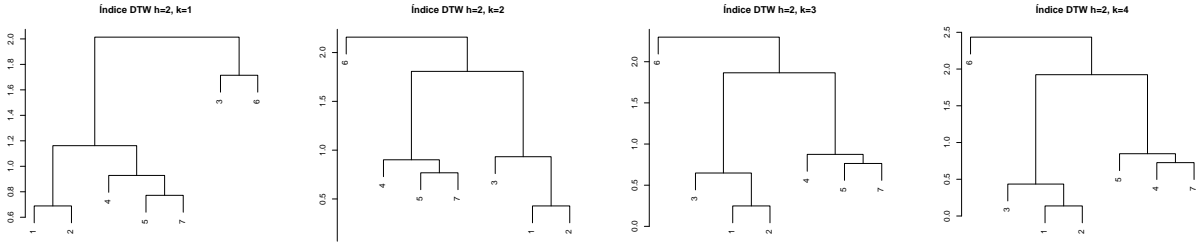


Figura 3.9: Dendogramas Índice con δ_{DTW} con $h = 2$, $k = 1, 2, 3, 4$.

De igual manera sucede con $h = 2$ y $k = 1$ el algoritmo clasifica dos de las tres series. Con $h = 2$, $h = 3$ y $k = 2, 3$ el algoritmo vuelve agrupar las series se simularon.

3.2.4. Índice de Disimilaridad Adaptativo con Distancia de Frechét

Finalmente si se calcula el Índice de Disimilaridad con Distancia se Frechét, se pueden observar los siguientes resultados. El Índice calculado con $h = 1, 2$ y $k = 1, 2, 3$ muestra algunas variaciones

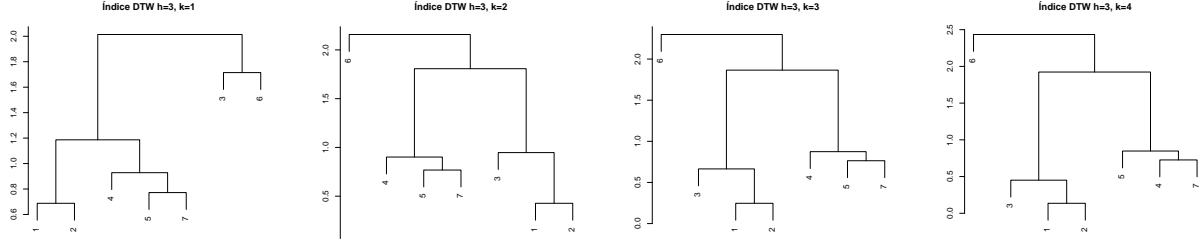


Figura 3.10: Dendogramas Índice con δ_{DTW} con $h = 3$, $k = 1, 2, 3, 4$.

con los resultados esperados, ya que, el índice no logra percibir las diferencias entre las correlaciones. Por otra parte, con $h = 3$ y $k = 1$, el Índice agrupa las series X_t , Y_t y W_t . De igual forma se puede apreciar que el Índice va agrupar mejor con altos valores de h y k .

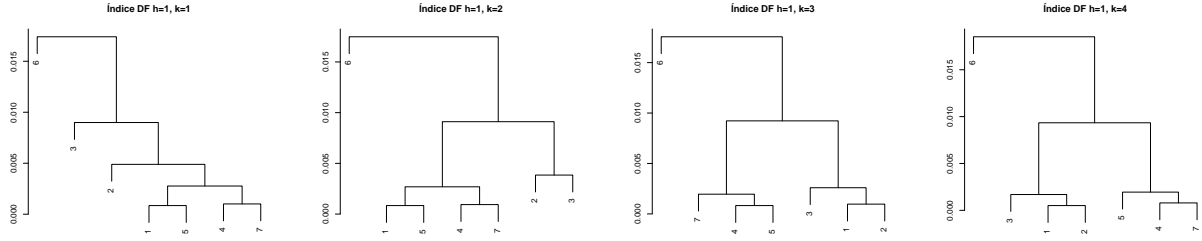


Figura 3.11: Dendogramas Índice con δ_F con $h = 1$, $k = 1, 2, 3, 4$.

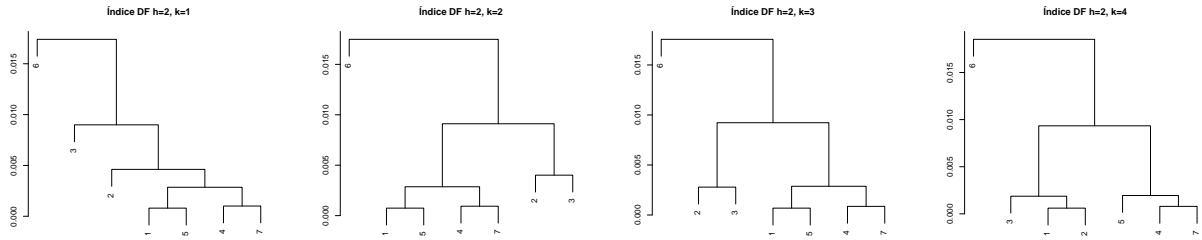


Figura 3.12: Dendogramas Índice con δ_F con $h = 2$, $k = 1, 2, 3, 4$.

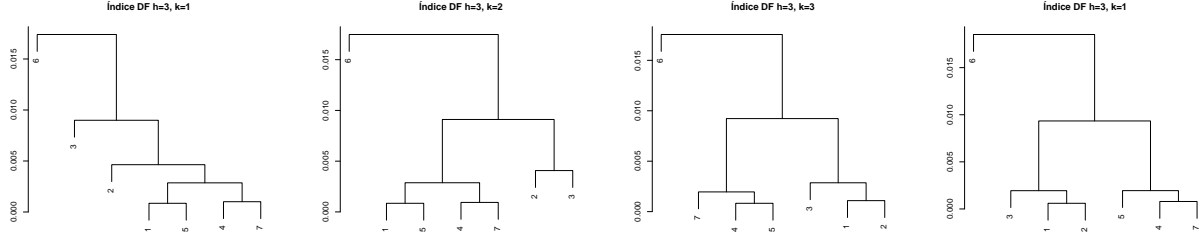


Figura 3.13: Dendrogramas Índice con δ_F con $h = 3$, $k = 1, 2, 3, 4$.

3.3. Conclusión

En los dendrogramas se puede apreciar que el Índice de Disimilaridad con distancia euclidiana y DTW agrupa las series correlacionadas.

El Índice con la distancia de Frechét presenta una variación en la agrupación, ya que para valores altos de h y k , los resultados son alcanzados.

La función de k , en el Índice de Disimilaridad modula la contribución del comovimiento respecto del comportamiento de los valores respecto a su cercanía, esto se puede apreciar mejor cuando las distancias son poco sensibles, ya que para valores altos de k , la distancia convencional se ve afectada por una mayor contribución al comportamiento respecto del comovimiento, es decir, para valores de $k \geq 3$ la distancia convencional sera mas sensible a los cambios del comportamiento respecto de su comovimiento.

Es importante, calcular el Índice con distintos valores de h y k , para verificar y sacar conclusiones respecto de los grupos establecidos por esta nueva medida.

La simulación a demostrado que el Índice de Disimilaridad agrupa series temporales considerando su estructura de comovimiento y el comportamiento a su cercanía, es decir, el Índice agrupa las series que están fuertemente correlacionadas.

Se demostró que las medidas convencionales ignoran la relación de interdependencia entre las medidas, ya que, son primordialmente basadas en la proximidad con relación a los valores.

Capítulo 4

Aplicación a datos de AFP

4.1. Introducción

En las siguientes secciones se introducirá algunos conceptos del Sistema de Pensiones, se hablará de su función, sus génesis y la reforma de Previsión Social.

Estos conceptos nos ayudarán para entender y analizar 7 AFP distintas en el período de Enero 1990-Febrero 2004 en su rentabilidad mensual.

Las AFP en Estudio son:

- 1 = AFP CUPRUM, 2 = AFP HABITAT, 3 = AFP PROVIDA, 4 = AFP PLANVITAL, 5=AFP STA MARIA, 6 = AFP SUMMA, 7 = AFP MAGISTER.

El ahorro de los trabajadores ha sido fundamental para el crecimiento y desarrollo económico de Chile, ya que ha sido mayoritariamente invertido en actividades productivas o crediticias haciendo posible el acceso a financiamiento de miles de chilenos, sea para comprar la casa propia o acceder a productos a los que nunca antes tuvo oportunidad.

El sistema de AFP ha operado en estos 26 años bajo el principio de Giro Único o exclusivo lo cual ha resguardado los fondos previsionales de problemas como las ventas atadas y los conflictos de interés, que surgen cuando un administrador de ahorros distribuye conjuntamente otros productos y además gestiona simultáneamente dineros propios y de terceros.

4.2. Sistema de AFP

En noviembre de 1980 se publicó el D.L. 3.500 que estableció un nuevo sistema de pensiones de Vejez, Invalidez y Sobrevivencia, sobre la base del ahorro de los trabajadores y la capitalización en cuentas individuales.

El nuevo Sistema de Pensiones incorpora el concepto de propiedad de los ahorros previsionales por parte de los trabajadores afiliados, enfatizando la estrecha correspondencia entre el esfuerzo de

ahorro realizado a lo largo de la vida activa de una persona y los beneficios en pensiones de vejez, invalidez y sobrevivencia que ésta recibe.

Así también, se instaura la administración de los ahorros por empresas privadas (AFP), con giro único y con el rol de otorgar beneficios y prestaciones previsionales. La normativa que regula a estas instituciones, además de la ley, es dictada por la Superintendencia de AFP, la que además fiscaliza el adecuado funcionamiento de estas sociedades.

El objetivo del Sistema de Pensiones es proveer ingresos de reemplazo para los trabajadores que dejan la vida activa o laboral y cubrir los riesgos de invalidez (total o parcial) y de muerte del trabajador (sobrevivencia), de manera de proteger al afiliado y a su grupo familiar.

Se basa en el ahorro y la Capitalización Individual. Los trabajadores dependientes cotizan obligatoriamente en las AFP y los independientes lo hacen en forma voluntaria. Los trabajadores son dueños de su ahorro previsional y en ellos recae la responsabilidad de preocuparse de su pensión, sin perjuicio que el Estado garantice pensiones mínimas.

Otorga libertad de elección a los afiliados. De este modo el trabajador puede elegir la administradora que gestione sus ahorros previsionales y cambiarse cuando lo desee, así como la edad a la que quiere pensionarse (jubilación por vejez o anticipada) y la modalidad de pago de pensión (retiro programado, renta vitalicia o retiro programado con renta vitalicia diferida). Asimismo, puede elegir el Tipo de Fondo en donde invertir sus ahorros.

Es uniforme en la aplicación de las normas para todos los afiliados y establece directa relación entre las contribuciones de los trabajadores y los beneficios obtenidos.

La administración es privada y está a cargo de sociedades anónimas especializadas.

4.3. Rentabilidad de los Fondos de Pensiones

La rentabilidad de los fondos es un factor determinante en un sistema de capitalización y ahorro individual, donde el saldo acumulado durante la vida activa del trabajador incidirá significativamente en el monto de las pensiones. De esta forma, un punto porcentual de mayor ganancia anual puede elevar el monto de la pensión al cabo de 30 ó 40 años de ahorro entre un 25 % y 30 %. Tras 26 años del sistema de pensiones la rentabilidad promedio real anual alcanza a UF +10 %, un resultado muy superior al 4 a 5 % de la rentabilidad estimada como suficiente para financiar pensiones de reemplazo con una tasa del orden del 70 % del promedio de las remuneraciones.

4.4. Contribución al Desarrollo Económico

Los fondos de pensiones han hecho un aporte significativo al desarrollo económico de Chile, otorgando, asimismo, acceso a amplios sectores de la población a servicios financieros a los que nunca habían podido acceder. Para un país en desarrollo como Chile, hoy economía emergente, disponer de ahorro fue siempre un problema y una aspiración, ya que es sabido que el círculo

virtuoso del progreso comienza con el ahorro y sigue con la inversión y el crecimiento. En la actualidad un 25 % del total de activos financieros nacionales pertenece a los fondos de pensiones de los trabajadores, lo cual constituye un aporte tangible al desarrollo económico.

El mercado de capitales chileno se ha expandido en forma significativa los últimos 26 años, progreso en el cual el ahorro de los fondos de pensiones ha jugado un rol clave. Se han convertido en una fuente de financiamiento para el Estado, para los bancos y para las empresas. El sistema de pensiones ha introducido competencia en el mercado de capitales.

Casi el 100 % del financiamiento de letras hipotecarias para la vivienda y bonos de bancos pertenece a los fondos de pensiones; también un 45 % de los bonos de empresas y un 55 % de los títulos estatales.

