



UNIVERSITAT DE VALÈNCIA
Programa de Doctorat en Estadística i Optimització

MODELOS Y ALGORITMOS DE PREDICCIÓN FUZZY

realizada por

Abel Rubio Fornés

Tesis doctoral
en Estadística y Optimización

Dirigida por

José D. Bermúdez Edo y
Enriqueta Vercher González

en la

Facultat de Matemàtiques
Departamento de Estadística e Investigación Operativa

Septiembre 2019

Dedicat a Joan:
"Tot és possible en aquesta vida si t'ho proposes"

UNIVERSITAT DE VALÈNCIA

Resumen

Facultad de Ciencias Matemáticas
Departamento de Estadística e Investigación Operativa

Programa de Doctorado en Estadística y Optimización

En un mundo tan globalizado la complejidad macro-económica, la volatilidad en los mercados y la estabilidad político-social de un país, afectan directamente a las cotizaciones de los valores bursátiles del país de referencia y añaden complejidad a la toma de decisiones.

La riqueza de un análisis o inferencia sobre las cotizaciones de valores en mercados bursátiles tiene que servir como soporte para la toma de decisiones, por lo que no sólo han de basarse en una predicción puntual, de ahí radica la gran relevancia de las series temporales difusas. Ya que las series temporales difusas pueden tratar entornos donde hay incertidumbre en los datos (comportamientos o relaciones entre los datos) y/o conceptos expresados en términos lingüísticos (riesgo de la inversión o rendimiento esperado), muy difíciles de traducir en términos matemáticos clásicos. Adicionalmente te permiten hacer predicciones difusas (conjuntos difusos) y de éstas extraer más información que la meramente puntual (predicción por intervalos).

Dada la ambigüedad de los datos (generalmente no lineales) y la asimetría de la volatilidad a lo largo del tiempo, en esta tesis se han analizado las series temporales de cotización de ciertos índices de referencia desde un punto de vista difuso (credibilista y posibilista). Aplicando la Teoría de Conjuntos difusos se ha observado que muchos investigadores han plasmado multitud de ideas y de líneas de investigación interesantes para la aplicación de modelos de series temporales difusas.

Así en esta tesis se han presentado varios modelos de predicción. Desde un punto de vista posibilista con unos modelos basados en nuevos operadores de ponderación para la predicción de series temporales difusas y otros modelos basados en una combinación lineal convexa de dichos operadores, pero

siempre con predicciones difusas a una etapa.

Y desde un punto de vista credibilista con modelos que utilizan variables difusas, rompiendo con los esquemas clásicos de predicción de series temporales difusas. Trabajar desde este punto de vista ha permitido hacer predicciones a varias etapas y, adicionalmente, predicciones por intervalos.

En los modelos basados en series temporales difusas ponderadas, se ha propuesto utilizar un conjunto de pesos obtenidos principalmente de tres formas: la primera forma sería a partir de la secuencia cronológica de relaciones lógicas difusas, la segunda forma utilizando la información proporcionada por las relaciones lógicas difusas una a una con los saltos y en la tercera forma mediante una versión generalizada del modelo de saltos utilizando rachas. En estos modelos, para considerar la posible tendencia de las series temporales difusas, se han presentado un conjunto de ponderaciones que miden la frecuencia relativa y la magnitud de los saltos observados en la serie temporal difusa. Estos modelos también proporcionan números difusos trapezoidales como resultado del proceso de predicción.

Nuestro enfoque ha sido probado utilizando datos históricos de series temporales de algunos valores que componen el IBEX35 y de cuatro índices bursátiles (IBEX35, NIKKEI225 japonés, DAX30 alemán y TAIEX taiwanés), para los modelos de series temporales difusas ponderadas y la base de datos de la Competición **M4** para los modelos basados en variables difusas. En los experimentos numéricos, se han comparado nuestras propuestas con otros métodos bien conocidos de series temporales difusas, series temporales difusas ponderadas y modelos clásicos estadísticos de predicción, y, generalmente, se han conseguido buenos resultados, obteniendo una mayor precisión de predicción.

También se ha presentado un predictor y un SAD (sistema de ayuda a la decisión) para el análisis de series temporales difusas, con las diferentes estrategias o modelos incluidos en el árbol de decisión. Aplicándose a series de cotizaciones diarias de índices del mercado de valores y obteniéndose predicciones difusas (números difusos trapezoidales) de una etapa, donde se han observado resultados prometedores.

Índice general

Lista de figuras	V
Lista de tablas	VII
1. Introducción	1
1.1. Teoría de Conjuntos difusos	6
1.2. Series Temporales difusas (STD)	13
1.3. Modificación del esquema clásico de STD	19
2. Predicción difusa de rendimientos de carteras	33
2.1. Formulación del problema de selección de carteras	34
2.1.1. Aproximación difusa al problema de selección de carteras	38
2.2. Predicción puntual en series temporales difusas	42
2.3. Predicción difusa del rendimiento de una cartera	50
2.3.1. Nuevo algoritmo propuesto	50
2.4. Experimentos numéricos	55
2.4.1. Análisis de una cartera	56
2.4.2. Análisis comparativo	59
2.5. Selección de carteras	64
2.6. Conclusiones	66
3. Series temporales con tendencia	69
3.1. Métodos de ponderación de series temporales con tendencia . .	70
3.1.1. Nueva propuesta de ponderación	71
3.1.2. Reglas heurísticas para la predicción difusa	72
3.1.3. Aproximación del parámetro γ	74
3.1.4. Predicción de una etapa: un ejemplo	74

3.2.	Experimentos numéricos	77
3.2.1.	Predicción del índice IBEX35	79
3.2.2.	Predicción del índice NIKKEI225	83
3.2.3.	Predicción del índice DAX30	86
3.2.4.	Predicción del índice TAIEX	89
3.3.	Conclusiones	92
4.	Minería de datos en series temporales	95
4.1.	Rachas y patrones	96
4.1.1.	Predicción difusa con patrones	97
4.1.2.	Ejemplo de aplicación del modelo de patrones	99
4.1.3.	Suavizado de la matriz de patrones	101
4.2.	Sistema de ayuda a la decisión (SAD)	105
4.3.	Experimentos numéricos	108
4.3.1.	Predicción del índice DAX30	109
4.3.2.	Predicción del índice TAIEX	112
4.4.	Conclusiones	115
5.	Predicción difusa credibilista	117
5.1.	Variable difusa y distribución de credibilidad	118
5.1.1.	Estimación de la variable difusa	120
5.1.2.	Media credibilista e intervalo de cubrimiento	121
5.2.	Modelos difusos de predicción	122
5.2.1.	Predicción de una serie: un ejemplo	125
5.3.	Experimentos numéricos	130
5.3.1.	Predicción Puntual	132
5.3.2.	Predicción Intervalar (cubrimiento)	135
5.3.3.	Coste Computacional	136
5.3.4.	Resultados de los otros modelos propuestos	137
5.4.	Conclusiones	138
6.	Conclusiones y líneas de investigación futuras	141

Índice de figuras

1.1.	Representación gráfica de un conjunto de nivel superior (α -corte)	8
1.2.	Gráfica de un número difuso trapezoidal	11
1.3.	Gráfica número difuso de tipo LR potencia donde $b=a$	12
2.1.	Trazo del conjunto de datos de rendimientos de una cartera X.	42
2.2.	Función de pertenencia de la predicción difusa \hat{A}_{153} para la serie de rendimientos de la cartera #1.	58
2.3.	Diagrama de cajas de los errores de predicción de una etapa obtenidos en el experimento de simulación de 2000 carteras.	62
2.4.	Gráfica del rendimiento esperado y riesgo lateral de carteras generadas aleatoriamente, la última población obtenida por la aplicación del algoritmo genético y la envoltura convexa.	66
3.1.	Esquema de flujo del Algoritmo 3	75
3.2.	Serie temporal de la cotización diaria del IBEX35 desde Enero del 2013 hasta Diciembre del 2015.	76
3.3.	Representación parcial de una serie temporal difusa de cotizaciones diarias del IBEX35 y la predicción difusa, y puntual de una etapa.	78
3.4.	Trazado conjunto de la serie temporal de las cotizaciones diarias de IBEX35 y de las predicciones proporcionadas por nuestro método STD de tendencia Algoritmo 3, de enero de 2015 a diciembre de 2015.	80
3.5.	Trazo de la serie temporal de cotizaciones diarias del NIK-KEI225 japonés, de enero de 2013 a diciembre de 2015.	84
3.6.	Trazo de la serie temporal de cotizaciones diarias del DAX30.	87
3.7.	Trazo de la serie temporal de cotizaciones diarias del TAIEX.	90

3.8.	Trazo de la serie temporal de cotizaciones diarias del TAIEX y predicciones, desde Enero del 2015 hasta Diciembre del 2015.	90
4.1.	Gráfica de v_0	103
4.2.	Gráfica de v_0 suavizado	104
4.3.	Esquema de flujo del Algoritmo 4	106
4.4.	Trazo de las 400 primeras observaciones de la serie temporal de las cotizaciones diarias del índice alemán de DAX30, entre los años 2011 y 2013.	109
4.5.	Representación parcial de la serie temporal difusa de 20 cotizaciones diarias del índice DAX30 alemán y la predicción difusa a una etapa.	110
4.6.	Trazo de la serie temporal de cotizaciones diarias del TAIEX.	113
5.1.	Gráfica de la variable difusa D^1	126
5.2.	Variables difusas de los conjuntos 2-diferencias y 3-diferencias	127

Índice de cuadros

2.1. Intervalos u_i y frecuencias de datos observados	43
2.2. Predicciones puntuales de nuestro método para las series temporales difusas: Algoritmo 2.	57
2.3. Representación matricial de los pesos W de las relaciones lógicas difusas de grupos correspondiente a la cartera #1.	58
2.4. Errores de ajuste obtenidos en el experimento de simulación para los métodos de STD ponderadas.	60
2.5. Precisión de las predicciones a una etapa	62
2.6. P-valores ajustados para comparaciones dos a dos de los errores de predicción puntual (MAD).	63
2.7. P-valores ajustados de la comparación por pares del sesgo de los errores de predicción puntual de una etapa.	63
3.1. Errores de ajuste de los datos de entrenamiento usando diferentes γ -valores.	76
3.2. Tamaños de datos experimentales y tiempos de computación.	79
3.3. Promedio de errores de predicción para diferentes modelos de STD, con el conjunto de datos IBEX35.	81
3.4. P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos de STD, con el conjunto de datos IBEX35.	81
3.5. Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del IBEX35.	82
3.6. P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos IBEX35.	83

3.7.	Promedio de errores de predicción del índice NIKKEI225, con conjunto de datos del año 2015 y para diferentes modelos de STD.	85
3.8.	P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos STD, con el conjunto de datos NIKKEI225 del año 2015.	85
3.9.	Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del NIKKEI225.	85
3.10.	P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos NIKKEI225.	86
3.11.	Promedio de errores de predicción del índice alemán DAX30, con conjunto de datos del año 2015 y para diferentes modelos de STD.	87
3.12.	P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos STD, con el conjunto de datos DAX30 2015.	88
3.13.	Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del DAX30.	88
3.14.	P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos DAX30.	89
3.15.	Promedio de errores de predicción del índice TAIEX, con conjunto de datos del año 2015.	91
3.16.	P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos STD.	91
3.17.	Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del DAX30.	91
3.18.	P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos TAIEX.	92
4.1.	Matriculados de la Universidad de Alabama	99
4.2.	Matriz de frecuencias de saltos asociada a rachas de longitud 2	101
4.3.	Matriz suavizada de frecuencias de saltos asociada a rachas de longitud 1	104
4.4.	Errores de predicción del índice DAX30, con conjuntos de datos de tamaño 300.	111

4.5.	P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos STD.	111
4.6.	Errores de predicción para modelos estadísticos y para el Algoritmo 4, con conjuntos de datos del DAX30 de tamaño 300.	111
4.7.	P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos estadísticos, junto con el Algoritmo 4.	112
4.8.	Errores de predicción del índice TAIEX, con conjuntos de datos de tamaño 350.	113
4.9.	P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos STD.	113
4.10.	Errores de predicción para modelos estadísticos y para el Algoritmo 4, con conjuntos de datos del TAIEX de tamaño 350.	114
4.11.	P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos estadísticos, junto con el Algoritmo 4.	114
5.1.	Serie de datos anuales y diferencias de distintos órdenes	125
5.2.	Parámetros de la variable difusa D^1	126
5.3.	Errores de predicción varias etapas	133
5.4.	Media de errores de predicción para las 6 etapas	133
5.5.	P-valores del test Wilcoxon dos a dos ajustados con el método de Holm del MAPE, para las etapas 1, 2, 3, 4, 5 y 6	134
5.6.	Cubrimiento varias etapas en tantos por uno.	135
5.7.	Estadísticos descriptivos de los tiempos de computación para las 9060 series	136
5.8.	Errores de predicción varias etapas	137
5.9.	Media de errores de predicción para las 6 etapas	137
5.10.	Cubrimiento varias etapas en tantos por uno.	138
5.11.	Estadísticos descriptivos de los tiempos de computación para las 9060 series	138

Introducción

En un mundo tan globalizado y de una creciente complejidad en las estructuras organizativas, en particular, en entornos industriales o empresariales, se ha convertido la toma de decisiones, a todos los niveles de la escala organizativa, en una herramienta de suma importancia. Siendo la tecnificación de la toma de decisiones uno de los grandes retos, no sólo presentes, sino también futuros, donde gran parte de la investigación actual, en muchos campos, tiene puesto su foco.