4.5. Antiguo Sistema Previsional

Chile fue el primer país de América Latina que creó un Sistema de Seguridad Social, a comienzos del siglo XX. A lo largo de los años se fueron creando diversos regímenes de pensiones diferenciados por el tipo de actividad o grupos ocupacionales con reglas y beneficios distintos.

Es así como llegan a coexistir 52 "Cajas" o Instituciones de Previsión que operaban bajo el esquema de reparto. Esto significa que los aportes de los afiliados activos financiaban las pensiones de los pasivos y, por tanto, la subsistencia del Sistema estaba supeditada a la relación trabajador/pensionado existente en la población en cada momento del tiempo.

Durante los primeros años de existencia del sistema, la proporción de trabajadores fue suficiente para financiar los beneficios de los pensionados. Sin embargo, los cambios demográficos, que fueron reflejando una permanente disminución de la natalidad y un aumento en las expectativas de vida revirtieron esta relación, provocando un fuerte desfinanciamiento del Sistema.

Mientras que en el año 1955, por cada 12,2 trabajadores cotizantes había 1 pensionado, en 1980 por cada 2,5 trabajadores cotizantes había 1 pensionado. Es decir, sólo en 25 años el costo de los trabajadores cotizantes, se incrementó casi 5 veces.

Un agravante del problema del financiamiento lo constituyó la fuerte evasión previsional, ya que a trabajadores y empleadores les resultaba más económico hacer imposiciones por el mínimo legal, preocupándose sólo de imponer por valores reales los últimos años de la vida activa del trabajador, cuando las imposiciones eran consideradas para la jubilación. Esta situación obligaba al Estado a elevar las imposiciones lo que, a su vez, incentivaba una mayor evasión previsional y así sucesivamente.

A ellos se agrega que el Estado fue proclive a otorgar beneficios sin el adecuado financiamiento, lo que acentuó el problema reseñado, ocasionando un déficit fiscal creciente, equivalente a un 28 % del gasto en la década 1970 - 1980.

El Antiguo Sistema se caracterizó también por su falta de equidad. Dado que no existía una relación directa entre los aportes de los trabajadores y los beneficios percibidos, se apreciaban notables

desigualdades entre los múltiples grupos cubiertos. Esta situación se sustentaba en la facultad del poder político para definir quién se beneficiaba y cuánto, quedando de manifiesto el otorgamiento de mayores concesiones a los grupos que ejercían mayor presión. En efecto, en el año 1965 (1), los obreros chilenos, que representaban el 70 % de los cotizantes y cuyos ingresos eran los más bajos, percibían en términos absolutos, pensiones equivalentes a la mitad de lo que obtenían los empleados privados y a 1/14 de lo obtenido por los empleados públicos, los cuales eran grupos de mayores ingresos. Al mismo tiempo, el aporte efectuado por estos trabajadores en el mismo período (2), era equivalente a más del doble del que realizaban los empleados públicos y sólo un 10 % inferior al de los privados.

Desde hacía mucho tiempo en nuestro país se presentaba la necesidad de introducir cambios al sistema de Seguridad Social. Ya en la década del '60 se elaboraron diversos informes sobre las falencias del antiguo sistema de Seguridad Social Chileno en los que se proponían cambios profundos.

El desfinanciamiento y la inequidad del esquema de reparto dieron origen a una reforma Previsional que creó, mediante el D.L. 3.500 de 1980, un Nuevo Sistema de Pensiones basado en la Capitalización Individual y administrado por entidades privadas denominadas Administradoras de Fondos de Pensiones (AFP). El Antiguo Sistema continuó funcionando, principalmente a través de un ente único, denominado Instituto de Normalización Previsional (INP), el cual fusionó a las principales Cajas de Previsión.

El Estado se hizo responsable del financiamiento de las cotizaciones pagadas en el Antiguo Sistema por aquellas personas que se cambiaron al Nuevo Sistema. Ello se materializa a través de unos instrumentos financieros denominados Bonos de Reconocimiento, los cuales son representativos de dichos períodos de cotizaciones y que el trabajador hace efectivo al instante de pensionarse o fallecer.

El Bono de Reconocimiento se reajusta de acuerdo a la variación de la inflación y devenga un interés del 4 % real anual, el cual se capitaliza cada año.

4.6. Efecto Manada

Uno de los vicios que se le han imputado al sistema de AFP es que si todas no invirtieran en los mismos activos, quizás inyectarían competencia al mercado por la vía de ofrecer mejores retornos a los afiliados.

Es por eso que las distintas AFP invierten en acciones en las mismas empresas, revelando que la rentabilidad en las distintas AFP son similares. Este fenómeno se llama Efecto Manada.

4.6.1. Contexto y consecuencias de la competencia en rentabilidad vía ranking

Para entender el contexto de esta competencia, es útil tener en mente dos hechos. Primero, las magnitudes de las diferencias en rentabilidad tienden a ser pequeñas. De hecho, 10 puntos base de

mayor o menor rentabilidad pueden hacer la diferencia entre el primero y el último en el ranking en un año determinado.

Segundo, hay extrema transparencia en las decisiones de inversión. Las carteras de todos los fondos de pensiones son conocidas 10 días después de cerrado el mes. Si se imagina que una buena decisión de inversión es una especie de invento, entonces, en este caso, no habría un período de protección o patente de uso; todos podrían conocerlo y utilizarlo en forma gratuita pocos días después de haber sido creado.

Lo anterior, en forma absolutamente natural, genera incentivos a situarse muy cerca del benchmark (o vara de medida), que en este caso es la cartera de la competencia relevante, y a tomar decisiones en el margen, relativamente pequeñas en magnitud. Nótese que la cantidad de trabajo o estudio necesarios para tomar una buena decisión de inversión es la misma, independientemente de la magnitud de la decisión de inversión tomada. El riesgo pertinente pasa a ser el comercial, que consiste en quedar en una mala posición en el ranking de rentabilidad. Una medida que refleja adecuadamente que el riesgo se mide en términos relativos es el llamado tracking error o variabilidad de la diferencia de retorno con el sistema. El fenómeno descrito genera el llamado efecto manada o herding (lo que, en todo caso, no es exclusividad del sistema adoptado en Chile, ya que, se da en casi toda la industria de administración de carteras).

Algo que sí es particular del caso chileno es que existe por ley una rentabilidad mínima que se calcula con respecto al promedio del sistema de AFP. Esta norma tiende a reafirmar el comportamiento descrito en el párrafo anterior, pero es probable que incluso sin rentabilidad mínima la forma de competir fuera similar.

4.7. Acciones favoritas de las AFP

Según Econsult, en un Artículo publicado el 8 de Abril de 2008, en el Diario La Tercera, revisado el Miércoles 23 de Julio 2008, 16:45 hrs, mostró que Provida, apuesta más que el resto en CMPC, donde mantiene una sobreexposición de 1.81 % respecto de la Industria. Su segunda apuesta más fuerte comparativamente hablando es Endesa y AntarChile. La segunda AFP por fondo administrativo, Habitat, tiene como favorita a Telefónica CTC, seguida por Endesa y AntarChile.

Planvital e ING Santa María, por ejemplo, favorecen más a Soquimich que el sistema como un todo, aunque luego sus preferencias difieran. Cuprum, por su parte, favorece a Colbún, las acciones de Telefónica CTC y a Masisa; y al cierre de Junio Summa Bansander mostraba una mayor inclinación por Endesa, D&S y Copec.

4.7.1. El método de Encosult

Enconsult sumó la inversión directa de cada AFP en acciones mantenidas en los fondos A, B, C, D y E. Con ello pudo determinar el porcentaje del fondo invertido en cada título y observar cuáles son las acciones que más pesan en el.

Luego la consultora tomo esa exposición total del sistema en una acción y le resto la exposición particular de cada AFP. El calculo se efectuó excluyendo del total a cada administradora según fuera el caso (sistema sin Bansander, luego sin Provida y así). De esta forma pudo determinar en que porcentaje cada administradora esta sobre o bajo la media exposición del sistema para cada inversion, evitando la distorsión de que la AFP se compare en cierto modo consigo misma

4.8. Análisis Descriptivo

En esta sección se hará un sondeo, para entender como se han comportado las 7 AFP, desde un punto de vista descriptivo. En esta primera parte se realizará un conteo para saber cuantas veces han estado primeras, es decir, con la mayor rentabilidad, en el caso que una AFP tiene la mayor rentabilidad se le asignara un valor 1 y al resto de las AFP 0, entonces, finalmente se sumara todos los casos para saber cuantas veces estuvo primera en rentabilidad, si hay mas de un empate, es decir, que mas de dos AFP tienen la mayor rentabilidad, se le asignaran a cada una el valor 1.

La siguiente tabla muestra los resultados obtenidos, para las AFP en estudio.

AFP	CONTEO
Cuprum	35
Habitat	38
Magister	44
Planvital	26
Provida	37
Sta Maria	21
Summa	38

Se observa que la AFP que estuvo con mayor rentabilidad y esta primera con 44 veces dentro de este conteo es la AFP Magister, seguido de la AFP Provida que estuvo 37 veces primera y la AFP Cuprum con 35 veces.

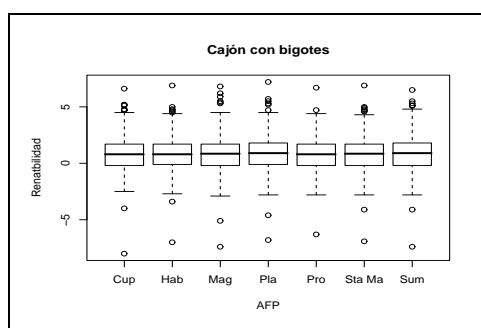


Figura 4.1: Dispersión de la rentabilidad de las AFP.

Se puede observar que en gráfico de Cajón con bigotes, que hay datos que escapan a estos Cajones con bigotes. Además, las media son similares en todas las AFP. Acontinuación, se mostrara un sumario con algunas medidas descriptivas.

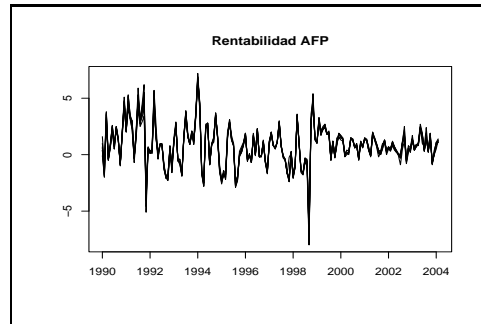


Figura 4.2: 7 Series durante 1990 - 2004.

En el siguiente gráfico se muestra el comportamiento de la rentabilidad en el tiempo de las 7 AFP, si se observa el gráfico se puede apreciar una línea gruesa, este fenómeno se denomina efecto manada el cual se ha mencionado anteriormente y en consecuencia no se puede notar grandes diferencias. En la siguiente tabla se muestra un análisis descriptivo por AFP.

Estadísticos	Cuprum	Habitat	Magister	Planvital	Provida	ING Sta María	Summa
Media	0,83	0,82	0,84	0,86	0,78	0,80	0,85
Mediana	0,81	0,83	0,86	0,90	0,80	0,87	0,90
Moda	-1,10	0,80	-1,70	1,60	-1,60	0,90	-1,70
Desv Est	1,82	1,77	1,92	1,88	1,68	1,80	1,82
Varianza	3,31	3,12	3,70	3,53	2,82	3,24	3,33
Coef Var	220,07	215,03	230,05	219,74	215,82	224,85	214,81
Rango	14,60	13,90	14,20	14,00	13,00	13,80	13,90
Mínimo	-8,00	-7,00	-7,40	-6,80	-6,30	-6,90	-7,40
Máximo	6,60	6,90	6,80	7,20	6,70	6,90	6,50
Cuenta	170,00	170,00	170,00	170,00	170,00	170,00	170,00

Si se observa la tabla se puede destacar que la rentabilidad media de Cuprum es de 0.8265, las rentabilidades extremas se pueden apreciar con un mínimo de -6.3 y un máximo de 6.7. Además, se puede apreciar que en un 25 % la rentabilidad llega a -0.175 y en el 75 % llega a 1.7. Es claro, mencionar que para el resto de las AFP la interpretación es la misma.

Más adelante se estudiarán estas AFP, para representar, a través, de modelos paramétricos y ver cual modelo puede representar mejor esta situación.

4.9. Aplicación Índice de Comovimiento o Codispersión

Para el conjunto de 7 AFP, se calculara el Índice de Comovimiento para determinar en que grado comueven, los resultados son:

Índice de Comovimiento con $h = 1$.