Esta tesis se centra en dos aspectos de gran relevancia en la toma de decisiones. El primero de los aspectos proviene del desconocimiento al trabajar con conceptos imprecisos que, generalmente, tienen naturaleza lingüística aunque también pueden provenir de la incertidumbre que envuelve los datos o de ciertas relaciones existentes entre ellos (los datos). En particular se observan las decisiones financieras que están basadas en premisas ambiguas y que devengan, en general, modelos matemáticos poco robustos (en sentido estadístico), por lo que se hace necesario nuevos modelos que manejen esa imprecisión dando apoyo a la toma de decisiones. Un factor a tener presente es que a partir de premisas e implicaciones difusas, en entornos con incertidumbre, las conclusiones que se obtienen o deducen debieran ser igualmente difusas (Rutkowska, 2002). Y el segundo de los aspectos nace de la necesidad de poseer información adicional sobre el futuro, intentando que dicha información no sólo sea puntual, englobando un concepto más amplio de predicción. Siendo todo un reto, obtener más información sin que ello perjudique a la calidad de la predicción.

Se plantea inicialmente esta tesis abordando uno de los problemas más ambiciosos de la toma de decisiones y de la lógica difusa, cuando se está en un entorno financiero y/o industrial, tomando como referencia el problema

de selección de carteras, con el fin de desarrollar metodología difusa para dar soporte a la toma de decisiones financieras. Teniendo en cuenta que para la toma de decisiones no sólo es importante saber aproximar la incertidumbre presente sino también futura, por ello se introducen las series temporales difusas (STD) en el problema de selección de carteras y se continua con el desarrollo de una serie de procedimientos para la predicción difusa que rompen con los modelos clásicos de predicción en el campo de las series temporales difusas.

Por lo tanto esta memoria tiene como objeto principal realizar predicciones difusas (a una o varias etapas) que permitan obtener predicciones puntuales o por intervalos. Un segundo objetivo de esta tesis es tratar los modelos de predicción difusa desde distintos puntos de vista (posibilista y credibilista) dentro del marco de la teoría de conjuntos difusos. Toda la tesis, tanto los algoritmos como los experimentos numéricos se implementan en el entorno de programación R (R Core Team, 2014).

La presente memoria está estructurada en seis capítulos. En este primer capítulo introductorio, se presentan el concepto de lógica difusa, el problema de selección de carteras y las series temporales. También se revisan las principales definiciones y propiedades de la teoría de conjuntos difusos, y se finaliza el presente capítulo con la introducción y revisión bibliográfica de las series temporales difusas. En un segundo capítulo se ven los principales modelos de resolución del problema de selección de carteras así como una aproximación difusa al mismo desde un punto de vista posibilista, se realiza una predicción difusa y se extraen medidas predictoras para su uso en el problema de selección de carteras. En el tercer capítulo se estudian las series con tendencia y se introduce en el modelo de predicción difusa un parámetro que estima la tendencia de la serie. En un cuarto capítulo se propone una generalización de las series temporales difusas con tendencia y se presenta un sistema de ayuda a la decisión que pretende dar soporte, ante la diversidad de modelos presentados en esta memoria, a la elección del modelo de predicción difusa más adecuado para los datos en estudio. En el quinto capítulo se cambia el enfoque de la predicción difusa, estudiándola desde un punto de vista credibilista y bajo un cambio de paradigma con respecto a los esquemas clásicos de series temporales difusas. Y finalmente, en el sexto y último capítulo, vienen recogidas las conclusiones de la tesis y algunas líneas de investigación futuras.

Lógica difusa

En un trabajo inicial, Zadeh (1965) se interesa en cómo representar sistemas complejos mediante el uso de modelos simples. Dado que las herramientas matemáticas tradicionales le resultaron inadecuadas para este propósito, Zadeh introduce el concepto de conjunto difuso, como un conjunto sin un límite definido. En este contexto el hecho de "pertenecer" o no a un conjunto tiene una transición gradual, estando caracterizada dicha transición por una función de pertenencia. Los conjuntos definidos de forma imprecisa desempeñan un papel importante en el pensamiento humano, particularmente en los dominios del reconocimiento de patrones, de la comunicación de la información y de la abstracción (Zadeh, 1975a).

Una de las principales motivaciones de la teoría de conjuntos difusos es la formulación de un marco formal y cuantitativo que capte la vaguedad del conocimiento humano tal como se expresa a través de los lenguajes naturales. Cualquier progreso significativo a lo largo de esa línea requiere un estudio formal de modelos de razonamiento humano, en concreto los aspectos del razonamiento humano que vienen representados por categorías inciertas o vagas.

Hay conceptos inherentes a sistemas reales, los cuales su explicación y definición mediante palabras resulta más sencilla. La lógica difusa posibilita trasladar enunciados del lenguaje natural a un formalismo matemático. Es por ello que la lógica difusa ayuda a construir mejores modelos de la realidad y es uno de los componentes principales del 'soft computing'.

Otro de los conceptos a tener en cuenta es el concepto de variable lingüística, que fue introducido por Zadeh (1975b) para proporcionar una base para el razonamiento aproximado. Se entiende por variable lingüística aquella variable que toma como valores palabras u oraciones en un lenguaje natural o artificial (Rutkowska, 2002). El concepto de variable lingüística proporciona un medio de caracterización para aproximar fenómenos que son demasiado complejos o están mal definidos como para ser susceptibles de descripción en términos cuantitativos convencionales. El uso de palabras u oraciones en lugar de números se fundamenta en que las caracterizaciones lingüísticas son generalmente menos precisas que los valores numéricos, dando mayor versatilidad para adaptarse a la realidad que envuelve.

Series temporales

Se entiende por serie temporal a la colección de observaciones de una variable aleatoria recogidas secuencialmente en el tiempo. Normalmente dichas observaciones son equiespaciadas aunque también es posible una observación de forma continua, acumulándose los valores en intervalos de tiempo. Por lo tanto una serie temporal puede ser: discreta o continua dependiendo de cómo sean las observaciones. Si se pueden predecir exactamente los valores, se dice que las series son deterministas, si a partir de las observaciones pasadas el futuro sólo se puede determinar de forma parcial entonces las series son estocásticas.

En esta memoria, los métodos de predicción que se proponen se comparan con otros métodos de predicción difusa y con algunos de los modelos estadísticos de predicción clásicos de referencia, como son: *ARIMA*, suavizado exponencial (SE) y medias móviles (MM). Los modelos *ARIMA* son una clase de modelos estadísticos para hacer análisis y predicciones de series temporales (Box y Jenkins, 1976). *ARIMA* es el acrónimo de "*Autoregressive Integrated Moving Average*" (modelo autorregresivo integrado de media móvil).

Los métodos de suavizado exponencial (SE) se utilizan en el análisis univariante, debido a su simplicidad y robustez (en sentido estadístico) en los procedimientos automáticos de predicción (Hyndman et al., 2002), y que tienen en cuenta la tendencia (esto es si las observaciones presentan una forma creciente o decreciente) y los efectos estacionales (influencia de la serie en ciertos períodos de tiempo) de los datos (Gardner, 1985). Tienen sus orígenes en los trabajos de Brown (1959) y Holt (1957), aunque fueron más conocidos a partir de Winters (1960). La forma general de la función de predicción del suavizado exponencial, que involucra a un conjunto de coeficientes determinados, fue dada por primera vez en Box y Jenkins (1976).

Otro método para la predicción que se utiliza en esta tesis es el de medias móviles (MM), donde la predicción viene determinada por la media aritmética de un subconjunto de datos precedentes.

Selección de carteras

Una cartera es una colección de inversiones pertenecientes a una institución o a un inversor privado. El problema de selección de carteras, trata la elección de qué activos y en qué proporción intervienen en una cartera, según las preferencias del inversor, para lograr un alto rendimiento esperado asumiendo un riesgo bajo.

El origen de la teoría moderna de cartera de valores, es el modelo probabilístico de media-varianza (MV), donde la relación entre el rendimiento y riesgo de la inversión es de relevancia para la toma de decisiones, utilizando la teoría de la probabilidad y la optimización matemática como las principales herramientas para lograr una cartera más eficiente. El trabajo de Markowitz (1952), estableció la relación entre el rendimiento esperado y la varianza del rendimiento de la inversión en un esquema de intercambio riesgo-rendimiento, asumiendo una distribución normal multivariante de los rendimientos de los activos.

En la formulación estándar de dicho problema el inversor elige la proporción de la inversión total, la cual va a ser invertida en cada activo, y también decide acerca de su objetivo para el rendimiento esperado, mientras se trata de minimizar el riesgo bajo ciertas restricciones para lograr su objetivo. De hecho, se asume que el inversor quiere una cartera de mínimo riesgo compatible con restricciones de rentabilidad.

Markowitz (1952) propone que el inversor debe abordar la cartera como un todo, estudiando las características de riesgo y rendimiento global, en lugar de utilizar el rendimiento esperado de cada valor en particular. Para Markowitz, el proceso de selección de una cartera óptima puede dividirse en dos etapas: la primera etapa comienza con la observación del comportamiento del rendimiento de los valores financieros, incorporando la experiencia y conocimiento de los expertos, y termina con estimaciones del rendimiento futuro de las carteras (a través del rendimiento individual de los valores y del rendimiento de la cartera en su conjunto). La segunda etapa se refiere a los procedimientos diseñados para encontrar y seleccionar la cartera que sea óptima con respecto a un único objetivo u objetivos múltiples, supeditado a los requisitos impuestos por el inversor. En dicho artículo se ocupa de estudiar la segunda parte del proceso de selección de una cartera, asumiendo normalidad.

Otras medidas del riesgo de la inversión como la semi-varianza, la des-

viación media absoluta o el valor-en-riesgo (Markowitz, 1959; Konno y Yamazaki, 1991; Speranza, 1993; Jorion, 1997, respectivamente), entre otras, se utilizan como medidas alternativas en el modelo MV. También se ha tratado de modelizar las demandas tanto del inversor como de los mercados financieros a través de los costes de transacción, el horizonte multi-período, la cardinalidad restringida, etc. (Mansini y Speranza, 1999; Zhang y Li, 2012; Georgieva y Popchev, 2013; Chang et al., 2000, respectivamente).

A partir del conocimiento imperfecto de los rendimientos de los activos y la incertidumbre subyacente en el comportamiento de los mercados financieros es posible utilizar medidas y/o restricciones difusas en el problema de selección de carteras (Dubois y Prade, 1980).

En el contexto de problemas de optimización bajo incertidumbre, la programación matemática difusa incorpora la incertidumbre y la ambigüedad mediante conceptos difusos (Bellman y Zadeh, 1970), otros trabajos donde también se usa una formulación difusa para problema de selección de carteras son, entre otros, Watada (1997); Tanaka y Guo (1999); Inuiguchi y Ramík (2000); León et al. (2000); Carlsson et al. (2002); Vercher et al. (2007). También se proponen algunas aproximaciones que tratan modelos estocásticos y difusos simultáneamente basados en la teoría de la posibilidad (Li et al., 2010).

Habiendo realizado una introducción a los principales conceptos a tratar en esta tesis (lógica difusa, series temporales y selección de carteras), a continuación véanse las principales definiciones y propiedades de la teoría de los conjuntos difusos.

1.1. Teoría de Conjuntos difusos

Las definiciones básicas de conjuntos difusos y números difusos se extraen de los trabajos de Zadeh (1965, 1975a) y Goguen (1967).

Definición 1.1.

Dado un conjunto de objetos X . Se llama *conjunto difuso* A a la colección de objetos pertenecientes a X , conectados por una relación de incertidumbre a través de una expresión que caracteriza a los objetos por grados de pertenencia al conjunto A en el intervalo $[0, 1]$. El grado de pertenencia viene dado por la función denominada *función de pertenencia* $\mu_A : X \mapsto [0, 1]$.

Por lo tanto el conjunto difuso se define como:

$$A = \{(x, \mu_A(x)) / x \in X\}$$

□

En otras palabras, un conjunto difuso es un conjunto que puede contener elementos de forma parcial. Es decir que la propiedad $x \in A$ puede ser cierta con un grado de verdad. Se mide esta posibilidad de pertenecer (o pertenencia) con un número $\mu_A(x)$, entre 0 y 1, llamado función o grado de pertenencia de x en A . Si es 0, x no pertenece a A , si es 1, entonces $x \in A$, totalmente, y si $0 < \mu_A(x) < 1$, x pertenece a A de una manera parcial.

Véase ahora algunos de los conceptos básicos introducidos por (Zadeh, 1965) y desarrollados en Zadeh (1975a,b,c) y Zimmermann (2001). Sea A un conjunto difuso:

Definición 1.2.

Soporte de un conjunto difuso A : $supp(A) = \{x \in X / \mu_A(x) > 0\}$

Definición 1.3.

Altura de un conjunto difuso A : $hgt(A) = sup\{\mu_A(x) / x \in X\}$

Definición 1.4.

Un conjunto difuso A es *normal* si $\exists x_0 \in X / \mu_A(x_0) = 1$

Definición 1.5.

Puntos de cruce de un conjunto difuso A : $\{x \in X / \mu_A(x) = 1/2\}$

Definición 1.6.

Al conjunto compuesto por todos los conjuntos difusos de X se le llama conjunto potencia y se denota $\mathcal{P}(X)$.

Definición 1.7.