$\rho(h = 1)$	Cuprum	Habitat	Provida	Planvital	Sta María	Summa
Habitat	0,990					
Provida	0,984	0,982				
Planvital	0,983	0,987	0,993			
Sta Maria	0,982	0,993	0,970	0,979		
Summa	0,991	0,995	0,988	0,992	0,989	
Magister	0,986	0,985	0,983	0,986	0,975	0,987

Se puede observar que el comovimiento entre las AFP es similar, es decir, estas comueven casi perfectamente en cualquier instante de tiempo.

Índice de Comovimiento con $h = 2$. De igual forma si se calcula ahora con $h = 2$, se puede apreciar los siguientes resultados:

$\rho(h = 2)$	Cuprum	Habitat	Provida	Planvital	Sta María	Summa
Habitat	0,990					
Provida	0,987	0,985				
Planvital	0,989	0,987	0,993			
Sta Maria	0,983	0,994	0,970	0,979		
Summa	0,992	0,995	0,988	0,992	0,989	
Magister	0,990	0,985	0,983	0,986	0,975	0,987

Los valores están cercanos a 1, es decir, que las series se mueven conjuntamente iguales en cualquier periodo de tiempo $[t_i, t_{i+2}]$ En este caso los valores son similares a la tabla anterior.

Índice de Comovimiento con $h = 3$. Ahora con $h = 3$, se pueden apreciar valores similares.

$\rho(h = 3)$	Cuprum	Habitat	Provida	Planvital	Sta María	Summa
Habitat	0,988					
Provida	0,986	0,987				
Planvital	0,988	0,987	0,993			
Sta Maria	0,981	0,994	0,970	0,979		
Summa	0,990	0,995	0,988	0,992	0,989	
Magister	0,983	0,985	0,983	0,986	0,975	0,987

Índice de Comovimiento con $h = 4$. Finalmente con $h = 4$, las diferencias entre valores no son significativa por simple inspección.

$\rho(h = 4)$	Cuprum	Habitat	Provida	Planvital	Sta María	Summa
Habitat	0,990					
Provida	0,986	0,987				
Planvital	0,989	0,989	0,993			
Sta Maria	0,982	0,993	0,970	0,979		
Summa	0,992	0,995	0,988	0,992	0,989	
Magister	0,990	0,985	0,983	0,986	0,975	0,987

Este alto grado de comovimiento, se puede deber a que las AFP bajo el fenómeno de efecto manada, comuevan casi perfectamente en cualquier instante de tiempo $[t_i, t_{i+h}]$.

4.10. Modelamiento

Como se mencionó anteriormente, se estudiará la estacionariedad y el comportamiento de la rentabilidad. Para poder modelar esta situación se usará las definiciones de series de tiempo mencionadas anteriormente y se estudiará como se comportan respecto al índice o coeficiente de Codispersión para ver cual es su grado de comovimiento.

Por otra parte, se aplicará el índice de disimilaridad, para clasificar las AFP respecto a su rentabilidad y considerando la estructura de correlación temporal, esto ayudará a responder los objetivos de este Proyecto de Título.

Las hipótesis que contempla este Proyecto de Título, es que las series que se estudiarán son estacionarias.

Para analizar la estacionalidad, hay varios aspectos que se tienen que considerar, proponer transformaciones que ayudarán a estabilizar la varianza unas propuestas son $\log x$, \sqrt{x} , $\frac{1}{x}$ y desplazar la serie en c por nombrar algunas. Otra alternativa es, a través, de la estructura de correlación temporal que en este caso es la función de autocorrelación (ACF) y función de autocorrelación parcial (Partial ACF), estas estructuras ayudarán a proponer un modelo que se pueda ajustar mejor a esta situación.

Para esto, existen valores que nos ayudan a comparar cuales de los modelos se ajusta mejor, como es el criterio Akaike (AIC), donde el AIC más pequeño evidencia un mejor ajuste. Observando las series por separado, como muestra la siguiente figura.

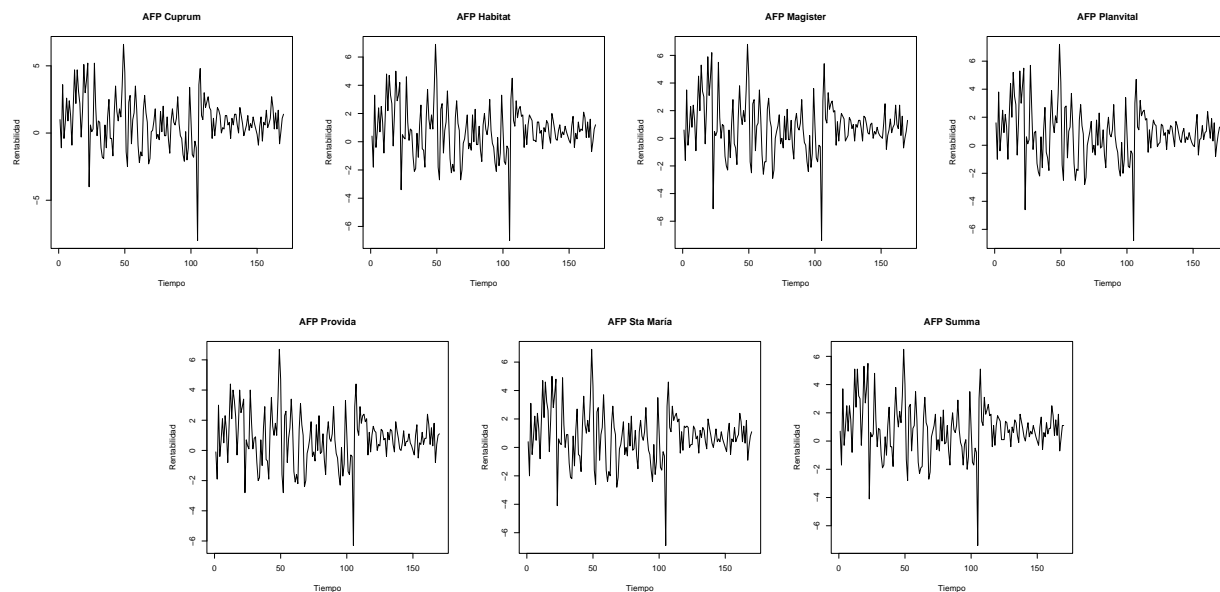


Figura 4.3: Rentabilidad AFP.

En estos Gráficos, se puede apreciar el comportamiento similar de las 7 AFP. Donde se puede observar que en todas las AFP, en el período de Noviembre 1997 y en Enero, Junio y Septiembre de 1998, las AFP muestran sus caídas mas significativas. Con Septiembre de 1998 la caída mas importante de estas, este dato atípico podría afectar la modelación en sus parámetros o quizás a validar el supuesto de normalidad.

Como las AFP muestran rentabilidades con valores negativos y además, se puede notar que la varianza de la rentabilidad de las AFP no es constante, proponer una transformación que nos ayude a homogenizar la varianza y dejar las series estacionarias, se tienen que considerar estos valores. Por lo general, para homogenizar la varianza se ocupan \sqrt{x} y $\log(x)$, pero como existen valores negativos, habrían casos indeterminados, es por eso que se trasladara esta serie en 10 unidades, un número un poco mayor al de los valores extremos que es -7.4, para así tener valores positivos y homogenizar la serie con el $\log(x)$. A continuación, se muestran las series transformadas:

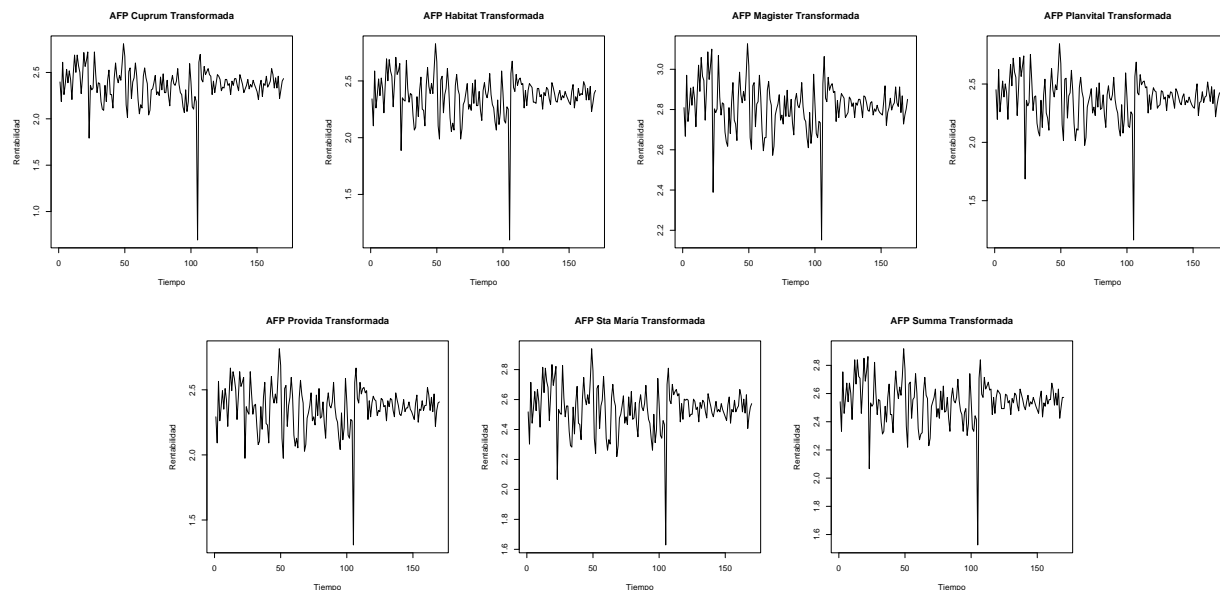


Figura 4.4: Transformación AFP.

En la Figura 4.4 se puede apreciar, que si bien la serie esta más estacionaria, aún se puede apreciar que la caída de septiembre 1998 afecta, al intentar homogenizar la varianza.

Luego de analizar la estacionariedad de las AFP, propondremos un modelo que represente esta situación. Si consideramos las funciones ACF y Partial ACF, los resultados son los siguientes.

Se observa que la memoria del proceso es de un retardo. Como se presento anteriormente en la Figura 4.5 de la rentabilidad de las AFP se puede observar una cierta estacionalidad, propia de los modelos de media móvil. Ahora bien, se analizará que sucede con la función de autocorrelación

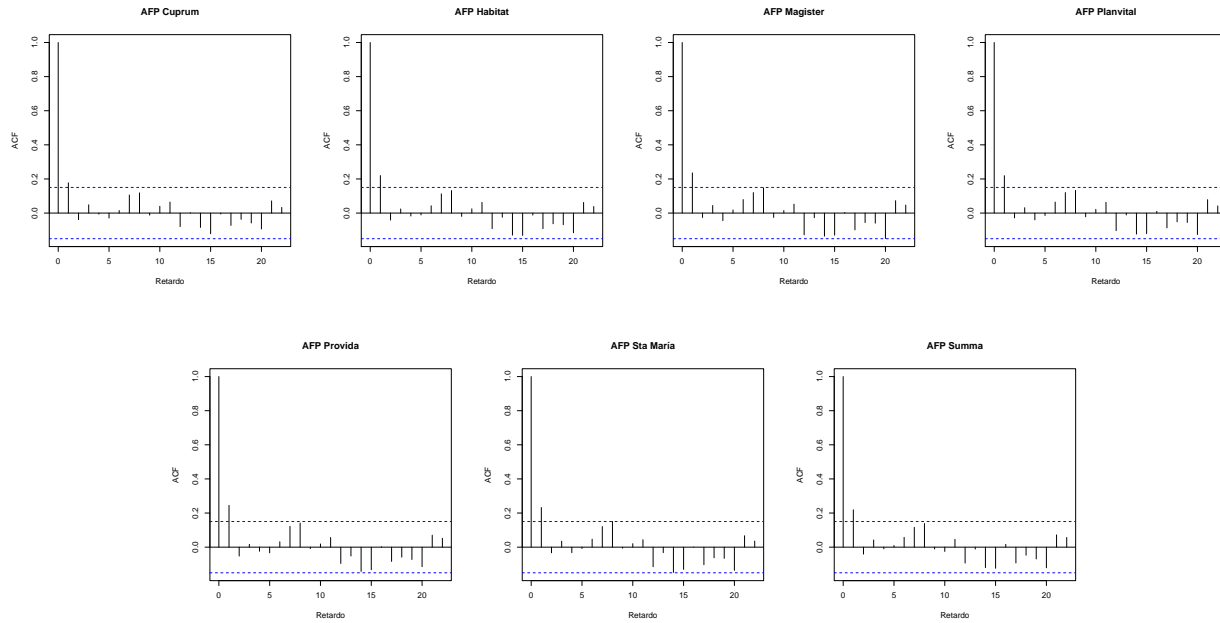


Figura 4.5: Función de Autocorrelación AFP.

parcial

En base a la Figura 4.6 se pueden proponer algunos modelos a priori, para representar la rentabilidad de las AFP.

La parte autoregresiva contribuye en 2 lag o retardos, Por otra parte, la estructura de media móvil contribuye con 1 lag o retardo. Acontinuación algunas propuestas de modelos para esta situación.

Modelos Propuestos	AIC
ARMA(1,1)	-60.71
ARMA(2,1)	-58.73
AR(1)	-60.98
MA(1)	-61.84

Si se observa el AIC de los modelos propuestos, el modelo que minimiza la varianza de los residuos es el modelo de media móvil MA(1). Las AFP son modelados a través de MA(1), desde aquí, se puede analizar supuesto de normalidad.

4.11. Estimación de parámetros MA(1) y análisis de residuos

En esta seccion se presentaran las estiamciones de los coeficiente de los modelos MA(1) para cada una de las AFP.

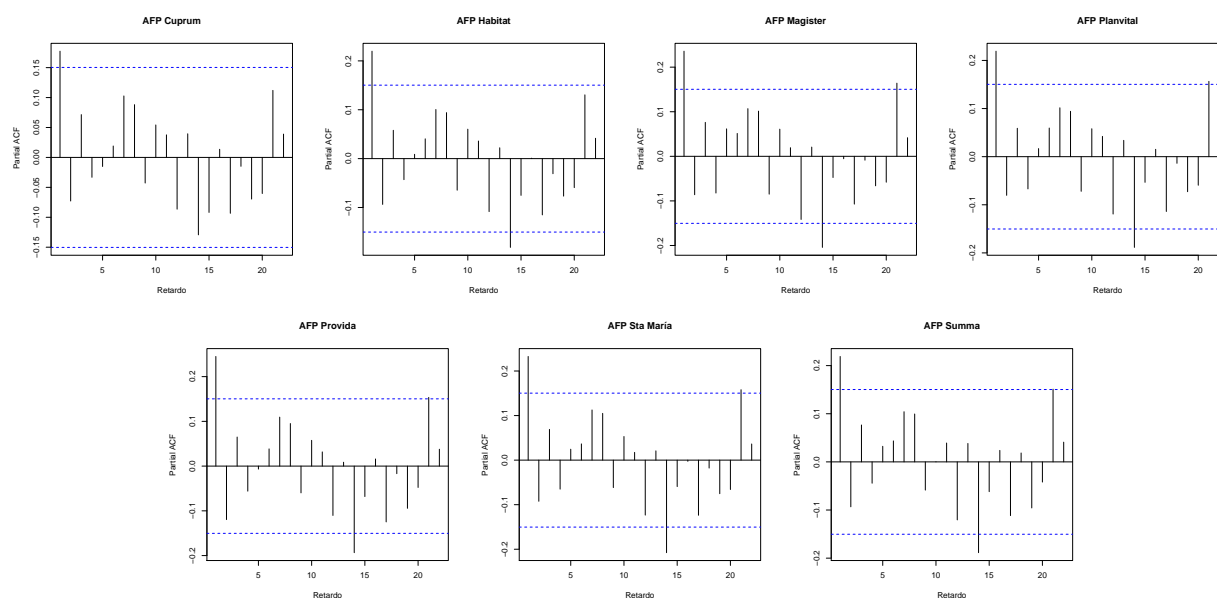


Figura 4.6: Función de Autocorrelación Parcial AFP.

Estimación de parámetros AFP Cuprum

Parámetros	MA(1)	Intercepto
Coefficiente	-0,9684	-1,00E-04
s.e	0,0356	3,00E-04
$\hat{\sigma}^2$	Log Ver	AIC
0,01031	145,37	-284,75

Estimación de parámetros AFP Habitat

Parámetros	MA(1)	Intercepto
Coefficiente	-0,9675	-1,00E-04
s.e	0,0371	3,00E-04
$\hat{\sigma}^2$	Log Ver	AIC
0,0084	162,69	-319,39

Estimación de parámetros AFP Provida

Parámetros	MA(1)	Intercepto
Coeficiente	-0,9629	-1,00E-04
s.e	0,0379	2,00E-04
$\hat{\sigma}^2$	Log Ver	AIC
0,003579	234,85	-463,71

Estimación de parámetros AFP Planvital

Parámetros	MA(1)	Intercepto
Coeficiente	-0,9621	-2,00E-04
s.e	0,0346	3,00E-04
$\hat{\sigma}^2$	Log Ver	AIC
0,009118	155,84	-305,69

Estimación de parámetros AFP Sta. María

Parámetros	MA(1)	Intercepto
Coeficiente	-0,96	4,20E-03
s.e	0,0437	6,40E-03
$\hat{\sigma}^2$	Log Ver	AIC
3,128	-337,44	680,88

Estimación de parámetros AFP Summa

Parámetros	MA(1)	Intercepto
Coeficiente	0,2786	1,27E+00
s.e	0,0805	7,10E-03
$\hat{\sigma}^2$	Log Ver	AIC
0,005291	204,29	-402,58

Estimación de parámetros AFP Magister

Parámetros	MA(1)	Intercepto
Coeficiente	-0,9626	-1,00E-04
s.e	0,0346	3,00E-04
$\hat{\sigma}^2$	Log Ver	AIC
0,005998	191,22	-376,44

Test de bondad de ajuste En esta parte se presentará la verificación del supuesto de normalidad de los residuos para cada modelo a través del test de Shapiro Wilks, presentando su estadística y el valor- p .

AFP	W	p-valor
Cuprum	0,7983	0
Habitat	0,8817	0
Provida	0,8946	0
Planvital	0,8924	0
Sta. Maria	0,9773	0,007089
Summa	0,9095	0
Magister	0,8946	0

A continuación se presentan histogramas de los residuos de las 7 AFP.

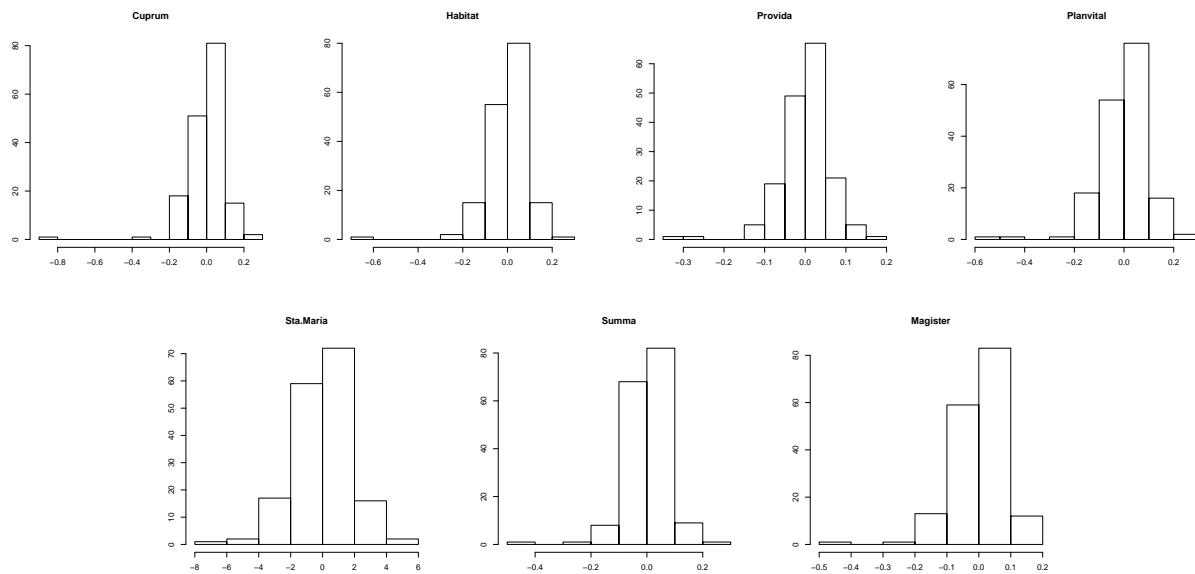


Figura 4.7: Histograma de los residuos AFP.

4.12. Análisis de Cluster

En esta sección se implementará el algoritmo de clasificación para series temporales con el Índice de Disimilitud Adaptativo para distancias convencionales.

Para entender mejor los dendogramas se han definido los siguientes valores que estarán referenciando a las AFP.

- 1 = AFP CUPRUM, 2 = AFP HABITAT, 3 = AFP PROVIDA, 4 = AFP PLANVITAL, 5 = AFP STA MARIA, 6 = AFP SUMMA y 7 = AFP MAGISTER

Aplicando el algoritmo de clasificación para series temporales con las distancias convencionales se tiene que:

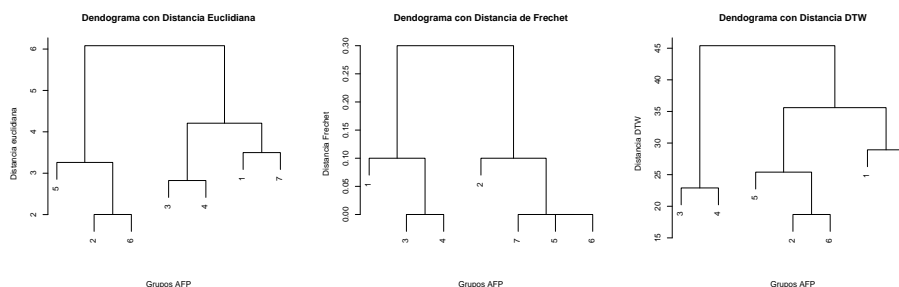


Figura 4.8: Dendogramas.

En estos gráficos se puede apreciar los distintos grupos con las tres distancias: La distancia euclidiana forma dos grupos:

Grupo 1	Grupo 2
(2)(6)(5)	(3)(4)(3)(1)(7)

La distancia de Frechét agrupa las AFP:

Grupo 1	Grupo 2
(3)(4)(1)	(2)(7)(5)(6)

La distancia DTW agrupa las AFP:

Grupo 1	Grupo 2
(3)(4)	(2)(6)(3)(1)(7)

4.12.1. Camino Distorsionado entre AFP

La siguiente Figura muestran la alineación entre las AFP, se presentan algunas combinaciones de la AFP CUPRUM con el resto de las AFP, y se destaca presencia de una perfecta alineación, es decir, las unidades experimentales entre par de punto son similares.

Esta gran similitud se debe al efecto manada, el cual ha sido detectado usando el Índice de Comovimiento.

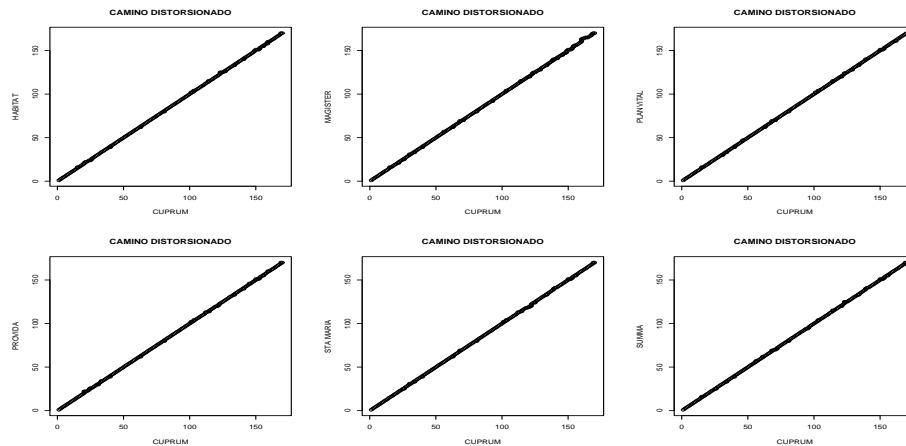


Figura 4.9: Alineamiento entre AFP.

4.12.2. Aplicación Índice de Disimilaridad con Distancia Euclidiana

Si se calcula el Índice usando la distancia euclidiana y se aplica el algoritmo de clasificación se puede observar que de acuerdo a la definición esta calcula y reúne las AFP que están más cerca respecto a cada una de las unidades experimentales, este algoritmo se calculo con $h = 1, 2$ y $k = 1, 2, 3$.