Dado $\alpha \in [0, 1]$ y un conjunto difuso $A \in \mathcal{P}(X)$ se le llama α -*corte* o *nivel* α de un conjunto difuso A al conjunto de nivel superior:

$$A_\alpha = \{x \in X / \mu_A(x) \geq \alpha\}$$

Definición 1.8. (Zimmermann, 2001)

Sean $A, B \in \mathcal{P}(X)$, se definen las funciones de pertenencia de:

- $A \cup B$ como $\mu_{A \cup B}(x) = max\{\mu_A(x), \mu_B(x)\}$

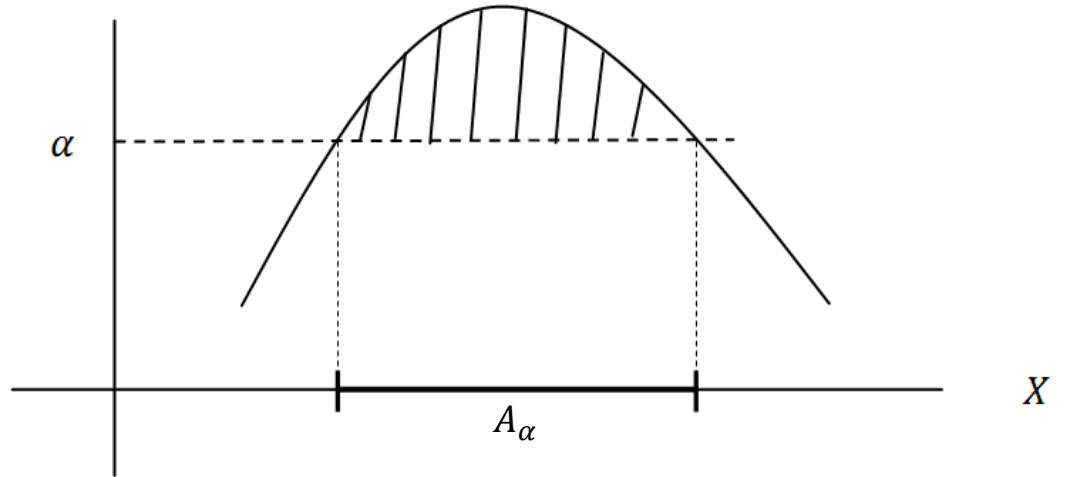


Figura 1.1: Representación gráfica de un conjunto de nivel superior (α -corte)

- $A \cap B$ como $\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}$
- A^c como $\mu_{A^c}(x) = 1 - \mu_A(x)$ (complementario de A)

Definición 1.9.

Sean $A, B \in \mathcal{P}(X)$

- Identidad: $A = B \iff \mu_A(x) = \mu_B(x) \quad \forall x \in X$
- Inclusión: $A \subseteq B \iff \mu_A(x) \leq \mu_B(x) \quad \forall x \in X$

A partir de estas definiciones es fácil probar las siguientes propiedades:

Proposición 1.10. (Bellman y Giertz, 1973)

Sean $A, B \in \mathcal{P}(X)$

- I) $A \subseteq B \wedge B \subseteq A \iff A = B$
- II) $B \subseteq A \iff B_\alpha \subseteq A_\alpha \quad \forall \alpha \in [0, 1]$
- III) $A \subseteq B \rightarrow \text{supp}(A) \subseteq \text{supp}(B)$
- IV) $A \subseteq B \wedge B \subseteq C \rightarrow A \subseteq C$ (transitividad)

Definición 1.11. (Zimmermann, 2001)

Sea A un conjunto difuso, entonces A es convexo si

$$\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \min\{\mu_A(x_1), \mu_A(x_2)\} \quad \forall x_1, x_2 \in X, \text{ y } \forall \lambda \in [0, 1]$$

□

En el caso que A sea convexo, se puede escribir el α -corte como:

$$A_\alpha = [\underline{a}(\alpha), \bar{a}(\alpha)]$$

Definición 1.12. (Dijkman et al., 1983)

Sea $X = \mathbb{R}$ y A un conjunto difuso. Se le llama A número difuso si A es convexo y normal.

Definición 1.13. (Zadeh, 1975b)

Sean $A \in \mathcal{P}(X)$ y $B \in \mathcal{P}(Y)$ dos conjuntos difusos, entonces se define el conjunto difuso *producto cartesiano* $A \times B$ con función de pertenencia $\mu_{A \times B}(x, y) = \min\{\mu_A(x), \mu_B(y)\}$

□

En el campo de los conjuntos difusos otra forma más sencilla de denotar un conjunto difuso cuando el soporte es un conjunto con cardinal finito es como sigue:

Sea U el conjunto de datos y $T \in \mathcal{P}(U)$ un conjunto difuso entonces

$$T = \{(u, \mu_T(u)) | u \in U\} = \mu_T(u_1)/u_1 + \mu_T(u_2)/u_2 + \mu_T(u_3)/u_3 + \dots$$

Se propone un ejemplo para ver cómo se construye el producto cartesiano:

$$\text{Sea } X = \{x_1, x_2, x_3\} \text{ y } A \in \mathcal{P}(X) \quad \longrightarrow \quad A = 0.2/x_1 + 0.5/x_2 + 1/x_3$$

$$\text{Sea } Y = \{y_1, y_2\} \text{ y } B \in \mathcal{P}(Y) \quad \longrightarrow \quad B = 0.3/y_1 + 0.9/y_2$$

El producto cartesiano $A \times B$ tendría los siguientes puntos:

$$\{(x_1, y_1), (x_2, y_1), (x_3, y_1), (x_1, y_2), (x_2, y_2), (x_3, y_2)\}$$

y de la definición de función de pertenencia del producto cartesiano, se obtiene:

$$\mu_{A \times B}(x_1, y_1) = \min(\mu_A(x_1), \mu_B(y_1)) = \min(0.2, 0.3) = 0.2$$

$$\mu_{A \times B}(x_2, y_1) = \min(\mu_A(x_2), \mu_B(y_1)) = \min(0.5, 0.3) = 0.3$$

$$\mu_{A \times B}(x_3, y_1) = \min(\mu_A(x_3), \mu_B(y_1)) = \min(1, 0.3) = 0.3$$

...

Definición 1.14. (Dubois y Prade, 1980)

Sean los conjuntos $X, Y, Z \subseteq \mathbb{R}$ y sea $f : X \times Y \rightarrow Z$ una función lineal tal que $z = f(x, y)$ con $x \in X, y \in Y$ y $z \in Z$. Por el principio de extensión de Zadeh (1975a), los conjuntos difusos $A \in \mathcal{P}(X), B \in \mathcal{P}(Y)$ aplicados bajo la función lineal f , da como resultado otro conjunto difuso $C = f(A, B)$ que está en $\mathcal{P}(Z)$, tal que:

$$\mu_C(z) = \sup_{(x,y) \in X \times Y | z=f(x,y)} \text{mín}[\mu_A(x), \mu_B(y)]$$

Véase a continuación la composición suma con el operador sup-mín, sean $Y = \{1, 2\}, X = \{1, 2, 3\}$ tal que $X, Y \in \mathbb{R}$ y sean los conjuntos difusos $A = 0.2/1 + 0.5/2 + 1/3$ y $B = 0.3/1 + 0.9/2$, sea $f(x, y) = x + y$ entonces el operador suma (+) de conjuntos difusos se define como:

$$\mu_{A+B}(z) = \sup_{(x,y) \in X \times Y | x+y=z} [\text{mín}[\mu_A(x), \mu_B(y)]]$$

En este caso se tiene $\{2, 3, 4, 5\}$ como todas las posibles combinaciones de la suma aritmética $x + y$ con $x \in X$ e $y \in Y$, y por lo tanto el operador suma resultaría:

$$\mu_{A+B}(2) = \sup[\text{mín}[\mu_A(1), \mu_B(1)]] = \sup[\text{mín}[0.2, 0.3]] = 0.2$$

de forma análoga se calculan los demás puntos

$$\mu_{A+B}(3) = \sup[\text{mín}[\mu_A(1), \mu_B(2)], \text{mín}[\mu_A(2), \mu_B(1)]] = \sup[0.2, 0.5] = 0.5$$

$$\mu_{A+B}(4) = \sup[\text{mín}[\mu_A(2), \mu_B(2)], \text{mín}[\mu_A(3), \mu_B(1)]] = \sup[0.5, 0.3] = 0.5$$

$$\mu_{A+B}(5) = \sup[\text{mín}[\mu_A(3), \mu_B(2)]] = \sup[0.9] = 0.9$$

□

Definición 1.15.

Un *número difuso trapezoidal* se denota como $D = (a, b, \gamma, \beta)$ donde el núcleo (o corazón) del número difuso trapezoidal D está en el intervalo $[a, b]$ y las cantidades γ y β son las amplitudes derecha e izquierda (longitud que hay entre el extremo inferior o superior del núcleo y el límite inferior o superior del soporte), respectivamente. Y donde la función de pertenencia es:

$$\mu_D(x) = \begin{cases} \frac{x-(a-\gamma)}{\gamma} & \text{si } x \in [a - \gamma, a] \\ 1 & \text{si } x \in [a, b] \\ \frac{(b+\beta)-x}{\beta} & \text{si } x \in [b, b + \beta] \\ 0 & \text{en otro caso} \end{cases}$$

Como caso particular, si $a = b$ entonces μ_T es la función de pertenencia de un número difuso triangular, denotado como $D = (a, a, \gamma, \beta)$.

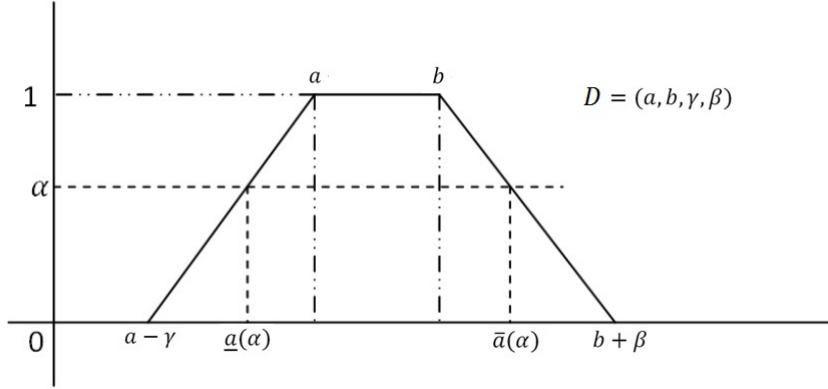


Figura 1.2: Gráfica de un número difuso trapezoidal

Proposición 1.16. (Dubois y Prade, 1980)

Aritmética difusa: sin pérdida de generalidad, sean $A = (a_A, b_A, \gamma_A, \beta_A)$ y $B = (a_B, b_B, \gamma_B, \beta_B)$ dos números difusos trapezoidales y sea $\lambda \in \mathbb{R}$. Entonces:

- I) *La suma de números difusos trapezoidales también es un número difuso trapezoidal. Y se calcula como sigue:*

$$A + B = (a_A + a_B, b_A + b_B, \gamma_A + \gamma_B, \beta_A + \beta_B)$$

- II) *La multiplicación de un número difuso trapezoidal por un escalar también es un número difuso trapezoidal. Y se calcula como sigue:*

$$\lambda A = \begin{cases} (\lambda a_A, \lambda b_A, \lambda \gamma_A, \lambda \beta_A) & \text{si } \lambda \geq 0 \\ (\lambda b_A, \lambda a_A, |\lambda| \beta_A, |\lambda| \gamma_A) & \text{si } \lambda < 0 \end{cases}$$

Nótese que la suma y multiplicación por un escalar se deduce por el principio de extensión (Zadeh, 1975a) y el operador sup-mín de la definición 1.14.

Definición 1.17. (Dubois y Prade, 1980)

Sean $L, R : [0, +\infty] \rightarrow [0, 1]$ donde L es una función real semi-continua decreciente, y R es una función real semi-continua decreciente, y $L(0) = R(0) = 1$

Sean $a, b, \gamma, \beta \in \mathbb{R}^+$ con $a \leq b$ entonces

A es un número difuso de tipo LR si y solo si

$$\mu_A(x) = \begin{cases} L\left(\frac{a-x}{\gamma}\right) & \text{si } x \leq a \\ 1 & \text{si } a \leq x \leq b \\ R\left(\frac{x-b}{\beta}\right) & \text{si } x \geq b \end{cases}$$

L y R son las funciones de referencia (izquierda L del inglés 'left' y derecha R del inglés 'right') del número difuso A . Así m y n son los extremos del núcleo de A ($[a, b] = \{x | \mu_A(x) = 1\}$), y γ, β son las amplitudes de las colas izquierda y derecha, respectivamente.

□

Como caso particular de número de tipo LR, se tiene los números LR potencia, donde las funciones L y R están definidas como funciones potencia de forma genérica:

$$L(x) = 1 - \left(\frac{a-x}{\gamma}\right)^\alpha \qquad R(x) = 1 - \left(\frac{x-b}{\beta}\right)^\delta$$

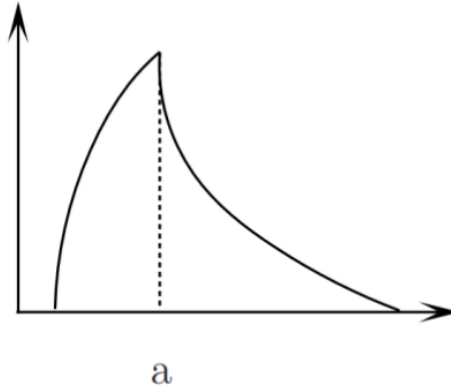


Figura 1.3: Gráfica número difuso de tipo LR potencia donde $b=a$

Definición 1.18. (Delgado et al., 1998)

Sean A un conjunto difuso, $A_\alpha = [\underline{a}(\alpha), \bar{a}(\alpha)]$ un α -corte de A y f una función de reducción entonces el *valor* de A (con respecto a f) se define como:

$$Val_f(A) = \int_0^1 f(\alpha) [\bar{a}(\alpha) + \underline{a}(\alpha)] d\alpha \qquad (1.1)$$

Delgado et al. entienden como función reductora a: $f : [0, 1] \rightarrow [0, 1]$ tal que $\int_0^1 f(x)dx = 0.5$. Por otro lado $Val_f(A)$ puede verse como un valor central que representa desde un punto de vista global el valor de la magnitud que representa el número difuso.

Definición 1.19. (Delgado et al., 1998)

Sean A un conjunto difuso, $A_\alpha = [\underline{a}(\alpha), \bar{a}(\alpha)]$ un α -corte de A y f una función de reducción entonces la *ambigüedad* de A (con respecto a f) se define como:

$$Amb_f(A) = \int_0^1 f(\alpha)[\bar{a}(\alpha) - \underline{a}(\alpha)]d\alpha \quad (1.2)$$

□

$Amb_f(A)$ puede verse como la "amplitud global" de la función de pertenencia de A (ponderada a través de f); esto es obviamente una medida de la vaguedad de A .

Como caso particular se tiene que el valor de A coincide con el valor medio possibilístico al tomar como función reductora constante 0.5 (Bermúdez et al., 2007b).

En esta sección se ha introducido la teoría de conjuntos difusos, viendo las principales definiciones y los conceptos más importantes que serán utilizados a lo largo de esta memoria. A continuación véase una introducción a las series temporales difusas y el principal esquema de predicción en este tipo de series.

1.2. Series Temporales difusas (STD)

Basado en los trabajos de Zadeh (1965, 1975b), se define y estudia el proceso dinámico que toma valores lingüísticos como observaciones (Song y Chissom, 1993a). Este proceso dinámico se denomina serie temporal difusa. Para describir series temporales difusas, se emplean modelos basados en ecuaciones relacionales difusas.

Véase brevemente algunos conceptos fundamentales de series temporales difusas:

Definición 1.20. (Song y Chissom, 1993a)

Sea Y un universo de objetos y $U(t) \subseteq Y$ con $t = 1, 2, \dots$ siendo t el instante en el que se toma la observación. $U(t)$ es un subconjunto denominado universo de discurso donde se definen los conjuntos difusos $A_i(t)$ con $i = 1, 2, \dots$ siendo

i la enumeración de los conjuntos difusos.

A la colección de $A_i(t)$ para $t = 1, 2, \dots$, denotado como $F(t)$, se le llama *serie temporal difusa* en $U(t)$.

□

Donde se entiende que $F(t)$ es una variable difusa (según definición de Zadeh (1975c)) pudiendo ser lingüística o no y que toma por valores los conjuntos difusos $A_i(t)$.

Por lo tanto $A_i(t)$ es la realización de la variable difusa $F(t)$ en el instante t . En todos los casos considerados en esta memoria, $\forall t = 1, 2, \dots, N$ se tiene que $U(t) = U(t+1)$ y $A_i(t) = A_i(t+1)$ son invariantes temporales, y por lo tanto se denota el universo de los datos como U y los conjuntos difusos como A_i . Nótese que una observación concreta de la serie temporal difusa en el instante t_0 se denota como $F(t_0) = A_i$.

Para ejemplificar la noción de serie temporal difusa, véase a continuación un ejemplo:

Sea Y el conjunto de todos los posibles enunciados del lenguaje natural que pudiera realizar un individuo acerca del parámetro meteorológico de temperatura $Y = \{\text{frío; muy frío; cálido; muy cálido; congelado; } \dots\}$ donde cada uno de los enunciados puede ser representado mediante un conjunto difuso. Ahora sea $U = \{\text{cálido; muy cálido}\} \subseteq Y$ el conjunto de todas las observaciones para $t = 1, 2, 3, 4$. Considérese, por ejemplo, que $\{F(t), t = 1, \dots, 4\} = \{\text{cálido; muy cálido; muy cálido; cálido}\}$ es una serie temporal difusa. En este ejemplo $F(t)$ funciona como una variable lingüística, siendo los términos lingüísticos "cálido" y "muy cálido" realizaciones de dicha variable lingüística.

Sean μ_c y μ_{mc} funciones de pertenencia de los términos lingüísticos "cálido" y "muy cálido" respectivamente:

$$\mu_c(x) = \begin{cases} 1 & \text{si } 25 \leq x \leq 30 \\ 0.5 & \text{si } 30 < x \leq 35 \\ 0 & \text{en otro caso} \end{cases} \quad \mu_{mc}(x) = \begin{cases} 0.5 & \text{si } 25 \leq x \leq 30 \\ 1 & \text{si } 30 < x \leq 35 \\ 0 & \text{en otro caso} \end{cases}$$

siendo x la variable temperatura en grados centígrados. Así la serie temporal difusa $F(t)$ podría venir inducida por una serie temporal de números que provienen de temperaturas observadas.

Definición 1.21. (Song y Chissom, 1994)

Sea $\{F(t), t = 0, 1, 2, \dots\}$ una serie temporal difusa, sin pérdida de generalidad, sean $F(t_0) = A_i$ y $F(t_0 - 1) = A_j$ con $t_0 \in \mathbb{N}^+$ y con A_i, A_j conjuntos difusos definidos sobre el universo U . Entonces A_i, A_j se dicen que están relacionados si:

$$A_j = A_i \circ R_{ij}(t_0, t_0 - 1) \quad (1.3)$$

donde \circ es el operador composición max-min y $R_{ij} = A_i \times A_j$ (producto cartesiano) es la *relación difusa* entre ellos, entonces se dice que A_j es consecuencia de A_i , llamándose A_i antecedente y A_j consecuente. La relación difusa (según la definición de Song y Chissom (1991)) se denota como:

$$A_i \longrightarrow A_j \quad (1.4)$$

El modelo de Song y Chissom se basa en las relaciones difusas de primer orden existentes entre diferentes observaciones en diferentes tiempos.

Revisando la literatura se pueden ver algunos ejemplos de composición de relaciones difusas mediante operadores como la composición max-min (Song y Chissom, 1993a), min-max (Song y Chissom, 1994) o el operador aritmético (Chen, 1996).

También el concepto de n-orden fue introducido por Song y Chissom (1993a) y mejora su definición en Song y Chissom (1993b). Los modelos n-orden de series temporales difusas están referenciados como modelos de orden superior.

Definición 1.22. (Song y Chissom, 1993b)

Sea $\{F(t), t = 0, 1, 2, \dots\}$ una serie temporal difusa, si $F(t)$ está causada por $F(t-1), F(t-2), F(t-3) \dots F(t-n)$, entonces la relación difusa se representa por $F(t-n) \dots F(t-2), F(t-1) \longrightarrow F(t)$, y $F(t)$ se denomina como *serie temporal de orden n*.

□

Esquema clásico de series temporales difusas (STD)

Debido a la gran cantidad de cálculos y a las complejas operaciones requeridas por los modelos de Song y Chissom, Chen (1996) usa una serie de operaciones aritméticas que simplifican el modelo y consigue que con menos cálculos sea un modelo más eficiente, obteniendo las predicciones más robustas (en sentido estadístico) frente al ruido del histórico de datos. El esquema propuesto por Chen establece un punto de partida para el desarrollo de toda la metodología de las STD.

A continuación se presenta el esquema clásico descrito por Chen (1996) para las STD, que es el esquema base que se sigue en esta tesis con algunas modificaciones, que serán introducidas en su momento:

Sea $\{y_t, t = 1, \dots, N\}$ un conjunto de observaciones de una serie temporal.

Etapa 1. Definir el universo.

Sean $y_{min} = \min_{1 \leq t \leq N} (y_t)$ y $y_{max} = \max_{1 \leq t \leq N} (y_t)$ dados $d_1, d_2 \in \mathbb{R}^+$ se define el universo como el intervalo

$$U = [\underline{u}, \bar{u}] = [y_{min} - d_1, y_{max} + d_2]$$

Etapa 2. Partir el universo de los datos.

Se subdivide el Universo U en 7 intervalos de longitudes iguales,

$$\{u_1, \dots, u_7\} \subset U \quad \text{y} \quad \bigcup_{1 \leq i \leq 7} u_i = U$$

La longitud del intervalo sería $l = \frac{\bar{u} - \underline{u}}{7}$, $u_i = [\underline{u}_i, \bar{u}_i] = [\underline{u} + (i - 1) * l, \underline{u} + i * l]$ para $i = 1, \dots, 7$. Y sea $I = \{1, \dots, 7\}$ el conjunto de subíndices.

Etapa 3. Determinar los valores lingüísticos.

Se determinan los conjuntos difusos definidos sobre cada una de las particiones, como representación de los valores lingüísticos.

Los conjuntos difusos normalmente se definen de manera sencilla, se tienen 7 conjuntos difusos:

$$A_i = \mu_{A_i}(u)/u_1 + \dots + \mu_{A_i}(u)/u_7$$

Chen propone las siguientes funciones de pertenencia

$$\mu_{A_1}(x) = \begin{cases} 1 & \text{si } x \in u_1 \\ 0.5 & \text{si } x \in u_2 \\ 0 & \text{en otro caso} \end{cases}$$

Para $i = 2, \dots, 6$

$$\mu_{A_i}(x) = \begin{cases} 0.5 & \text{si } x \in u_{i-1} \\ 1 & \text{si } x \in u_i \\ 0.5 & \text{si } x \in u_{i+1} \\ 0 & \text{en otro caso} \end{cases}$$

$$\mu_{A_7}(x) = \begin{cases} 0.5 & \text{si } x \in u_6 \\ 1 & \text{si } x \in u_7 \\ 0 & \text{en otro caso} \end{cases}$$

Etapa 4. Construir la STD.

Se establece la regla que determina cómo se observa un conjunto difuso en un instante t_0 .

Se determina la relación entre el histórico de datos y los conjuntos difusos, dados $1 \leq i \leq 7$ e y_{t_0} el dato observado en el instante t_0 con $1 \leq t_0 \leq N$:

$$F(t_0) = A_i \iff \mu_{A_i}(y_{t_0}) = \max_k(\mu_{A_k}(y_{t_0})) \quad \forall k \in I$$

Con las funciones de pertenencia definidas en la etapa anterior, por construcción se tiene que:

$$\text{si } y_{t_0} \in u_i \mapsto F(t_0) = A_i$$

Continuando con el ejemplo anterior (página 15) donde se tenían los conjuntos difusos (en dicho caso valores lingüísticos) definidos por las funciones de pertenencia μ_c y μ_{mc} . Dadas cuatro observaciones cronológicamente consecutivas de la temperatura ambiente

tomadas en grados centígrados $\{28, 31, 33, 29\}$, y siguiendo el procedimiento descrito en esta etapa 4 se construiría la serie temporal difusa $\{F(t), t = 1, \dots, 4\} = \{\text{cálido; muy cálido; muy cálido; cálido}\}$.

Etapa 5. Establecer las relaciones difusas.

Dichas relaciones están inducidas por las relaciones existentes de la serie temporal inicial. Las relaciones entre los datos de la serie temporal inducen a relaciones lógicas difusas entre los conjuntos difusos, esto es:

Sean $1 \leq t_0 \leq N$ y $1 \leq i, j \leq 7$ con $y_{t_0-1} \in u_i$ y $y_{t_0} \in u_j$

Por lo tanto se tiene que $F(t_0 - 1) = A_i$ y $F(t_0) = A_j$, entonces se establece una relación lógica difusa (RLD):

$$A_i \rightarrow A_j$$

donde A_i es el antecedente y A_j el consecuente. Nótese que la relación lógica difusa sólo se establece no se calcula, Song y Chissom (1994) sí que calculan relaciones difusas.

Etapa 6. Construir las *relaciones lógicas difusas de grupos* (RLDG).

Para construir las RLDG se agrupan todas aquellas relaciones que tienen el mismo conjunto difuso antecedente.

Esto es, si $A_i \rightarrow A_j$ y $A_i \rightarrow A_k$ entonces se pueden agrupar en una sola expresión: $A_i \rightarrow A_j, A_k$

Los subíndices de la RLDG asociado al conjunto difuso A_i conforman el conjunto $J_i = \{j, k\}$

Etapa 7. Predicción puntual.

Se calcula el valor puntual en el instante $N + 1$, a partir de unas reglas de predicción previamente establecidas.

Chen (1996) toma como representante puntual de los conjuntos difusos, los centros de los intervalos que conforman la partición inicial, $m_i = \frac{\bar{u}_i + u_i}{2}$ para $i = 1, \dots, 7$.

Finalmente quedando la predicción puntual como:

Sea $F(N) = A_i$ y J_i el conjunto de índices de las RLDG donde el antecedente es A_i , entonces la predicción puntual es:

$$\hat{y}_{N+1} = \frac{\sum_{j \in J_i} m_j}{|J_i|} \quad (1.5)$$

donde $|J_i|$ es el cardinal del conjunto J_i .

□

El esquema propuesto por Song y Chissom es similar al de Chen, diferenciándose en las etapas 5, 6 y 7.

Se ha visto una introducción teórica a las series temporales difusas (STD) y un esquema clásico de predicción para ese tipo de series. A continuación, haciendo una revisión bibliográfica, se verá el estado de la cuestión analizando las principales modificaciones sobre el esquema y resaltando las aportaciones de cada uno de los artículos y/o libros tomados como referencias bibliográficas en esta memoria.