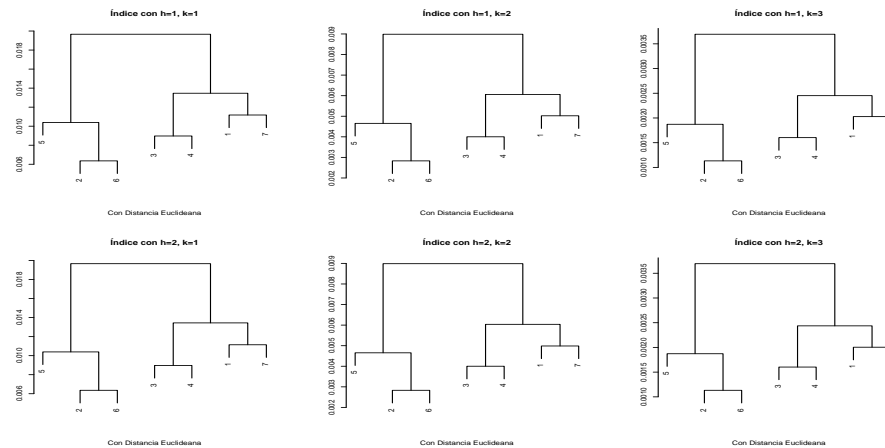


Figura 4.10: Dendrogramas de las AFP con Índice usando δ_E con $h = 1, 2$ y $k = 1, 2, 3$.

Los dendrogramas con el Índice Adaptativo no muestra cambios respecto de la distancia euclidiana.

4.12.3. Aplicación Índice de Disimilaridad con Distancia Frechét

Este Índice agrupa las AFP considerando las distancias más pequeñas dentro de las diferencias máximas de cada unidad experimental. Se puede apreciar que este Índice presenta algunas variaciones cuando $h = 2$, esta variación de agrupación es solo el intercambio de prioridad del algoritmo de Clasificación.

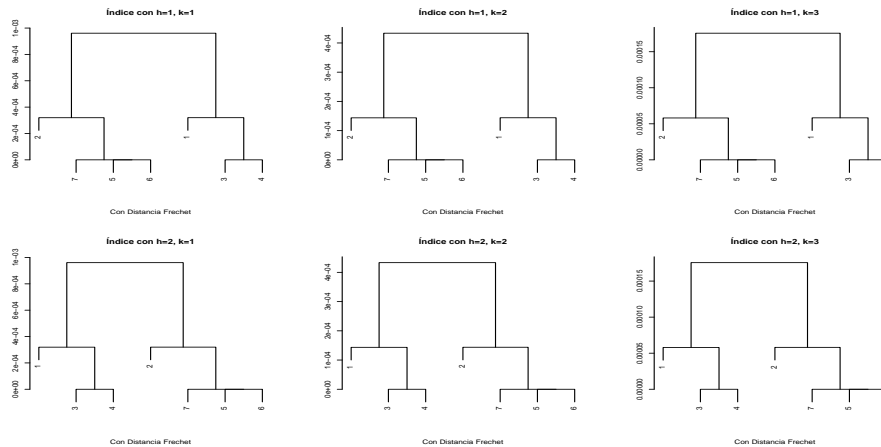


Figura 4.11: Dendrogramas de las AFP con Índice usando δ_F con $h=1,2$ y $k=1,2,3$

En general este Índice no presenta, cambios respecto de su distancia convencional.

4.12.4. Aplicación Índice de Disimilaridad con Distancia DTW

Con el Índice de Disimilaridad adaptativo usando la distancia δ_{DTW} , agrupa las AFP que están más alineadas. Este Índice no presenta cambios respecto de su medida convencional.

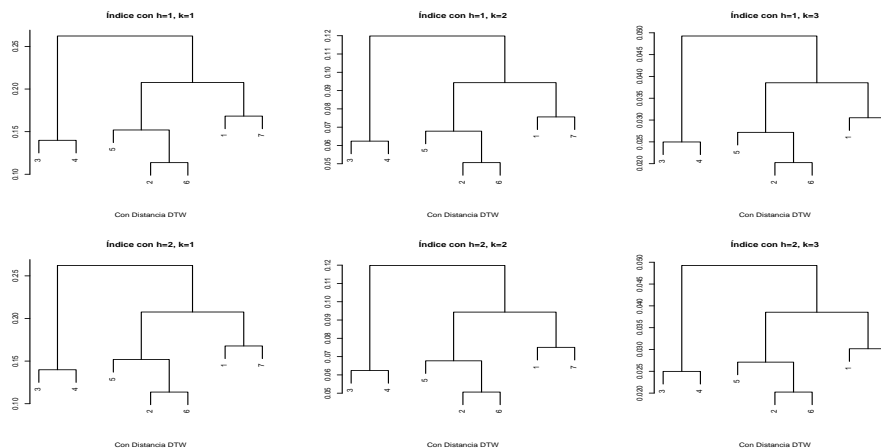


Figura 4.12: Dendrogramas de las AFP con Índice usando δ_{DTW} con $h=1,2$ y $k=1,2,3$

4.13. Conclusión

En este capítulo se pudo conocer como se comportan las AFP, el análisis descriptivo muestra evidencia de que las AFP tienen rentabilidad promedio bastante parecidas, así como en la desviación estándar. Seguidamente, se ha propuesto un modelo que mejor representa el comportamiento de la rentabilidad, basados en la información del AIC que son los modelos de media móvil de orden 1, es decir MA(1). Además, se han calculado los parámetros para cada una de las 7 AFP y se analizó los residuos, donde la normalidad se ve afectada fuertemente por la crisis asiática en los años de 1997-1998.

En el análisis de cluster, el Índice de Disimilaridad Adaptativo ha presentado pocos cambios respecto de sus medidas convencionales, dado la fuerte similitud de las AFP y en consecuencia, no ha podido detectar diferencias tanto en su nivel de comovimiento y de cercanía.

Capítulo 5

Conclusiones

Las ventajas que tiene el Índice de Comovimiento es la capacidad de capturar la información respecto del comovimiento conjunto de las series, esto se ha podido observar en el sistema Chileno de AFP, donde deja en evidencia que la rentabilidad en el tiempo de las 7 AFP en estudio tienen el mismo comportamiento, lo interesante de cuantificar este fenómeno es también, buscar las posibles causas de este fenómeno. Actualmente el modelo del sistema de pensiones ha sido cuestionado debido, a la poca competencia por abrir e investigar en otros mercados bursátiles, ya que las AFP en estudios invierten en las mismas acciones. Otra causa a considerar es la crisis económica mundial, donde se ha creado un ambiente de desconfianza entre los inversionistas, esto motiva a que las personas sean más conservadoras en las inversiones.

También se ha estudiado el Índice de Disimilaridad y se ha presentado la definición y las propiedades de algunas medidas convencionales ampliamente usadas.

Se ha estudiado el comportamiento de las AFP y se ha podido determinar que esta rentabilidad es representada por modelos de media móvil MA(1), esto motiva el estudio de una representación paramétrica asociada a esta medida de similaridad.

En la simulación se ha demostrado que las medidas convencionales δ_E , δ_M , δ_F y δ_{DTW} ignoran la relación de interdependencia entre las series y son primordialmente basadas en la proximidad con relación a los valores. Para evitar esta limitación, se ha estudiado el Índice de Disimilaridad que tiene en cuenta la proximidad con relación al comportamiento y con relación a los valores.

Esta medida de Disimilaridad es una función del Índice de Comovimiento y se ha mostrado una medida de similaridad bastante razonable en relación a los valores, la Disimilaridad propuesta es definida como una función adaptable de afinación que balancea la proximidad con relación a los valores y la proximidad con relación al comportamiento.

En la simulación se ha estudiado la contribución del comportamiento y componentes de valores para el Índice de Disimilaridad y se comparó con las medidas convencionales.

Las desventajas del Índice de Comovimiento es que para series no estacionarias aún no se ha estudiado un estimador adecuado. Además, la presencia de outliers afecta fuertemente la estimación

de su comovimiento, lo que provocaría sesgo en las estimaciones dando conclusiones erróneas.

Por otra parte el Índice de Comovimiento o Codispersión tiene su extensión para procesos espaciales intrínsecamente estacionarios, como son los procesos autoregresivos, media móvil y ARMA espacial.

En este proyecto de título, no se a estudiado una version no paramétrica, debido a que este coeficiente esta basado el estimador de momentos de $\rho_{X,Y}$, pero con esto, no se dice que no se pueda implementar un coeficiente de codispersión no paramétrico, pero esto motivaría otro estudio.

Las aplicaciones que tiene el coeficiente de codispersión se pueden ver en Vallejos. (2008) *Similarity Coefficients for Spatial or Temporal Sequences*.

En trabajos posteriores se podría estudiar la robustez de Coeficiente de Codispersión, para estar seguro que los distintos grados de comovimiento estén bien representados y evitar sobre estimaciones de la codispersión. Además, estudiar la estacionariedad de las series, proponer y estudiar en Coeficiente de Codispersión para series no estacionarias. Otro aspecto a considerar es la comparación de el Índice de Disimilaridad Adaptativo $D(X_t, Y_t)$ con diferentes funciones que ayuden a detectar y modular pequeñas diferencias en el comportamiento de las series.

Bibliografía

- [1] Rukhin and Vallejos. (2008) Assessing the Asociation Between Two Spatial or Temporal Sequences. Journal of Applied Statistics. Pág.1-2.5-7. Página web www.deuv.cl/vallejos
- [2] Vallejos. (2008) Similarity Coefficients for Spatial or Temporal Sequences. Journal of Applied Statistics. Pág.1-4
- [3] A.D. Choaukria and P.N. Nagabbhushan. (2007) Adaptive dissimilarity index for measuring time series proximity. Springer-Verlag. Pág. 1-17.
- [4] T. Warren Liao. (2005) Clustering of time data- a survey. The Journal of de Pattern Recognition Society. Pág.1-6
- [5] Dallas E.Johnson. (1998) Métodos Multivariados Aplicados Al Análisis de Datos. International Thompson Editores. México. Pág.323-329.
- [6] E. Keogh and A. Ratanamahatana. (2004) Exact indexing of dynamic time warping. Springer-Verlag. Pág. 1-4
- [7] T. Eiter and H. Mannila. (1994). Computing Discrete Frechet Distance. Technischen Universität Wien. Pág. 1-5
- [8] S. Salvador and P. Chan (2006) FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space. Institute of Technology. Melbourne. Pág. 1-2.
- [9] E. Walker (2006). Aspectos Financieros del Sistema de AFP. Universidad Católica de Chile. Pág.1-3.
- [10] Memoria Corporativa (2007). Asociacion Gremial de Administradores de Fondos de Pensiones. Pág. 4-6.7-11.
- [11] Ngai Hang Chan, Time Series: Applications to Finance, Willey Series in probability and Statistics.(2002). Pág. 39-54.
- [12] Dag Tjøstheim. A measure of association for spatial variables. Biometrika (1978). Pág. 109-114.

Anexo

Rutinas para R

```
#####  
library(dtw)  
  
#####Base de datos AFP#####  
X<-matrix(0,ncol=7,nrow=170)  
X<-read.table(file='C:/Users/Juan Carlos Herrera/  
Documents/Seminario de Tesis/Datos afp.txt')  
X<-as.matrix(X)  
ts.plot(X,gpars=list(xlab="Tiempo", ylab="",main="Rentabilidad AFP" ))  
boxplot(as.data.frame(X))  
#####  
  
X<-read.table(file='C:/Users/Juan Carlos Herrera/Documents  
/Seminario de Tesis/Datos_CUPRUM.txt')  
X<-as.matrix(X)  
  
#####Vectorizamos#####  
l<-dim(X)[1]  
v1<-matrix(0,nrow=1,ncol=1)  
v2<-matrix(0,nrow=1,ncol=1)  
v3<-matrix(0,nrow=1,ncol=1)  
v4<-matrix(0,nrow=1,ncol=1)  
v5<-matrix(0,nrow=1,ncol=1)  
v6<-matrix(0,nrow=1,ncol=1)  
v7<-matrix(0,nrow=1,ncol=1)  
  
for(i in 1:l)  
{  
  v1[i]<-X[i,1]  
  v2[i]<-X[i,2]  
  v3[i]<-X[i,3]  
  v4[i]<-X[i,4]  
  v5[i]<-X[i,5]
```



```

    v6[i]<-X[i,6]
    v7[i]<-X[i,7]
  }
#####
####Algunas funciones para calcular distancia#####
#####Distancia Euclidiana#####
de<-function(a,b)
{
  n<-length(a)
  s<-0
  vec<-matrix(0,n,1)
  for(i in 1:n)
  {
    vec[i]<-(a[i]-b[i])^2
    s<-sqrt(sum(vec))
  }
  s
}
#####Distancia de Frechet#####
d<-function(a,b)
{
  c<-abs(a-b)
  c
}

x<-c(1.4,2.3,3.09,4.3,1.3)
y<-c(6.02,7.1,9.01,0,1,6,2.1,3.3)

df<-function(x,y)
{
  p<-length(x)
  q<-length(y)
  ca<-matrix(-1,nrow=p,ncol=q)
  c<-matrix(0,nrow=p,ncol=q)
  ca[1,1]<-d(x[1],y[1])
  k<-1
  l<-1
  s<-0
  for(i in 2:p)
  {
    ca[i,k]<-max(ca[i-1,k],d(x[i],y[k]))
  }
  for(j in 2:q)
  {
    ca[l,j]<-max(ca[l,j-1],d(x[l],y[j]))
  }
}

```

```

}
for(i in 2:p)
{
  for(j in 2:q)
  {
    ca[i,j]<-min(max(ca[i-1,j],ca[i-1,j-1],ca[i,j-1]),d(x[i],y[j]))
  }
}
for(i in 1:p)
{
  for(j in 1:q)
  {
    if(ca[i,j]>-1)
    {
      a<-ca[i,j]
      a
    }
    else
    {
      s<-100000
      s
    }
  }
}
}
a
}

#####Coeficiente de Codispersion#####

cod<-function(a,b,h)
{
  n<-length(a)
  m<-length(b)
  vec<-matrix(0,n,1)
  vec1<-matrix(0,n,1)
  vec2<-matrix(0,n,1)
  s1<-0
  s2<-0
  s3<-0
  r<-0
  for(i in 1:n-h)
  {
    vec[i]=(a[i+h]-a[i])*(b[i+h]-b[i])
    vec1[i]=(a[i+h]-a[i])^2
    vec2[i]=(b[i+h]-b[i])^2
  }
}

```

```

        s1<-sum(vec)
        s2<-sqrt(sum(vec1))
        s3<-sqrt(sum(vec2))
    }
r<-(s1/(s2*s3))
r
}
#####Indice de disimilaridad#####
D<-function(a,b,h,k)
{
l<-length(a)
p1<-cod(a,b,h)
p2<-de(a,b)/(l)
v<-(2/(1+exp(p1*k)))
f<-(v*p2)
f
}

#####

D1<-function(a,b,h,k)
{
l<-length(a)
p1<-cod(a,b,h)
p2<-dtw(a,b)$distance/(l)
v<-(2/(1+exp(k*p1)))
f<-(v*p2)
f
}

#####

D2<-function(a,b,h,k)
{
l<-length(a)
p1<-cod(a,b,h)
p2<-df(a,b)/l
v<-(2/(1+exp(k*p1)))
f<-(v*p2)
f
}

#####Funcion de disimilaridad que transforma matriz de
distancia en objeto dist#####
#h=h1

```

```

#t=h2
#k=k

cl<-function(A,B,C,V,E,F,G,h,k)
{
a<-D(A,B,h,k)
b<-D(A,C,h,k)
y<-D(A,V,h,k)
d<-D(A,E,h,k)
m<-D(A,F,h,k)
l<-D(A,G,h,k)
e<-D(B,C,h,k)
g<-D(B,V,h,k)
h<-D(B,E,h,k)
o<-D(B,F,h,k)
p<-D(B,G,h,k)
q<-D(C,V,h,k)
z<-D(C,E,h,k)
t<-D(C,F,h,k)
u<-D(C,G,h,k)
n<-D(V,E,h,k)
r<-D(V,F,h,k)
s<-D(V,G,h,k)
w<-D(E,F,h,k)
f1<-D(E,G,h,k)
w1<-D(F,G,h,k)

A1<-matrix(c(0,a,b,y,d,m,l),7,1)
B2<-matrix(c(a,0,e,g,h,o,p),7,1)
C3<-matrix(c(b,e,0,q,z,t,u),7,1)
D4<-matrix(c(y,g,q,0,n,r,s),7,1)
P5<-matrix(c(d,h,z,n,0,w,f1),7,1)
P6<-matrix(c(m,o,t,r,w,0,w1),7,1)
P7<-matrix(c(l,p,u,s,f1,w1,0),7,1)

MD<-matrix(cbind(A1,B2,C3,D4,P5,P6,P7),7,7)
md<-as.dist(MD)
md
}

#####

cl1<-function(A,B,C,V,E,F,G,h,k)
{

```

```

a<-D1(A,B,h,k)
b<-D1(A,C,h,k)
y<-D1(A,V,h,k)
d<-D1(A,E,h,k)
m<-D1(A,F,h,k)
l<-D1(A,G,h,k)
e<-D1(B,C,h,k)
g<-D1(B,V,h,k)
h<-D1(B,E,h,k)
o<-D1(B,F,h,k)
p<-D1(B,G,h,k)
q<-D1(C,V,h,k)
z<-D1(C,E,h,k)
t<-D1(C,F,h,k)
u<-D1(C,G,h,k)
n<-D1(V,E,h,k)
r<-D1(V,F,h,k)
s<-D1(V,G,h,k)
w<-D1(E,F,h,k)
f1<-D1(E,G,h,k)
w1<-D1(F,G,h,k)

```

```

A1<-matrix(c(0,a,b,y,d,m,l),7,1)
B2<-matrix(c(a,0,e,g,h,o,p),7,1)
C3<-matrix(c(b,e,0,q,z,t,u),7,1)
D4<-matrix(c(y,g,q,0,n,r,s),7,1)
P5<-matrix(c(d,h,z,n,0,w,f1),7,1)
P6<-matrix(c(m,o,t,r,w,0,w1),7,1)
P7<-matrix(c(l,p,u,s,f1,w1,0),7,1)

```

```

MD<-matrix(cbind(A1,B2,C3,D4,P5,P6,P7),7,7)
md<-as.dist(MD)
md
}

```

```
#####
```

```

cl2<-function(A,B,C,V,E,F,G,h,k)
{
a<-D2(A,B,h,k)
b<-D2(A,C,h,k)
y<-D2(A,V,h,k)
d<-D2(A,E,h,k)
m<-D2(A,F,h,k)

```

```

l<-D2(A,G,h,k)
e<-D2(B,C,h,k)
g<-D2(B,V,h,k)
h<-D2(B,E,h,k)
o<-D2(B,F,h,k)
p<-D2(B,G,h,k)
q<-D2(C,V,h,k)
z<-D2(C,E,h,k)
t<-D2(C,F,h,k)
u<-D2(C,G,h,k)
n<-D2(V,E,h,k)
r<-D2(V,F,h,k)
s<-D2(V,G,h,k)
w<-D2(E,F,h,k)
f1<-D2(E,G,h,k)
w1<-D2(F,G,h,k)

```

```

A1<-matrix(c(0,a,b,y,d,m,l),7,1)
B2<-matrix(c(a,0,e,g,h,o,p),7,1)
C3<-matrix(c(b,e,0,q,z,t,u),7,1)
D4<-matrix(c(y,g,q,0,n,r,s),7,1)
P5<-matrix(c(d,h,z,n,0,w,f1),7,1)
P6<-matrix(c(m,o,t,r,w,0,w1),7,1)
P7<-matrix(c(l,p,u,s,f1,w1,0),7,1)

```

```

MD<-matrix(cbind(A1,B2,C3,D4,P5,P6,P7),7,7)
md<-as.dist(MD)
md
}

```

```

#####Funcion que transforma matriz de distancia
en objeto dist para DE#####

```

```

m_d<-function(A,B,C,V,E,F,G)
{
a<-de(A,B)
b<-de(A,C)
y<-de(A,V)
d<-de(A,E)
m<-de(A,F)
l<-de(A,G)
e<-de(B,C)
g<-de(B,V)
h<-de(B,E)
o<-de(B,F)

```

```

p<-de(B,G)
q<-de(C,V)
z<-de(C,E)
t<-de(C,F)
u<-de(C,G)
n<-de(V,E)
r<-de(V,F)
s<-de(V,G)
w<-de(E,F)
f1<-de(E,G)
w1<-de(F,G)

A1<-matrix(c(0,a,b,y,d,m,l),7,1)
B2<-matrix(c(a,0,e,g,h,o,p),7,1)
C3<-matrix(c(b,e,0,q,z,t,u),7,1)
D4<-matrix(c(y,g,q,0,n,r,s),7,1)
P5<-matrix(c(d,h,z,n,0,w,f1),7,1)
P6<-matrix(c(m,o,t,r,w,0,w1),7,1)
P7<-matrix(c(l,p,u,s,f1,w1,0),7,1)

MD<-matrix(cbind(A1,B2,C3,D4,P5,P6,P7),7,7)
md<-as.dist(MD)
md
}
#####
m_df<-function(A,B,C,V,E,F,G)
{
a<-df(A,B)
b<-df(A,C)
y<-df(A,V)
d<-df(A,E)
m<-df(A,F)
l<-df(A,G)
e<-df(B,C)
g<-df(B,V)
h<-df(B,E)
o<-df(B,F)
p<-df(B,G)
q<-df(C,V)
z<-df(C,E)
t<-df(C,F)
u<-df(C,G)
n<-df(V,E)
r<-df(V,F)

```

```

s<-df(V,G)
w<-df(E,F)
f1<-df(E,G)
w1<-df(F,G)

A1<-matrix(c(0,a,b,y,d,m,l),7,1)
B2<-matrix(c(a,0,e,g,h,o,p),7,1)
C3<-matrix(c(b,e,0,q,z,t,u),7,1)
D4<-matrix(c(y,g,q,0,n,r,s),7,1)
P5<-matrix(c(d,h,z,n,0,w,f1),7,1)
P6<-matrix(c(m,o,t,r,w,0,w1),7,1)
P7<-matrix(c(l,p,u,s,f1,w1,0),7,1)

MD<-matrix(cbind(A1,B2,C3,D4,P5,P6,P7),7,7)
md<-as.dist(MD)
md
}
#####
m_dtw<-function(A,B,C,V,E,F,G)
{
a<-dtw(A,B)$distance
b<-dtw(A,C)$distance
y<-dtw(A,V)$distance
d<-dtw(A,E)$distance
m<-dtw(A,F)$distance
l<-dtw(A,G)$distance
e<-dtw(B,C)$distance
g<-dtw(B,V)$distance
h<-dtw(B,E)$distance
o<-dtw(B,F)$distance
p<-dtw(B,G)$distance
q<-dtw(C,V)$distance
z<-dtw(C,E)$distance
t<-dtw(C,F)$distance
u<-dtw(C,G)$distance
n<-dtw(V,E)$distance
r<-dtw(V,F)$distance
s<-dtw(V,G)$distance
w<-dtw(E,F)$distance
f1<-dtw(E,G)$distance
w1<-dtw(F,G)$distance

A1<-matrix(c(0,a,b,y,d,m,l),7,1)
B2<-matrix(c(a,0,e,g,h,o,p),7,1)

```



```

C3<-matrix(c(b,e,0,q,z,t,u),7,1)
D4<-matrix(c(y,g,q,0,n,r,s),7,1)
P5<-matrix(c(d,h,z,n,0,w,f1),7,1)
P6<-matrix(c(m,o,t,r,w,0,w1),7,1)
P7<-matrix(c(l,p,u,s,f1,w1,0),7,1)

MD<-matrix(cbind(A1,B2,C3,D4,P5,P6,P7),7,7)
md<-as.dist(MD)
md
}
#####
m_cod<-function(A,B,C,V,E,F,G,h)
{
a<-cod(A,B,h)
b<-cod(A,C,h)
y<-cod(A,V,h)
d<-cod(A,E,h)
m<-cod(A,F,h)
l<-cod(A,G,h)
e<-cod(B,C,h)
g<-cod(B,V,h)
h<-cod(B,E,h)
o<-cod(B,F,h)
p<-cod(B,G,h)
q<-cod(C,V,h)
z<-cod(C,E,h)
t<-cod(C,F,h)
u<-cod(C,G,h)
n<-cod(V,E,h)
r<-cod(V,F,h)
s<-cod(V,G,h)
w<-cod(E,F,h)
f1<-cod(E,G,h)
w1<-cod(F,G,h)

A1<-matrix(c(0,a,b,y,d,m,l),7,1)
B2<-matrix(c(a,0,e,g,h,o,p),7,1)
C3<-matrix(c(b,e,0,q,z,t,u),7,1)
D4<-matrix(c(y,g,q,0,n,r,s),7,1)
P5<-matrix(c(d,h,z,n,0,w,f1),7,1)
P6<-matrix(c(m,o,t,r,w,0,w1),7,1)
P7<-matrix(c(l,p,u,s,f1,w1,0),7,1)