1.3. Modificación del esquema clásico de STD

Con objeto de ampliar el campo de estudio de la teoría de conjuntos difusos a nuevas disciplinas, Song y Chissom (1993a) introducen el concepto difuso en el campo de las series temporales, basándose en las ecuaciones relacionales difusas y el razonamiento aproximado (Wu, 1986; Zadeh, 1975a,b,c). Song y Chissom (1993a) plantean por primera vez, lo que acuñan con el término de series temporales difusas, un nuevo enfoque que trata problemas enmarcados en contextos difusos, donde el histórico de datos se expresa en valores lingüísticos o difusos y que permiten dar un nuevo tratamiento para series temporales.

Song y Chissom (1993a,b, 1994) presentan las definiciones de series temporales difusas para modelos invariantes y variantes temporales, donde aplican distintos operadores max-min (Song y Chissom, 1993a) y el min-max (Song y Chissom, 1994) para aproximar las relaciones difusas entre los conjuntos, ambos operadores requieren de muchos cálculos.

De forma paralela a Song et al. e independientemente, Wang y Mendel (1992) formulan algunos de los procesos desarrollados en la algoritmia de las STD, como la definición y posterior partición del universo de los datos, el uso de números difusos triangulares para definir los conjuntos difusos, y una serie de reglas heurísticas para la predicción puntual.

Sullivan y Woodall (1994) haciendo un repaso crítico al trabajo presentado por Song et al. en años anteriores sobre las STD para los modelos invariantes y variantes temporales, y haciendo comparaciones con modelos de probabilidad, concretamente con el modelo de Markov de estados discretos

(Kemeny y Snell, 1960; Sullivan y Woodall, 1994) proponen tratar los valores lingüísticos como datos categóricos o estados discretos y aplican el modelo de Markov, obteniendo buenos resultados que mejoran a los obtenidos por Song y Chissom (1993a).

En Song et al. (1995) y en Song y Leland (1996) presentan las series temporales difusas homogéneas y dan la idea de utilizar números difusos como observaciones así como una estimación de la predicción puntual, a la vez que matizan y explican la definición de STD dada en trabajos anteriores.

Las etapas descritas en la algoritmia de las STD, tanto de Song y Chissom como el esquema de Chen, abren diferentes líneas de investigación que han sido desarrolladas durante los últimos años o décadas, con trabajos que se centran en las diferentes etapas del modelo básico. A continuación, en los siguientes apartados se comentan las principales aportaciones a las distintas etapas del esquema clásico de predicción de STD. Casi todos los artículos que se mencionan modifican varias de las etapas del esquema clásico, pero se clasifican por etapas según la relevancia o importancia que ha tenido cada modificación.

1.3.1. Aportaciones relevantes en la Etapa 2: Partir el universo de datos

Huarng (2001a) prueba que la partición inicial del universo afecta en gran manera a las predicciones. Para tratar la partición inicial propone dos algoritmos (*distribution-based length* y *average-based length*) que determinan la longitud de los intervalos que conforman dicha partición. El modelo propuesto por Huarng sigue el esquema básico de Chen y se verifica una disminución notable de los errores de ajuste. De esta manera Huarng prueba la existencia de una relación entre el tamaño de los intervalos que conforman la partición del universo y los resultados de la predicción.

Huarng (2001b) también aplica a las STD unas reglas heurísticas e introduce las relaciones lógicas difusas heurísticas para grupos. Unos años más tarde, Huarng y Yu (2006b) introducen una nueva forma de calcular la partición inicial, basándose en las tasas de cambios de observaciones consecutivas y en los percentiles, dando como consecuencia una partición en intervalos de longitud variable, y razona como ésta partición refleja mejor las variaciones de los datos en diferentes contextos.

Cheng et al. (2006) proponen dos modelos que se diferencian en la par-

tición inicial del universo y en los conjuntos difusos empleados (son de tipo LR), y que utilizan la definición de relación difusa descrita por Ross (1997). Introducen por primera vez el concepto de entropía (Shannon, 1948) en el campo de las STD. Los principios de maximización de la entropía de Jaynes (1957) y el principio de minimización de la divergencia de Kullback y Leibler (1951) se aplican en multitud de campos (Kapur, 1989), ahora también en las STD. En el primer modelo denominado como MEPA (*Minimize Entropy Principle Approach*), toma la entropía de una distribución de probabilidad como medida de la incertidumbre de la distribución (Christensen, 1985) y analiza la minimización de la entropía (ganancia de información) para determinar la cantidad de información contenida en cada conjunto. Así el principio minimax de la entropía formulado por Christensen (1985) es usado para crear un proceso de segmentación del universo, que repetido 7 veces da como resultado una partición del universo en 7 intervalos de longitud variable, en este modelo utiliza números triangulares para definir los conjuntos difusos. El segundo modelo denominado como TFA (*Trapezoid Fuzzification Approach*) sigue el esquema básico descrito por (Chen, 1996), pero utilizando números trapezoidales difusos.

Chen y Chung (2006b) usan por primera vez algoritmos genéticos (Holland, 1975; Goldberg, 1989; Banzhaf, 2009), buscando la longitud óptima del intervalo de la partición del universo que minimiza el error de ajuste.

Li y Cheng (2007) haciendo una recapitulación de los grandes problemas que tienen las STD, presentan un modelo que intenta controlar la incertidumbre de la predicción, mejorar la partición de los intervalos y mejorar la robustez de la predicción para diferentes longitudes de intervalos.

Un año más tarde, Cheng et al. (2008b) introducen una nueva herramienta para trabajar las STD, con la aplicación de técnicas de clustering. Concretamente la c-media difusa, técnica introducida por Bezdek (1981), con la que se pretende buscar una partición del histórico de datos por medio de la minimización de una función objetivo que tiene como parámetro principal las funciones de pertenencia de los conjuntos difusos y los centros de los clústers. Uno de los principales problemas que tienen es determinar el número de clústers adecuado, razonando a partir de los postulados de Miller (1956) y su número mágico, determinan que el número de clústers a introducir en el algoritmo es 7. En este trabajo, utilizan los centros de los clústers como los representantes de los distintos conjuntos difusos y que por lo tanto son usados posteriormente como los valores puntuales del proceso de predicción puntual.

En sintonía con Cheng et al. (2008a) y con Yu (2005b), se encuentra el trabajo desarrollado por Lee et al. (2009), que simplemente tomando como referencia los mencionados trabajos anteriores, introducen un pequeño ajuste en la predicción final.

Wang y Chen (2009) siguiendo el trabajo de Lee et al. (2006) con algunas modificaciones e introduciendo las técnicas de clustering estudiadas por Cheng et al. (2008b), presentan su modelo bi-factor de orden superior utilizando las citadas técnicas de clustering para la partición inicial del universo.

A partir de los primeros algoritmos heurísticos de optimización aplicados a STD (Chen y Chung, 2006a,b; Lee et al., 2006) otros caminos se exploran por Kuo et al. (2009) utilizando el algoritmo de optimización PSO del inglés 'Particle Swarm Optimization'(Kennedy y Eberhart, 1995, 2001), con el fin de obtener una partición del universo de los datos. El trabajo de Kuo et al. (2009) está centrado en las dos partes que consideran más influyentes en las STD, una es la partición inicial del universo y la otra parte son las reglas de predicción, proponiendo un nuevo esquema para tratar las reglas de predicción difusa.

Wang et al. (2013, 2014) partiendo del algoritmo de Chen (1996) y de las STD ponderadas de Lee et al. (2009) e incidiendo en el hecho de que cada conjunto difuso se tendría que adaptar más a los datos que representa, proponen una partición del universo en intervalos de longitud variable, usando técnicas de clustering como FCM del inglés 'Fuzzy C-Mean'(Bezdek, 1981) o GG (Gath y Geva, 1989); y usando también información granulada (Pedrycz y Vukovich, 2001).

Continuando con los algoritmos heurísticos de optimización, Enayatifar et al. (2013) aplican para la búsqueda del tamaño óptimo de los intervalos el algoritmo imperialista competitivo (Atashpaz-Gargari y Lucas, 2007), que combinado con un modelo de STD ponderada de orden superior intenta hacer predicciones a corto plazo.

Chen y Chen (2014) utilizan técnicas de discretización de la entropía, como en Chen et al. (2011), y también la transformada de Fourier para transformar los datos en señales de onda. El modelo propuesto está implementado usando métodos de *bootstrapping*, para los cuales usan períodos inferiores a un mes. La entropía se utiliza como parámetro de referencia para hacer la partición inicial, anteriormente de forma análoga ya había sido utilizada la entropía (Cheng et al., 2006). En cuanto a la predicción, similar a Cheng

et al. (2008a), tiene en cuenta los errores de predicción cometidos. Finalmente se reajusta la predicción puntual con el método de FFT (*Fast Fourier Transform*, Raynolds y Mullin (2005)).

Duru y Bulut (2014) buscan el número eficiente de clústers en la STD que mejoran las predicciones, para ello proponen la utilización de un nuevo método de clustering llamado HDP (*Histogram Damping Partition*) que define sub-clusters a partir de la desviación típica muestral, esta técnica esta basada en el trabajo descrito por Pathak (2011). El mismo año Bulut (2014) ve necesario la integración de la estacionalidad en los modelos de predicción difusa, proponiendo un modelo estacional volviendo a utilizar la técnica HDP. También mencionar otros trabajos interesantes posteriores, basados en técnicas de clustering y clustering difuso, como Askari y Montazerin (2015).

Lu et al. (2015) siguiendo el trabajo de Wang et al. (2014) y el campo de la información granulada, proponen un nuevo método para partir el universo basado en los intervalos de información granulada. Después de encontrar los puntos del universo que lo dividen en intervalos de longitud variable, se aplica o un algoritmo PSO (*Particle Swarm Optimization*) o bien un GA (algoritmo genético), para optimizar cierta función dependiente de las amplitudes de los intervalos.

En la misma línea de investigación que Chen y Hsiao (2006) se tiene a Chen y Chen (2015), que minimizan la entropía para definir los intervalos, y a la predicción final añaden un factor de corrección dependiente de errores anteriores de predicción, como en Cheng et al. (2008a).

En Ye et al. (2016) se plantea un modelo multi-orden que combina el uso de algoritmos genéticos y el análisis técnico. En cuanto al análisis técnico se pueden ver trabajos similares en Chu et al. (2009) y Sadaei y Lee (2014). Y en cuanto al algoritmo genético para determinar la partición inicial, en este sentido ya en Chen y Chung (2006b) se pudo ver el funcionamiento de este tipo de algoritmos.

El modelo de Sadaei et al. (2016) está basado en la tendencia de los datos y en la utilización de algoritmos evolutivos (ICA del inglés *Imperialist Competitive Algorithm*), como en Enayatifar et al. (2013)). Este interesante trabajo propone tratar la tendencia a partir de las primeras diferencias de los subíndices de las STD y también propone una nueva forma de predicción a partir de esas primeras diferencias, mediante una combinación lineal entre el valor observado actual y la predicción con las primeras diferencias. El al-

goritmo evolutivo lo aplica sobre dicha combinación lineal para mejorar la precisión de la predicción.

Singh y Dhiman (2017) presentan un modelo de programación lineal para determinar la longitud de los intervalos de la partición del universo de los datos. Basado en un modelo de orden superior, usa el grado de pertenencia de los puntos centrales de los intervalos para generar predicciones puntuales.

1.3.2. Aportaciones relevantes en la Etapa 3: Determinar los valores lingüísticos.

Lee et al. (2006); Cheng et al. (2006) y Liu (2007) fueron los primeros en proponer el uso de números trapezoidales difusos como conjuntos difusos definidos sobre cada una de las particiones. La interpretación del problema a tratar debe afectar a la forma en que vas a modelizar o representar la incertidumbre subyacente de los datos, por ello se debe seleccionar el conjunto difuso más adecuado para dicha incertidumbre, ya que este hecho puede alterar el funcionamiento y los resultados que proporciona el modelo de predicción de STD, y hasta ese momento no se había prestado demasiada atención a este hecho.

Chou (2017) utiliza las series temporales difusas basadas en el concepto de nivel de significación predictiva a largo plazo para poder evaluar la auto-correlación del histórico de datos. Utiliza números triangulares como representación de los conjuntos difusos, aproximándolos mediante un nuevo método de integración gradual.

1.3.3. Aportaciones relevantes en la Etapa 4: Construir la STD.

En Teoh et al. (2009) se utiliza la teoría de aproximación de conjuntos (Pawlak, 1982; Walczak y Massart, 1999). Esta teoría es un procedimiento lógico de razonamiento para analizar la información de un sistema, que en este caso se utiliza para el análisis de las relaciones lógicas difusas de orden superior, extrayendo patrones de fluctuaciones para generar reglas de predicción y pesos. Finalmente aplica un ajuste sobre la predicción, el modelo de esperanza adaptada de Kmenta (1986), ya aplicado en anterioridad en STD (Cheng et al., 2008a), con el objeto de mejorar la precisión de la predicción.

1.3.4. Aportaciones relevantes en la Etapa 5: Establecer las relaciones difusas.

Huarng y Yu (2006a) introducen por primera vez la inteligencia artificial (McCulloch y Pitts, 1943; Turing, 1950) en el campo de las STD, utilizan una red neuronal (McCarthy et al., 1955) para la predicción, y también proponen un modelo híbrido para el reconocimiento de patrones (Weiss y Kulikowski, 1991). En trabajos anteriores ya se usaron las redes neuronales para la predicción (Indro et al., 1999), aunque en este caso particularmente se usan las redes neuronales para establecer las relaciones lógicas difusas de las STD. El trabajo de Huarng y Yu (2006a) se basa sobre todo en Zhang (2004), tanto la adecuación de los datos como en la preparación de la red o el modelo de selección y evaluación.