```

```

MD<-matrix(cbind(A1,B2,C3,D4,P5,P6,P7),7,7)
md<-as.dist(MD)
md
}
#####
#series_corr<-function(p,q)
#{
rho<-0.9
rho1<-0.7
N<-200 # number of observations
phi1<-(-0.5)
phi2<-0.3
phi3<-(-0.8)
phi4<-0.7
phi5<-0.1
phi6<-(-0.9)
phi7<-0.2
x1<-matrix(0,nrow=N,ncol=1)
x2<-matrix(0,nrow=N,ncol=1)
x3<-matrix(0,nrow=N,ncol=1)
x4<-matrix(0,nrow=N,ncol=1)
x5<-matrix(0,nrow=N,ncol=1)
x6<-matrix(0,nrow=N,ncol=1)
x7<-matrix(0,nrow=N,ncol=1)
e1<-matrix(rnorm(N,0,1),N,1)
e2<-matrix(rnorm(N,0,1),N,1)
e3<-matrix(0,N,1)
e4<-matrix(0,N,1)
e5<-matrix(rnorm(N,0,1),N,1)
e6<-matrix(rnorm(N,0,1),N,1)
e7<-matrix(rnorm(N,0,1),N,1)
e8<-matrix(rnorm(N,0,1),N,1)
for (k in 1:N)
{
  e3[k]<-rho*e1[k]+sqrt(1-rho^2)*e2[k]
  e4[k]<-rho1*e1[k]+sqrt(1-rho1^2)*e2[k]
  x1[1]<-e1[1]
  x2[1]<-e3[1]
  x3[1]<-e4[1]
  x4[1]<-e5[1]
  x5[1]<-e6[1]
  x6[1]<-e7[1]
  x7[1]<-e8[1]
}
  for (j in 2:N)

```

```

    {
      x1[j]<-phi1*x1[j-1]+e1[j]
      x2[j]<-phi2*x2[j-1]+e3[j]
      x3[j]<-phi3*x3[j-1]+e4[j]
      x4[j]<-phi4*x4[j-1]+e5[j]
      x5[j]<-phi5*x5[j-1]+e6[j]
      x6[j]<-phi6*x6[j-1]+e7[j]
      x7[j]<-phi7*x7[j-1]+e8[j]
    }
r<-matrix(cbind(x1,x2,x3,x4,x5,x6,x7),N,7)
ts.plot(r,xlab="Tiempo",ylab="Series",main="Grfico Series Simuladas")
#}
#ts.plot(r)
#####
#####
x<-as.data.frame(X)
boxplot(x,names=c("Cup","Hab","Mag","Pla","Pro","Sta Ma","Sum"),xlab="AFP",
ylab="Renatbilidad", main="Cajn con bigotes")

#####Vectorizamos#####
l<-170
v1<-matrix(0,nrow=l,ncol=1)
v2<-matrix(0,nrow=l,ncol=1)
v3<-matrix(0,nrow=l,ncol=1)
v4<-matrix(0,nrow=l,ncol=1)
v5<-matrix(0,nrow=l,ncol=1)
v6<-matrix(0,nrow=l,ncol=1)
v7<-matrix(0,nrow=l,ncol=1)

for(i in 1:l)
{
  v1[i]<-X[i,1]
  v2[i]<-X[i,2]
  v3[i]<-X[i,3]
  v4[i]<-X[i,4]
  v5[i]<-X[i,5]
  v6[i]<-X[i,6]
  v7[i]<-X[i,7]
}

v11<-ts(v1,start=c(1990,1),frequency=12)
v21<-ts(v2,start=c(1990,1),frequency=12)
v31<-ts(v3,start=c(1990,1),frequency=12)
v41<-ts(v4,start=c(1990,1),frequency=12)

```

```

v51<-ts(v5,start=c(1990,1),frequency=12)
v61<-ts(v6,start=c(1990,1),frequency=12)
v71<-ts(v7,start=c(1990,1),frequency=12)

ts.plot(v11,xlab="Tiempo",ylab="Rentabilidad",main="AFP Cuprum")
ts.plot(v21,xlab="Tiempo",ylab="Rentabilidad",main="AFP Habitat")
ts.plot(v31,xlab="Tiempo",ylab="Rentabilidad",main="AFP Magister")
ts.plot(v41,xlab="Tiempo",ylab="Rentabilidad",main="AFP Planvital")
ts.plot(v51,xlab="Tiempo",ylab="Rentabilidad",main="AFP Provida")
ts.plot(v61,xlab="Tiempo",ylab="Rentabilidad",main="AFP Sta Mara")
ts.plot(v71,xlab="Tiempo",ylab="Rentabilidad",main="AFP Summa")

#####Transformaciones#####
t0<-v1+10
t1<-log(t0^0.5)
#t2<-diff(t1)

t3<-v2+10
t4<-log(t3^0.5)
#t5<-diff(t4)

t6<-v3+16
t7<-log(t6^0.5)
#t8<-diff(t7)

t9<-v4+10
t10<-log(t9^0.5)
#t11<-diff(t10)

t12<-v5+10
t13<-log(t12^0.5)
#t14<-diff(t3)

t15<-v6+12
t16<-log(t15^0.5)
#t17<-diff(t16)

t18<-v7+12
t19<-log(t18^0.5)
#t20<-diff(t19)

ts.plot(t1,xlab="Tiempo",ylab="Rentabilidad",main="AFP Cuprum Transformada")
ts.plot(t4,xlab="Tiempo",ylab="Rentabilidad",main="AFP Habitat Transformada")
ts.plot(t7,xlab="Tiempo",ylab="Rentabilidad",main="AFP Magister Transformada")
ts.plot(t10,xlab="Tiempo",ylab="Rentabilidad",main="AFP Planvital Transformada")

```

```

ts.plot(t13,xlab="Tiempo",ylab="Rentabilidad",main="AFP Provida Transformada")
ts.plot(t16,xlab="Tiempo",ylab="Rentabilidad",main="AFP Sta Mara Transformada")
ts.plot(t19,xlab="Tiempo",ylab="Rentabilidad",main="AFP Summa Transformada")
#####Funcion ACF y PACF#####

a1<-acf(t1,xlab="Retardo",ylab="ACF",main="AFP Cuprum")
a2<-acf(t4,xlab="Retardo",ylab="ACF",main="AFP Habitat")
a3<-acf(t7,xlab="Retardo",ylab="ACF",main="AFP Magister")
a4<-acf(t10,xlab="Retardo",ylab="ACF",main="AFP Planvital")
a5<-acf(t13,xlab="Retardo",ylab="ACF",main="AFP Provida")
a6<-acf(t16,xlab="Retardo",ylab="ACF",main="AFP Sta Mara")
a7<-acf(t19,xlab="Retardo",ylab="ACF",main="AFP Summa")


pa1<-pacf(t1,xlab="Retardo",ylab=" Partial ACF",main="AFP Cuprum")
pa2<-pacf(t4,xlab="Retardo",ylab="Partial ACF",main="AFP Habitat")
pa3<-pacf(t7,xlab="Retardo",ylab="Partial ACF",main="AFP Magister")
pa4<-pacf(t10,xlab="Retardo",ylab="Partial ACF",main="AFP Planvital")
pa5<-pacf(t13,xlab="Retardo",ylab="Partial ACF",main="AFP Provida")
pa6<-pacf(t16,xlab="Retardo",ylab="Partial ACF",main="AFP Sta Mara")
pa7<-pacf(t19,xlab="Retardo",ylab="Partial ACF",main="AFP Summa")


#####Modelos Propuestos#####
m0<-arima(t1,order=c(1,0,1))
m01<-arima(t1,order=c(0,0,1))
m02<-arima(t1,order=c(2,0,1))

m1<-arima(t1,order=c(0,0,1))
m1
m2<-arima(t4,order=c(0,0,1))
m2
m3<-arima(t7,order=c(0,0,1))
m3
m4<-arima(t10,order=c(0,0,1))
m4
m5<-arima(t13,order=c(0,0,1))
m5
m6<-arima(t16,order=c(0,0,1))
m6
m7<-arima(t19,order=c(0,0,1))
m7
#####Residuos#####
tsdiag(m1)
tsdiag(m2)

```

```
tsdiag(m3)
tsdiag(m4)
tsdiag(m5)
tsdiag(m6)
tsdiag(m7)
```

```
shapiro.test(m1$residual)
shapiro.test(m2$residual)
shapiro.test(m3$residual)
shapiro.test(m4$residual)
shapiro.test(m5$residual)
shapiro.test(m6$residual)
shapiro.test(m7$residual)
```

Apéndice

Análisis Descriptivo AFP CUPRUM

En la siguiente tabla se muestra la rentabilidad de la AFP CUPRUM con las medias por año y mes.

	Mes	E	F	M	A	M	J	J	A	S	O	N	D	
Año	n	1	2	3	4	5	6	7	8	9	10	11	12	Media año
1990	1	1	-1,1	3,6	-0,4	0,8	2,6	0,9	2,4	1,3	-0,9	2	4,7	1,40
1991	2	2,2	4,7	3,1	2,2	-0,3	1,6	5,1	3	3,9	5,2	-4	0,6	2,27
1992	3	0,1	0,3	5,2	1,9	-0,2	0,9	0,8	-1,1	-1,8	-1,9	0,6	-1,1	0,30
1993	4	1,2	2,5	-0,4	-0,4	-1,7	1,5	3,5	1,6	0,9	1,8	1,2	3,5	1,26
1994	5	6,6	4,5	-1,4	-2,5	2,3	2,8	-0,8	1	1,5	3,5	1,7	-0,9	1,52
1995	6	-2,2	-1,4	-1,7	1,7	2,8	1,5	0,8	-2,3	-1,9	0,1	0,2	0,9	-0,12
1996	7	1,8	-0,4	-0,1	-0,5	1,6	0,1	2	-0,2	-0,2	1,2	-0,5	-1,5	0,27
1997	8	0,9	1,8	0,8	0,6	1,1	2,7	0,7	-0,2	-0,4	-1,6	-2,1	0,1	0,36
1998	9	-2	-1,1	3,4	1	-1,6	-1,8	-0,6	-1,1	-8	3,6	4,8	1,3	-0,17
1999	10	1,0	3,0	1,9	2,3	2,7	1,9	1,7	-0,4	1,1	-0,2	0,8	1,9	1,48
2000	11	1,7	1,3	0,0	0,4	0,3	1,3	1,3	0,6	0,8	-0,4	1,1	0,6	0,74
2001	12	1,4	1,4	0,6	0,0	1,9	1,3	0,7	-0,2	0,2	0,6	1,3	0,2	0,79
2002	13	0,7	0,4	1,2	0,7	0,3	-0,1	-0,9	0,5	1,2	-0,6	0,8	0,6	0,38
2003	14	1,7	0,4	0,7	1,1	2,7	1,9	0,3	1,5	0,3	1,7	-0,8	0,1	0,97
2004	15	1,1	1,4											1,20
	Media mes	1,14	1,17	1,20	0,58	0,90	1,30	1,10	0,36	-0,07	0,86	0,50	0,79	

Análisis Descriptivo AFP HABITAT

En la siguiente tabla se muestra la rentabilidad de la AFP HABITAT con las medias por año y mes.

	Mes	E	F	M	A	M	J	J	A	S	O	N	D	
Año	n	1	2	3	4	5	6	7	8	9	10	11	12	Media año
1990	1	0,4	-1,8	3,3	-0,4	0,8	2,4	0,7	2,5	1,4	-0,8	1,8	4,8	1,3
1991	2	2,2	4,7	3,3	2,7	-0,3	2	5	2,9	3,3	4,2	-3,4	0,5	2,3
1992	3	0,3	0,2	4,6	1,4	0,1	0,9	0,8	-1	-2,1	-1,9	0,6	-1,1	0,2
1993	4	1,2	2,6	-0,5	-0,6	-1,8	1,7	3,7	1,6	0,9	1,9	0,9	3,6	1,3
1994	5	6,9	4,4	-1,8	-2,7	2,3	2,7	-0,8	0,8	1,4	3,6	1,5	-1,2	1,4
1995	6	-2,2	-1,6	-2,1	1,7	2,9	1,3	0,8	-2,7	-2	-0,1	0,2	1	-0,2
1996	7	1,9	-0,5	-0,1	-0,6	1,9	0	2,3	-0,2	-0,2	1,1	-0,6	-1,4	0,3
1997	8	1,2	2	0,9	0,5	1,2	3	0,8	-0,1	-0,4	-1,5	-2,1	0,3	0,5
1998	9	-1,7	-0,8	3,3	1	-1,4	-1,6	-0,3	-0,5	-7	3	4,5	1,5	0,0
1999	10	1,1	2,9	1,7	2,3	2,5	1,8	1,9	-0,4	1,2	-0,2	1,2	1,9	1,5
2000	11	1,7	1,5	0,1	0,1	0,0	1,4	1,4	0,6	0,8	-0,5	1,0	0,7	0,7
2001	12	1,5	1,3	0,4	-0,1	2,0	1,5	0,9	0,2	0,1	0,8	1,2	0,3	0,8
2002	13	0,7	0,5	1,1	0,7	0,4	0,1	-0,1	1,1	1,8	-0,4	0,6	0,2	0,6
2003	14	1,4	0,7	0,9	0,8	2,1	1,7	0,3	1,4	0,3	1,6	-0,7	0,0	0,9
2004	15	0,9	1,2											1,0
	Media por mes	1,2	1,2	1,1	0,5	0,9	1,3	1,2	0,4	0,0	0,8	0,5	0,8	

Análisis Descriptivo AFP MAGISTER

En la siguiente tabla se muestra la rentabilidad de la AFP MAGISTER con las medias por año y mes.