Yu y Huarng (2008) siguiendo el trabajo realizado por Huarng y Yu (2006a) y basándose también en la utilización de redes neuronales en las STD (Smith y Gupta, 2002; Widrow et al., 1994); proponen el uso de redes neuronales en un modelo de STD de dos factores. Más tarde, Aladag et al. (2009) que al igual que Huarng y Yu (2006a) utilizan las redes neuronales para establecer las relaciones difusas y presentan otro modelo de predicción de orden superior.

1.3.5. Aportaciones relevantes en la Etapa 6: Construir las RLDG.

Chen (2002) desarrolla el concepto de series temporales difusas de orden superior, esto es, no sólo tiene en cuenta la última observación, también observaciones anteriores para la construcción de las RLDG y la obtención de predicciones.

Song (2003) utiliza funciones de auto-correlación con la intención de medir el grado de dependencia entre los conjuntos difusos. Como los conjuntos difusos se definen a partir de los intervalos, lo que finalmente va a medir es la dependencia entre los intervalos, y esta se define como una dependencia colectiva entre los diferentes puntos de cada intervalo. Dado que se tiene un número infinito de puntos en cada intervalo, Song hace una aproximación de la medida de dependencia al considerar sólo un número finito de puntos de cada intervalo. Como tanto el número como la selección de los puntos influirá en la medida de la dependencia, para simplificar el problema lo que hace es tomar un sólo punto de cada intervalo como representante de dicho intervalo, utilizando un operador para obtener valores puntuales. Song

entiende que ese punto es el que mejor representa cada conjunto difuso y la medida de dependencia entre los distintos puntos la asume como medida de dependencia entre los distintos conjuntos difusos. Si tiene un grado alto de dependencia, esto respondería a valores altos de la función de auto-correlación, es un indicador para la utilización de STD de orden superior como mejor modelo para la predicción, idea extraída de Brockwell y Davis (1996) que utilizan las funciones de auto-correlación para discriminar entre distintos modelos de predicción. Song (2003) propone dos funciones más para medir dicha dependencia la función de auto-correlación parcial y la función de auto-correlación aleatoria, con el fin de seleccionar el modelo más apropiado para las STD.

Otro de los trabajos relevantes de las series temporales difusas es el de Yu (2005b), que introduce las series temporales difusas ponderadas. Por primera vez, en las relaciones difusas de grupos se tienen en cuenta las recurrencias entre las relaciones difusas, propone una ponderación cronológica y finalmente introduce los pesos calculados en las reglas de predicción. Ese mismo año en Yu (2005a) plantea un modelo de STD tipo-2 para series financieras, utilizando conjuntos difusos tipo-2, donde en cada día se obtienen 3 datos: el máximo diario, el mínimo diario y el valor al cierre. Para adecuar el modelo a las STD, hace una definición particular de los operadores unión e intersección para las RLD y RLDG. Finalmente también propone reglas heurísticas de predicción para este nuevo modelo.

Cheng et al. (2008a) siguiendo la línea de investigación de Yu (2005a), e intentando mejorar el error de ajuste, plantean una nueva ponderación, basada en el cardinal de recurrencias de cada clase de relación difusa. Aunque también usan las mismas reglas de predicción que Yu, hacen un ajuste final de la predicción puntual (propuesto por Kmenta (1986) para disminuir los errores de ajuste). Este ajuste dependerá del valor actual de la serie y del error de predicción cometido en el paso anterior.

En un estudio posterior, Cheng et al. (2009), basándose en este mismo trabajo de Cheng et al. (2008a), aplican las STD al campo de la teoría de la difusión de la innovación (Bass, 1969; Mahajam et al., 1990), concretamente a productos ICT (*Information and Communication Technologies*) ya que se probó en Gharavi y Cheng (2004), que los productos ICT alteran los modelos operacionales de las empresas y conducen a una difusión de la innovación.

A partir de los trabajos presentados por Ting (2003) y Kim et al. (2006) muestran la fuerte correlación existente entre el volumen de operaciones

(compras-ventas) y las fluctuaciones de los precios. En este tipo de tratamiento de los modelos de STD se ve la gran importancia que tiene en el análisis de los indicadores técnicos, es por ello que Chu et al. (2009) incorporan este indicador como segundo factor para la ponderación de STD y presentan un modelo de predicción bi-factor. Este estudio sigue el esquema propuesto por Cheng et al. (2008a).

Chen y Chen (2015) presentan un modelo de dos factores, se habla de un modelo de segundo orden en las RLDG e introducen la tendencia mediante probabilidades extraídas de ciertas frecuencias que hay en RLD, clasificando la tendencia en tres tipos. También proponen una nueva fórmula que determina el tamaño que deben tener los intervalos de la partición del universo.

Cheng y Chen (2018) proponen un algoritmo para la partición inicial del universo y, en función de la longitud y el número de intervalos, definir el universo. También consideran las relaciones lógicas difusas de distintos órdenes. Finalmente, para la predicción, utilizan reglas de asociación difusas ponderadas que están basadas en las reglas de asociación Apriori. Las reglas de asociación, en minería de datos, se utilizan para expresar cuál es el elemento en común dentro de un determinado conjunto de datos.

1.3.6. Aportaciones relevantes en la Etapa 7: Predicción puntual.

Chen y Hsu (2004) introducen una serie de reglas nuevas para la partición del universo y para la predicción mediante un intento de integración de la tendencia dentro del modelo.

Huarng et al. (2007) presentan un modelo multivariante que integra varios modelos univariantes mediante una función heurística multivariante, en la cual están integradas variables heurísticas y los conjuntos difusos observados, y usan como modelo base el modelo clásico propuesto por Chen (1996).

Chen et al. (2007), en un interesante trabajo, proponen integrar en un modelo la serie de Fibonacci, el esquema de trabajo propuesto por Song y Chissom (1993a) y el método de ponderación de Yu (2005a). Representan con patrones las fluctuaciones existentes en el histórico de datos, también utilizan las amplitudes de cada valor lingüístico como valor para la predicción puntual y finalmente la serie de Fibonacci, basándose en el principio de Ondas de Elliot (Fischer y Fischer, 2001; Elliot, 1994; Mandelbrot y Hudson, 2004), que se introduce en el proceso de predicción.

Cheng et al. (2008c) trabajan con bases de datos de visitas externas a pacientes en un hospital, mostrando la versatilidad de las STD. En este trabajo se proponen dos modelos bastante similares en su forma y esquema pero distintos en la predicción: *Expectation Method* y *Grade-Selection Method*. Básicamente es una reinterpretación de los modelos propuestos por Chen (1996) y Chen y Hsiao (2006) haciendo algunas variaciones sobre todo en las observaciones que las clasifican entre visibles y ocultas, dependiendo del valor de la función de pertenencia. El *Expectation Method* se basa en la interpretación difusa del valor esperado de una función sobre cierta variable x , y el *Grade-Selection Method* consta de una serie de reglas heurísticas que intentan modelizar la tendencia de la serie.

Siguiendo el esquema básico descrito por Chen (1996), Stevenson y Porter (2009) proponen usar el porcentaje de cambio observación-a-observación como universo de datos, tomando de Chen y Hsu (2004), con pequeñas modificaciones, la partición del universo. Y para la predicción proponen una fórmula basada en las tasas de cambio de las transiciones entre dos observaciones consecutivas.

Leu et al. (2009) sigue el modelo propuesto por Lee et al. (2006), introduciendo como novedad lo que denomina distancia-base en series temporales difusas, este concepto se basa en la distancia euclídea entre los conjuntos difusos antecedentes y los consecuentes, introduciendo dicha distancia como parámetro para la predicción. Plantean un modelo multi-factor de orden superior, que también intenta resolver algunas de las carencias observadas en trabajos anteriores, también incide más en la importancia del segundo factor dedicándole más trabajo y análisis.

Arutchelvan et al. (2010) proponen un método de predicción en las STD de orden superior y variante en el tiempo, aplicado a un histórico de datos de accidentes. Un trabajo con gran similitud a Stevenson y Porter (2009), y salvando las distancias tomando ideas de Chen y Hsu (2004) y nutriéndose de la base de datos proporcionada por Jilani y Burney (2008).

Panigrahi y Behera (2018) proponen un método de orden superior para hacer predicciones simultáneas a varias series temporales. El orden del método lo determinan analizando las funciones de auto-correlación y auto-correlación parcial. Para la longitud del intervalo se usa un algoritmo de optimización y finalmente para la predicción puntual se utiliza la media de los intervalos.

1.3.7. Otras aproximaciones para la predicción difusa

Hwang et al. (1998) proponen un modelo que partiendo del trabajo anterior de Chen (1996), intentan trabajar con la tendencia de la serie, en el que se utilizaba las variaciones sufridas en la serie entre observaciones consecutivas como el histórico de datos, incluyendo una serie de reglas heurísticas que difieren de las RFG, y acaba con una predicción de la variación. Este tratamiento que hace del problema es novedoso desde el punto de vista conceptual de las series temporales, porque es la primera vez que en las STD se trata la tendencia subyacente en la serie temporal. En un trabajo anterior, Chang (1997) trata la estacionalidad difusa para hacer predicciones en series temporales pero no en el marco de las STD. Así tendencia y estacionalidad empiezan a introducirse en el campo de la lógica difusa.

Tseng et al. (1999) también estudiaron la estacionalidad difusa para series temporales, aunque tampoco en el ámbito de las STD. Este incipiente interés en la estacionalidad, hace que finalmente Song (1999) haciendo una serie de modificaciones de sus trabajos originales (Song y Chissom, 1993a,b), presente la estacionalidad en las STD.

Hasta este momento todos los trabajos de las STD eran modelos de una sola variable (rendimiento, precio, cotización, volumen de compra-venta, etc.), Chen y Hwang (2000) presentan por primera vez un modelo de dos variables basado en Hwang et al. (1998), dejando la puerta abierta a un modelo multi-variable más general. Y también abre el abanico de aplicaciones de las STD para otro tipo de datos, como son las series temporales de temperaturas; hasta este momento la mayoría de trabajos se centraban en la predicción del número de matriculados en la Universidad de Alabama (Song y Chissom, 1993a).

Lee et al. (2006) proponen un nuevo enfoque que va muy dirigido a la inversión en valores financieros. En ese artículo se utiliza la teoría de velas japonesas (Joan, 2009). Usan como referencia la series temporales simbólicas (Keogh y Kasetty, 2003; Daw et al., 2003), buscando patrones en las series temporales simbólicas tomando las velas japonesas como símbolos (Nison, 2001; Keogh et al., 2002). En cada observación obtienen cuatro datos: precio a la apertura, precio al cierre, máximo diario y mínimo diario; y con cada observación se construye una vela, posteriormente el conjunto de velas las recodifican para obtener la serie temporal simbólica, utilizando conjuntos difusos triangulares (para la tendencia) y trapezoidales (para los datos de

entrenamiento).

Egrioglu et al. (2009) incorporan métodos clásicos de predicción (como es el SARIMA) al campo de las STD, mediante un modelo híbrido de predicción, con la intención de tratar series temporales difusas estacionales. Utiliza el método de Box-Jenkins para determinar el orden del modelo (Box y Jenkins, 1976), por lo tanto el modelo SARIMA resultante de este método proporciona valores puntuales usados como residuos, introduciéndolos en el modelo de predicción como una segunda variable.

Sadaei y Lee (2014) queriendo resumir, en general, todas las investigaciones de las STD propone un nuevo esquema de trabajo. Aunque en líneas generales no difiere de la algoritmia descrita por Song y Chissom (1993a), Chen (1996) o Yu (2005b), lo único novedoso es la proposición de un paso previo antes de empezar a definir el universo del histórico de datos. Lo que hace es un procesamiento de los datos, con el cual pretende estabilizar la varianza y la media del histórico, esto lo consigue filtrando y eliminando el ruido de los datos, para ello propone usar el ROI (*Return on Investment*), aunque en un trabajo previo de Stevenson y Porter (2009) se había presentado un procedimiento similar.

También es interesante el trabajo de Efendi et al. (2015), proponiendo múltiples cambios. Primero en la partición del universo, donde proponen aplicar una fórmula para estimar el número de intervalos que debería tener el universo. Segundo, haciendo una ponderación basada en los subíndices de los conjuntos difusos de las RLDG. Y tercero con predicciones sobre los subíndices de los conjuntos difusos.

Nguyen y Novák (2019) se limitan a hacer predicciones en series temporales estacionales. Las series temporales son descompuestas en tendencias, una componente estacional y un término que representa las fluctuaciones irregulares. Se hace una predicción de la tendencia y la componente estacional mediante una transformación difusa, utilizando patrones y técnicas de lógica difusa natural. Finalmente la predicción es una combinación de las predicciones de la tendencia y la componente estacional, ajustando en última instancia con el término de las fluctuaciones irregulares.

Wu et al. (2019) introducen un paso más en el proceso de predicción utilizando el error de aprendizaje para mejorar las predicciones. Para el número de intervalos utilizan el criterio difuso de Silhouette (Campello y Hruschka, 2006) y para la longitud de los intervalos usan la técnica de clustering c-

media difusas. Finalmente mediante las reglas de predicción y junto con un modelo de optimización lineal utilizando los errores de ajuste, realizan las predicciones puntuales.

Una vez visto el estado de la cuestión y los principales conceptos empleados en esta memoria, en el siguiente capítulo se presentan las principales aproximaciones al problema de selección de carteras (clásicas y difusas) y la predicción difusa a partir de las series temporales de conjuntos difusos.

Predicción difusa de rendimientos de carteras

En este capítulo se explica cómo se calcula una predicción difusa del rendimiento de una cartera previamente determinada, a la vez que se proponen nuevas estrategias durante el proceso de predicción. Adicionalmente se introduce la formulación del problema de selección de carteras y los principales modelos de optimización asociados, ya que hay algunos conceptos importantes que se utilizan en esta memoria.

Introducimos nuestras aportaciones a las STD para hacer predicciones difusas de los rendimientos de carteras. Se ven cómo funcionan los principales algoritmos en series temporales difusas y se aplican sobre un ejemplo concreto de cartera. Se propone un experimento para poder realizar un análisis estadístico de los resultados obtenidos con las diferentes aplicaciones de las STD.

Se concluye con la aplicación de las medidas predictivas difusas al problema de selección de carteras, junto con un resumen del capítulo y las conclusiones.

Parte del contenido de este capítulo viene recogido en el artículo Rubio et al. (2016) publicado en la revista *International Journal of Approximate Reasoning (IJAR)*.

2.1. Formulación del problema de selección de carteras

Se plantea el problema de selección de carteras: considérense n activos financieros que cotizan en un mercado de valores, donde los rendimientos futuros que se obtienen son inciertos. Se pretende seleccionar una cartera óptima donde invertir cierto capital, que se asegure un rendimiento (lo más alto posible) y que también se minimice el riesgo de la inversión.

Se estudia primero el modelo clásico de Markowitz (1952) que presenta la formulación estándar del modelo de media-varianza.

El rendimiento de una inversión se define como $x'R$, siendo R el vector aleatorio de rendimientos de los n activos consecutivos y x el vector que representa la proporción de capital invertido en cada activo en tantos por uno. Entonces el rendimiento esperado de la inversión viene dado por:

$$E(x'R) = E\left[\sum_{i=1}^n R_i x_i\right] = \sum_{i=1}^n E[R_i] x_i \quad (2.1)$$

donde $E[\cdot]$ representa el valor esperado de una variable aleatoria. Y el riesgo de la inversión se calcula a partir de la matriz de varianzas-covarianzas de R :

$$V(x'R) = x'\Sigma x = \sum_{j=1}^n \sum_{i=1}^n \sigma_{ij} x_i x_j \quad (2.2)$$

donde σ_{ij} es la covarianza entre R_i y R_j ($\sigma_{ij} = E[(R_i - E(R_i))(R_j - E(R_j))]$). El modelo MV que Markowitz propone, es como sigue:

$$\text{Min}_x \sum_{j=1}^n \sum_{i=1}^n \sigma_{ij} x_i x_j$$

s.a.

$$\sum_{i=1}^n E(R_i) x_i = \rho \quad (2.3)$$

$$\sum_{i=1}^n x_i = 1 \quad (2.3a)$$

$$x_i \geq 0 \quad i = 1, \dots, n \quad (2.3b)$$

donde ρ es el nivel de rendimiento deseado. En este problema Markowitz añade las restricciones (2.3a) y (2.3b), con las que pretende por un lado ase-

gurar que no hayan proporciones negativas del capital invertido y por otro obliga a que se invierta el cien por cien del capital dispuesto, siendo la suma de todas las proporciones igual a uno.

A partir del modelo MV de Markovitz es posible plantear un problema de optimización bi-objetivo, minimizando el riesgo y maximizando el rendimiento esperado. Obsérvese que para diferentes elecciones de las proporciones x_1, x_2, \dots, x_n el inversor obtendrá diferentes combinaciones del rendimiento esperado y del riesgo. Al conjunto de todas las posibles combinaciones del rendimiento esperado y del riesgo se le llama conjunto realizable. De entre todas las combinaciones del rendimiento esperado de la cartera y del riesgo, el inversor puede elegir una cartera dependiendo de su aversión al riesgo. Una solución eficiente del problema bi-objetivo y una forma de tratar esta aversión al riesgo es considerar el problema de optimización:

$$\text{Min}_x \quad V(x'R) - AE(x'R)$$

s.a.

$$\sum_{i=1}^n x_i = 1$$

$$x_i \geq 0 \quad i = 1, \dots, n$$

donde $0 \leq A \leq \infty$ es un índice de aversión al riesgo. Estudiando los posibles valores de A se pueden obtener diferentes soluciones con diferentes niveles de riesgo, si $A = 0$ se obtendrá una cartera con un nivel de riesgo menor, a medida que se va incrementando el valor de A se irá incrementando el riesgo, aunque también corresponderá a un rendimiento esperado mayor.

En lo sucesivo, los momentos (esperanza y varianza) se estiman con los momentos muestrales de un histórico de datos. Sea T el período de tiempo a estudiar y sea r_{it} el rendimiento del activo i en los instantes $t = 1, 2, \dots, T$

$$r_i = E[R_i] = \frac{1}{T} \sum_{t=1}^T r_{it} \quad (2.4)$$

donde R_i es la variable aleatoria de rendimientos del activo i .

Konno y Yamazaki (1991) proponen una medida alternativa del riesgo de la inversión tomando la desviación media absoluta como nueva función objetivo (modelo MAD). La fácil transformación de este modelo en un problema de programación lineal hizo que fuera un adelanto muy importante

en la resolución de este tipo de problemas de selección de carteras, sobre todo cuando se trataban grandes bases de datos de rendimientos de activos financieros.

El modelo MAD es de la forma:

$$\begin{aligned} \text{Min}_x \quad & \frac{1}{T} \sum_{t=1}^T \left| \sum_{i=1}^n (r_{it} - r_{i\cdot}) x_i \right| \\ \text{s.a.} \quad & \\ & \sum_{i=1}^n x_i r_{i\cdot} \geq \rho C_0 \\ & \sum_{i=1}^n x_i = 1 \\ & 0 \leq x_i \leq u_i \quad i = 1, \dots, n \end{aligned}$$

donde C_0 es el capital disponible y u_i es la máxima cantidad de dinero que se puede invertir en el activo i .

Tomando $z_t = \left| \sum_{i=1}^n (r_{it} - r_{i\cdot}) x_i \right|$ donde $z_t \geq 0$ y añadiendo las restricciones al modelo MAD

$$\begin{aligned} z_t + \sum_{i=1}^n (r_{it} - r_{i\cdot}) x_i &\geq 0 \quad t = 1, \dots, T \\ z_t - \sum_{i=1}^n (r_{it} - r_{i\cdot}) x_i &\geq 0 \quad t = 1, \dots, T \end{aligned}$$

se obtiene un modelo lineal, denominado como LMAD, el cual tiene como función objetivo: $\frac{1}{T} \sum_{t=1}^T z_t$.

Konno y Yamazaki (1991) prueban que los modelos MV y LMAD son equivalentes cuando la distribución conjunta de los rendimientos de los n activos tiene una distribución normal multivariante.

Modificando el modelo LMAD, Speranza (1993) usa como medida del riesgo la semi-desviación absoluta:

$$z'_t = \left| \min \left\{ 0, \sum_{i=1}^n (r_{it} - r_{i\cdot}) x_i \right\} \right|$$

entonces en este modelo propuesto por Speranza tan solo se necesitarían las restricciones: $z'_t + \sum_{i=1}^n (r_{it} - r_i)x_i \geq 0 \quad t = 1, \dots, T$, lo que resulta en un menor número de restricciones que en el modelo LMAD. Speranza prueba que ambos modelos (con z_t y z'_t) son equivalentes siempre que se asuma la distribución normal multivariante de los rendimientos.

Es posible introducir nuevas restricciones operativas provenientes de especificaciones del mercado de valores, como los lotes de compras (el mínimo número de acciones que es posible comprar de cierto activo bursátil en el mercado de valores).

Sea el lote c_j el precio de compra del mínimo número de acciones del activo j . Donde $c_j = N_j p_j$ siendo p_j el precio de compra por acción del activo j y N_j el número mínimo de acciones que requiere el lote del activo j . Ahora la variable x_j es una variable entera y representa el número de lotes por activo j . Así se tiene el modelo MILP (modelo mixto de programación lineal entera) que con restricciones de lotes mínimos se quedaría un modelo semi-LMAD:

$$\begin{aligned}
\text{Min}_x \quad & \frac{1}{T} \sum_{t=1}^T z'_t \\
\text{s.a.} \quad & z'_t + \sum_{i=1}^n (r_{it} - r_i)x_i \geq 0 \quad t = 1, \dots, T \\
& \sum_{i=1}^n x_i r_i c_i \geq \rho C \\
& C_0 \leq \sum_{i=1}^n x_i c_i \leq C_1 \\
& 0 \leq x_i \leq u_i \quad x_i \in \mathbb{Z}^+, i = 1, \dots, n \\
& z'_t \geq 0 \quad t = 1, \dots, T
\end{aligned}$$

donde C es la inversión total de la cartera, nótese también que la inversión total no tiene porque coincidir con todo el capital disponible para invertir, esta delimitado entre C_0 y C_1 .

Posteriormente Mansini y Speranza (1999) desarrollaron un algoritmo de programación lineal entera mixta que resolvía una versión reducida del problema de selección de carteras, lo formularon como una desviación media

general absoluta que reemplaza a la función objetivo del modelo MV.

A continuación véase cómo la incertidumbre sobre el rendimiento futuro y el riesgo que subyace del comportamiento errático de una cartera, en un entorno financiero cambiante, pueden ser aproximados desde un punto de vista difuso.

2.1.1. Aproximación difusa al problema de selección de carteras

Los principales enfoques del problema de optimización de selección de carteras tienen como objetivo maximizar el rendimiento de la inversión, al mismo tiempo que se minimiza el riesgo. Como ya se sabe el riesgo y el rendimiento son medidas cuantitativas que guardan una cierta correlación. Además, está generalmente aceptado que el rendimiento futuro o el riesgo de la inversión son inciertos y, por lo tanto, la teoría de conjuntos difusos puede ser utilizada para describir el comportamiento de los rendimientos o riesgos de los activos, así como de otras variables que puedan afectar al mercado financiero.

La teoría de conjuntos difusos puede modelar el problema de selección de carteras por medio de objetivos y restricciones difusas, y resolver problemas de optimización utilizando técnicas de Soft Computing (Watada, 1997; Tanaka y Guo, 1999; Inuiguchi y Ramík, 2000; Carlsson et al., 2002).

Se plantean dos estrategias de modelización: una es construir conjuntos difusos a partir de cada uno de los rendimientos de los activos y luego conformar la cartera, y otra es construir el conjunto difuso que representa el rendimiento de toda la cartera una vez ya determinada.

Primero se estudia la primera de las estrategias de modelización del problema de selección de carteras construyendo conjuntos difusos a partir de cada uno de los rendimientos de activos. La incertidumbre alrededor del rendimiento del activo i -ésimo puede ser representado como el número difuso trapezoidal $D_i = (a_i, b_i, \gamma_i, \beta_i)$ (definido en 1.15).

Según la definición de número trapezoidal la función de pertenencia de un número difuso trapezoidal es como sigue:

$$\mu_{D_i}(x) = \begin{cases} \frac{x-a_i+\gamma_i}{\gamma_i} & \text{si } x \in [a_i - \gamma_i, a_i] \\ 1 & \text{si } x \in [a_i, b_i] \\ \frac{b_i+\beta_i-x}{\beta_i} & \text{si } x \in [b_i, b_i + \beta_i] \\ 0 & \text{en otro caso} \end{cases}$$

Tanto el rendimiento esperado como el riesgo del activo i -ésimo pueden ser estimados usando el intervalo posibilístico de medias y el riesgo lateral difuso de D_i respectivamente (Dubois y Prade, 1987). El intervalo posibilístico de medias del número difuso D_i (Dubois y Prade, 1987) se define como:

$$I_m(D_i) = \left[a_i - \frac{\gamma_i}{2}, b_i + \frac{\beta_i}{2} \right] \quad (2.5)$$

Con esta primera estrategia se está planteando minimizar el riesgo lateral de cada activo, a partir de la definición de cotas inferiores para los rendimientos esperados de cada uno de los activos. Por un lado se tiene que el valor del rendimiento esperado del activo i -ésimo es el punto medio del intervalo posibilístico de medias:

$$E(D_i) = \frac{a_i + b_i}{2} + \frac{\beta_i - \gamma_i}{4} \quad (2.6)$$

Por otro lado se tiene que el riesgo lateral del activo i -ésimo mide el fracaso para alcanzar el valor esperado del intervalo posibilístico de medias utilizando la semi-desviación media absoluta, la cual penaliza sólo las desviaciones negativas del rendimiento esperado (Bermúdez et al., 2007a).

Para el número difuso trapezoidal D_i , su riesgo lateral coincide con la longitud del intervalo de $I_m(D_i)$, es decir, la amplitud del intervalo del α_i -corte para $\alpha_i = 0.5$, mientras que la ambigüedad (definida en 1.19) corresponderá a la semi-amplitud de dicho intervalo $I_m(D_i)$, Vercher et al. (2007). Así el riesgo lateral sería:

$$w(D_i) = b_i - a_i + \frac{\beta_i + \gamma_i}{2} \quad (2.7)$$

Finalmente se obtiene un modelo MAD cambiando medidas probabilistas por medidas posibilistas:

$$\begin{aligned}
\text{Min}_x \quad & \sum_{i=1}^n (b_i - a_i + \frac{1}{2}(\gamma_i + \beta_i))x_i \\
\text{s.a.} \quad & \sum_{i=1}^n \left(\frac{1}{2}(a_i + b_i) + \frac{1}{4}(\beta_i - \gamma_i) \right) x_i \geq \rho \\
& \sum_{i=1}^n x_i = 1 \\
& 0 \leq l_i \leq x_i \leq u_i \quad i = 1, \dots, n
\end{aligned} \tag{2.8}$$

$$\tag{2.8a}$$

la última desigualdad (2.8a) representa una restricción de diversificación, donde l_i es la cantidad mínima de inversión sobre el activo i -ésimo (respectivamente u_i representa el máximo).