	Mes	E	F	M	A	M	J	J	A	S	O	N	D	
Año	n	1	2	3	4	5	6	7	8	9	10	11	12	Media año
1990	1	0,6	-1,6	3,5	-0,5	0,6	2,3	0,8	2,4	1,3	-0,9	1,9	4,5	1,2
1991	2	2	5,3	3,4	3	-0,4	2,1	5,9	3,1	4,4	6,2	-5,1	0,5	2,5
1992	3	0,2	0,5	5,5	1,5	0	1	0,9	-1,3	-2	-2,3	0,6	-1,4	0,3
1993	4	1,1	2,8	-0,4	-0,7	-1,9	1,5	3,8	1,7	1	2	1,2	3,5	1,3
1994	5	6,8	4,4	-1,7	-2,5	2,5	2,8	-0,9	0,9	1,1	3,5	1,5	-1,2	1,4
1995	6	-2,6	-1,7	-1,7	2,1	2,9	1,3	0,8	-2,9	-2,3	0	0,5	0,9	-0,2
1996	7	1,7	-0,4	0	-0,7	1,6	-0,1	2,1	-0,1	-0,1	1,3	-0,6	-1,5	0,3
1997	8	1,2	1,8	0,8	0,6	1,1	2,8	0,7	-0,3	-0,5	-1,7	-2,4	0,2	0,4
1998	9	-2,1	-1,1	3,6	0,9	-1,4	-1,7	-0,5	-0,6	-7,4	2,7	5,4	1,4	-0,1
1999	10	1,1	3,3	2,1	2,5	2,7	1,9	2,0	-0,5	1,2	-0,2	1,0	1,8	1,6
2000	11	1,6	1,4	-0,2	0,0	0,2	1,5	1,3	0,7	1,0	-0,2	1,0	0,8	0,8
2001	12	1,3	1,3	0,7	-0,2	1,6	1,5	1,0	0,4	0,3	1,0	1,1	0,0	0,8
2002	13	0,5	0,3	0,8	0,4	0,2	0,1	0,0	1,1	2,5	-0,8	0,2	0,6	0,5
2003	14	1,4	0,4	0,8	1,0	2,4	1,0	0,7	2,4	0,2	1,6	-0,7	0,0	0,9
2004	15	0,5	1,3											0,9
	Media por mes	1,0	1,2	1,2	0,5	0,9	1,3	1,3	0,5	0,0	0,9	0,4	0,7	

Análisis Descriptivo AFP PLANVITAL

En la siguiente tabla se muestra la rentabilidad de la AFP PLANVITAL con las medias por año y mes.

	Mes	E	F	M	A	M	J	J	A	S	O	N	D	
Año	n	1	2	3	4	5	6	7	8	9	10	11	12	Media año
1990	1	1,6	-1	3,8	-0,4	0,9	2,5	0,9	2,2	1,3	-1	2,1	4,4	1,4
1991	2	2	5,2	3,5	2,9	-0,7	1,6	5,3	3	4,2	5,5	-4,6	0,6	2,4
1992	3	0,1	0,4	5,7	2,6	-0,3	0,9	1	-1,3	-1,9	-2,2	0,6	-1,6	0,3
1993	4	1,1	2,7	-0,5	-0,9	-1,8	1,6	3,9	1,6	0,9	2,1	1,6	3,7	1,3
1994	5	7,2	4,5	-1,4	-2,5	2,7	2,8	-0,9	1	1,3	3,7	1,5	-1,3	1,6
1995	6	-2,5	-1,7	-1,8	1,8	2,9	1,4	0,8	-2,8	-2,3	0	0,5	1	-0,2
1996	7	1,7	-0,5	0	-0,7	1,8	0	2,3	-0,2	-0,1	1,1	-0,7	-1,6	0,3
1997	8	1,1	2	0,8	0,6	1,1	2,9	0,7	-0,1	-0,5	-1,7	-2,2	0,2	0,4
1998	9	-2	-1	3,4	0,9	-1,5	-1,6	-0,4	-0,6	-6,8	3	4,7	1,3	-0,1
1999	10	1,1	3,2	2,1	2,4	2,5	1,8	1,9	-0,5	1,1	-0,2	1,0	1,8	1,5
2000	11	1,5	1,4	-0,1	0,1	0,2	1,5	1,4	0,7	0,9	-0,3	1,1	0,8	0,8
2001	12	1,4	1,3	0,6	-0,1	1,7	1,4	0,9	0,4	0,2	0,9	1,2	0,2	0,8
2002	13	0,6	0,4	0,9	0,5	0,2	0,0	-0,1	1,2	2,2	-0,7	0,4	0,5	0,5
2003	14	1,4	0,4	0,9	1,0	2,4	1,4	0,5	1,9	0,3	1,6	-0,8	0,1	0,9
2004	15	0,8	1,3											1,0
	Media por mes	1,1	1,2	1,3	0,6	0,9	1,3	1,3	0,5	0,1	0,8	0,5	0,7	

Análisis Descriptivo AFP PROVIDA

En la siguiente tabla se muestra la rentabilidad de la AFP PROVIDA con las medias por año y mes.

	Mes	E	F	M	A	M	J	J	A	S	O	N	D	
Año	n	1	2	3	4	5	6	7	8	9	10	11	12	Media año
1990	1	-0,1	-1,9	3	-0,4	0,7	2,1	0,5	2,3	1,3	-0,8	1,6	4,4	1,1
1991	2	2,1	4	3,3	2,2	-0,3	1,5	4	2,5	2,9	3,4	-2,8	0,7	2,0
1992	3	0,3	0,1	4	1,8	0,1	0,8	0,9	-0,8	-2	-1,8	0,7	-1	0,3
1993	4	1,5	2,9	-0,6	-0,7	-1,9	1,4	3,5	1,6	1	1,8	1	3,5	1,3
1994	5	6,7	4,7	-1,6	-2,8	2,3	2,6	-0,8	0,8	1,5	3,4	1,6	-1,4	1,4
1995	6	-2,1	-1,6	-2,2	1,4	3,1	1,7	1	-2,4	-2	-0,2	0,2	0,8	-0,2
1996	7	1,9	-0,4	-0,1	-0,7	1,7	0	2,3	-0,2	-0,1	1,1	-0,6	-1,6	0,3
1997	8	1	1,9	0,8	0,6	1,1	2,9	0,7	-0,2	-0,5	-1,6	-2,3	0,2	0,4
1998	9	-1,7	-0,9	3,3	1,1	-1,4	-1,6	-0,3	-0,4	-6,3	2,9	4,4	1,4	0,0
1999	10	1,0	2,9	1,8	2,3	2,4	1,9	2,1	-0,3	1,2	-0,1	1,0	1,6	1,5
2000	11	1,3	1,2	0,0	0,4	0,3	1,4	1,3	0,7	0,8	-0,4	1,2	0,7	0,8
2001	12	1,4	1,3	0,4	-0,1	1,9	1,3	0,8	0,1	0,0	0,5	1,3	0,3	0,8
2002	13	0,6	0,6	1,1	0,5	0,3	0,0	-0,3	1,0	1,7	-0,5	0,5	0,4	0,5
2003	14	1,2	0,4	0,8	0,8	2,4	1,7	0,4	1,5	0,3	1,8	-0,8	0,4	0,9
2004	15	1,0	1,1											1,0
	Media por mes	1,1	1,1	1,0	0,5	0,9	1,3	1,2	0,4	0,0	0,7	0,5	0,7	

Análisis Descriptivo AFP STA MARIA

En la siguiente tabla se muestra la rentabilidad de la AFP STA MARIA con las medias por año y mes.

	Mes	E	F	M	A	M	J	J	A	S	O	N	D	
Año	n	1	2	3	4	5	6	7	8	9	10	11	12	Media año
1990	1	0,4	-2	3,1	-0,5	0,7	2,2	0,5	2,4	1,3	-0,8	1,9	4,7	1,2
1991	2	2,1	4,6	3,2	2,6	-0,3	1,9	5	2,8	3,7	4,8	-4,1	0,6	2,2
1992	3	0,3	0,2	4,9	1,6	0	0,9	0,9	-1,1	-2,1	-2,2	0,8	-1,3	0,2
1993	4	1,2	2,7	-0,5	-0,6	-1,7	1,5	3,6	1,6	1	1,9	1,1	3,4	1,3
1994	5	6,9	4,3	-1,6	-2,6	2,5	2,8	-0,9	0,9	1,4	3,7	1,4	-1,6	1,4
1995	6	-2,4	-1,7	-2	1,6	2,9	1,4	0,9	-2,8	-2,1	-0,1	0,2	0,9	-0,3
1996	7	1,8	-0,5	0	-0,6	1,7	0,1	2,2	-0,2	-0,1	1,2	-0,6	-1,5	0,3
1997	8	1,1	1,9	0,8	0,5	1,1	2,8	0,8	-0,2	-0,5	-1,6	-2,4	0,2	0,4
1998	9	-1,9	-1	3,5	1,1	-1,4	-1,6	-0,3	-0,6	-6,9	2,9	4,6	1,4	0,0
1999	10	1,1	2,9	1,9	2,2	2,4	1,8	2,0	-0,4	1,1	-0,2	1,5	1,4	1,5
2000	11	1,5	1,4	0,0	0,2	0,2	1,5	1,3	0,6	0,8	-0,4	1,2	0,7	0,8
2001	12	1,4	1,2	0,4	-0,1	2,0	1,4	0,9	0,3	0,0	0,6	1,3	0,4	0,8
2002	13	0,6	0,4	1,1	0,6	0,3	0,0	-0,3	1,0	1,7	-0,5	0,6	0,4	0,5
2003	14	1,4	0,4	0,7	0,9	2,4	1,8	0,4	1,5	0,3	1,9	-0,9	0,1	0,9
2004	15	0,8	1,1											1,0
	Media por mes	1,1	1,1	1,1	0,5	0,9	1,3	1,2	0,4	0,0	0,8	0,5	0,7	

Análisis Descriptivo AFP SUMMA

En la siguiente tabla se muestra la rentabilidad de la AFP SUMMA con las medias por año y mes.

	Mes	E	F	M	A	M	J	J	A	S	O	N	D	
Año	n	1	2	3	4	5	6	7	8	9	10	11	12	Media año
1990	1	0,7	-1,7	3,7	-0,3	1	2,5	0,7	2,5	1,5	-0,8	2	5,1	1,4
1991	2	2,4	5,1	3,2	3	-0,3	2	5,3	2,7	3,8	5,5	-4,1	0,6	2,4
1992	3	0,3	0,5	4,8	1,5	-0,4	0,9	0,8	-1	-1,9	-1,7	0,3	-1	0,3
1993	4	1,1	2,4	-0,4	-0,4	-1,8	1,6	3,8	1,9	1	2,1	1,3	3,2	1,3
1994	5	6,5	4	-1,3	-2,8	2,4	2,6	-0,7	0,9	1,1	3,5	1,5	-1,5	1,4
1995	6	-2,3	-1,9	-1,8	1,6	3,1	1,3	1	-2,7	-2,2	0,3	0,7	1,1	-0,2
1996	7	1,9	-0,6	0,1	-0,7	1,7	0	2,2	-0,2	-0,1	1,1	-0,7	-1,7	0,3
1997	8	1	2	0,8	0,6	1,1	2,9	0,8	-0,1	-0,4	-1,7	-0,23	0,1	0,6
1998	9	-2	-1,1	3,5	0,9	-1,5	-1,7	-0,5	-0,8	-7,4	3	5,1	1,4	-0,1
1999	10	1,1	3,1	1,9	2,2	2,6	1,8	1,9	-0,4	1,1	-0,3	1,1	1,8	1,5
2000	11	1,6	1,4	0,1	0,1	0,1	1,4	1,3	0,6	0,8	-0,4	1,0	0,6	0,7
2001	12	1,5	1,3	0,4	-0,1	1,9	1,4	0,9	0,3	0,0	0,7	1,3	0,3	0,8
2002	13	0,7	0,5	1,1	0,7	0,3	0,0	-0,3	1,0	1,7	-0,6	0,6	0,3	0,5
2003	14	1,3	0,5	0,9	0,9	2,5	1,8	0,4	1,5	0,4	1,9	-0,7	0,2	1,0
2004	15	1,1	1,1											1,1
	Media por mes	1,1	1,1	1,2	0,5	0,9	1,3	1,3	0,4	0,0	0,9	0,7	0,8	