Ahora véase la segunda estrategia de modelización difusa, que consiste en construir un conjunto difuso que represente a toda la cartera. Este enfoque es relativamente nuevo (Bermúdez et al., 2007a) y es un cambio importante en el concepto a la hora de tratar el problema de optimización.

Sea X una cartera, donde x_1, \dots, x_n es una elección de las proporciones de n activos financieros y sea r_{it} el rendimiento del activo i -ésimo en el instante t , se define el rendimiento de la cartera en el instante t como:

$$r_t = \sum_{i=1}^n x_i r_{it} \text{ para } t = 1, \dots, T \tag{2.9}$$

La serie temporal original de rendimientos de los activos, una vez establecidas las proporciones de cada uno de los activos, induce una serie temporal del rendimiento de una cartera. Análogamente a la estrategia anterior $D_X = (a, b, \gamma, \beta)$ corresponde a la representación difusa de los rendimientos de la cartera X .

Adicionalmente se añaden restricciones de cardinalidad (Gupta et al., 2011), controlando así el número total de activos que intervienen en la cartera. El modelo conjunto utilizando restricciones de cardinalidad, quedaría como

sigue:

$$\begin{aligned}
& \underset{x}{\text{Min}} \quad w(D_X) \\
& \text{s.a.} \\
& \quad E(D_X) \geq \rho \\
& \quad \sum_{i=1}^n x_i = 1 \\
& \quad y_i l_i \leq x_i \leq y_i u_i \quad i = 1, \dots, n \\
& \quad y_i \in \{0, 1\} \\
& \quad K_1 \leq \sum_{i=1}^n y_i \leq K_2
\end{aligned} \tag{2.10}$$

$$\tag{2.10a}$$

donde la última desigualdad (2.10a) es la restricción de cardinalidad y expresa que el número total de activos debe estar comprendido entre un valor mínimo K_1 y un valor máximo K_2 . También se tiene que $w(D_X) = b - a + \frac{\beta + \gamma}{2}$ y $E(D_X) = \frac{a+b}{2} + \frac{\beta - \gamma}{4}$.

Nótese que para cada cartera X se obtiene una serie temporal de rendimientos y que por lo tanto se construyen números difusos trapezoidales distintos. Además si se opta por usar, en el problema de selección de carteras las restricciones de cardinalidad, se convierte en un problema NP-hard, lo que imposibilita la resolución del problema de optimización mediante procedimientos clásicos de optimización (Bertsimas y Shioda, 2009), es por ello que se suelen utilizar algoritmos heurísticos o meta-heurísticos para abordar este tipo de problemas.

Dado que el objetivo principal del problema de la optimización de carteras es minimizar el riesgo, es importante decidir cuál de las estrategias anteriores se elige, para tratar con modelos de selección de carteras posibilísticos. Una explicación más detallada del uso de estas medidas difusas para los posibles problemas de selección de carteras se encuentra en Vercher et al. (2007); Bermúdez et al. (2012), donde también se presentan algunas relaciones entre diferentes definiciones del intervalo posibilístico de medias.

Uno de los objetivos de esta tesis es en el problema de selección de carteras sustituir las medidas posibilistas de rendimiento esperado y riesgo por medidas predictivas difusas (posibilistas o credibilistas). En este capítulo se trata la cartera en su conjunto, optando por la segunda de las estrategias, haciéndose predicciones difusas de la rentabilidad de una cartera con el objetivo

de utilizar los distintos momentos posibilísticos de la predicción difusa para introducirlos en el problema de selección de carteras como medidas alternativas. A continuación véase como funcionan algunos de los modelos clásicos de predicción de series temporales difusas.

2.2. Predicción puntual en series temporales difusas

A continuación se estudia cómo funciona el algoritmo básico de Chen (1996), introducido en la sección 1.2, sobre un ejemplo práctico de una cartera concreta. Adicionalmente también se estudian el funcionamiento de los principales algoritmos de series temporales difusas ponderadas Yu (2005b); Cheng et al. (2008a), ambos algoritmos están incluidos en la revisión bibliográfica sección 1.3.5.

A lo largo de este capítulo tanto en los ejemplos como en los experimentos numéricos, con tal de reducir el número de citas a los diferentes autores y que no se creen excesivas repeticiones de las citas, se denota: Chen (1996) como Chen, Yu (2005b) como Yu y Cheng et al. (2008a) como Cheng.

Ejemplo: considérese una cartera de valores X compuesta por 10 activos financieros incluidos en el IBEX35 entre los años 2011 y 2013, tómense 152 rendimientos semanales porcentuales $(r_t)_{t=1}^{152}$ de dicha cartera X , con $\text{rango}(r_t) = [-12.67, 12.95]$.

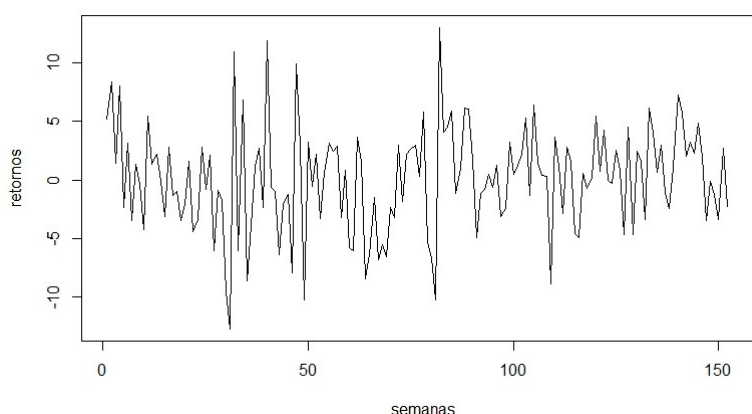


Figura 2.1: Trazo del conjunto de datos de rendimientos de una cartera X .

Esta cartera se obtiene por simulación, de hecho es la primera obtenida en

el experimento numérico descrito con detalle más adelante, en el apartado 2.4.

Se aplican las etapas principales de los algoritmos de Chen, Cheng y Yu:

Etapa 1. Se define el universo:

Los tres algoritmos definen el mismo universo $U = [-15, 15]$.

Etapa 2. Se hace una partición del universo y se definen los intervalos que recubrirán el universo U , de modo que alternativamente:

- El algoritmo de Chen parte en 7 intervalos iguales el universo, luego la longitud de los intervalos es 4.29 ($= 30/7$), obteniendo los siguientes intervalos:

	\underline{u}_i	\bar{u}_i	FDO
u_1	-15.00	-10.71	1
u_2	-10.71	-6.43	10
u_3	-6.43	-2.14	34
u_4	-2.14	2.14	56
u_5	2.14	6.43	43
u_6	6.43	10.71	5
u_7	10.71	15.00	3

Cuadro 2.1: Intervalos u_i y frecuencias de datos observados

donde \underline{u}_i es el extremo inferior del intervalo u_i , \bar{u}_i es el extremo superior del intervalo u_i y la columna FDO (frecuencia de datos observados) representa al número de datos observados que pertenecen al intervalo u_i .

- El algoritmo de Cheng para la partición del universo U aplica una idea propuesta por Chen y Hsu (2004). Se calcula la media de las frecuencias de datos observados (anteriormente considerado según Chen) que es igual a 21.71 ($= 152/7$), y a partir del cuadro 2.1, cada uno de los intervalos con mayor frecuencia que la media calculada (i.e. u_3, u_4, u_5) se subdivide en dos, obteniendo un total de 10 intervalos, quedando la siguiente partición:
- El algoritmo de Yu utiliza el procedimiento de Huarng (2001a) para realizar la partición. De dicho procedimiento resulta que las particiones tienen longitud 2.14, y que por lo tanto se obtiene un total de 14 intervalos.
 $u_1 = [-15, -12.86], u_2 = [-12.86, -10.71], \dots, u_{14} = [12.86, 15]$

	\underline{u}_i	\bar{u}_i
u_1	-15.00	-10.71
u_2	-10.71	-6.43
u_3	-6.43	-4.29
u_4	-4.29	-2.14
u_5	-2.14	0.00
u_6	0.00	2.14
u_7	2.14	4.29
u_8	4.29	6.43
u_9	6.43	10.71
u_{10}	10.71	15.00

Etapa 3. Se determinan las variables A_i representadas por conjuntos difusos definidos sobre cada una de las particiones.
 Los tres algoritmos definen los conjuntos difusos A_i de la misma forma, con función de pertenencia:

$$\mu_1(x) = \begin{cases} 1 & \text{si } x \in u_1 \\ 0.5 & \text{si } x \in u_2 \\ 0 & \text{resto de casos} \end{cases} \quad \mu_{k_{max}}(x) = \begin{cases} 0.5 & \text{si } x \in u_{k_{max}-1} \\ 1 & \text{si } x \in u_{k_{max}} \\ 0 & \text{resto de casos} \end{cases}$$

$$\mu_i(x) = \begin{cases} 0.5 & \text{si } x \in u_{i-1} \\ 1 & \text{si } x \in u_i \\ 0.5 & \text{si } x \in u_{i+1} \\ 0 & \text{resto de casos} \end{cases} \quad \text{para } 1 < i < k_{max}$$

Donde k_{max} es el mayor de los índices, en el ejemplo en estudio sería $k_{max} = 7, 10, 14$ (según el algoritmo utilizado Chen, Cheng o Yu).

Etapa 4. Se clasifican los datos entre los diferentes conjuntos difusos.
 Los tres algoritmos proceden de forma similar.

$$F(t) = A_i \iff \mu_{A_i}(r_t) = \max_k(\mu_{A_k}(r_t))$$

En el caso a tratar, por construcción se tiene que:

$$\text{si } r_t \in u_i \mapsto F(t) = A_i$$

Etapa 5. Se establecen las relaciones lógicas difusas (RLD) entre los diferentes conjuntos A_i , las cuales están inducidas por las relaciones existentes en el histórico de datos inicial.

Los tres algoritmos proceden de la misma forma.

si $F(t) = A_{i_0}$ y $F(t + 1) = A_{i_1}$ para $t = 1, \dots, 151$ entonces la relación lógica difusa es $A_{i_0} \rightarrow A_{i_1}$. Obteniendo un total de 151 relaciones lógicas difusas.

Se tiene que la serie de rendimientos es $r_1 = 5.21, r_2 = 8.36, \dots, r_{151} = 2.66, r_{152} = -2.24$. De modo que

- Aplicando Chen: se tiene que $r_1 \in u_5 \mapsto F(1) = A_5$ y $r_2 \in u_6 \mapsto F(2) = A_6$, induce a la relación lógica difusa $A_5 \rightarrow A_6$. Aplicado al conjunto de datos se observa:
 $A_5 \rightarrow A_6 \rightarrow A_4 \rightarrow \dots \rightarrow A_5 \rightarrow A_3$
- Aplicando Cheng: se tiene que $r_1 \in u_8 \mapsto F(1) = A_8$ y $r_2 \in u_9 \mapsto F(2) = A_9$, induce a la relación lógica difusa $A_8 \rightarrow A_9$. Aplicado al conjunto de datos se observa:
 $A_8 \rightarrow A_9 \rightarrow A_6 \rightarrow \dots \rightarrow A_7 \rightarrow A_4$
- Aplicando Yu: se tiene que $r_1 \in u_{10} \mapsto F(1) = A_{10}$ y $r_2 \in u_{11} \mapsto F(2) = A_{11}$, induce a la relación lógica difusa $A_{10} \rightarrow A_{11}$. Aplicado al conjunto de datos se observa:
 $A_{10} \rightarrow A_{11} \rightarrow A_8 \rightarrow \dots \rightarrow A_9 \rightarrow A_6$

Etapa 6. Se construyen las relaciones lógicas difusas de grupos (RLDG), agrupando todas aquellas relaciones difusas donde la parte izquierda de la relación es la misma, o sea, tienen el mismo antecedente. Sea la relación difusa $A_i \rightarrow A_k$, la parte izquierda de la relación difusa es A_i (o antecedente) y la parte derecha de la relación difusa es A_k (o consecuente).

Si $\dots A_i \rightarrow A_{k1} \rightarrow \dots \rightarrow A_i \rightarrow A_{k2} \dots$ las relaciones lógicas difusas se pueden agrupar: $A_i \rightarrow A_{k1}, A_{k2}$, conformando la RLDG asociada a A_i . En este caso la parte derecha de la RLDG asociada a A_i contiene dos conjuntos difusos (A_{k1}, A_{k2}).

Para los tres algoritmos desarrollados:

- En el algoritmo de Chen no hay recurrencias en las relaciones difusas de grupos, esto es, en la parte derecha de las relaciones difusas de grupos no se permiten repeticiones de conjuntos difusos y el orden o cronología de observación de los conjuntos difusos no es importante.
 $A_1 \rightarrow A_7$
 $A_2 \rightarrow A_1, A_2, A_3, A_5, A_6, A_7$
 $A_3 \rightarrow A_2, A_3, A_4, A_5, A_6, A_7$
 $A_4 \rightarrow A_2, A_3, A_4, A_5, A_6$

$$A_5 \rightarrow A_2, A_3, A_4, A_5, A_6$$

$$A_6 \rightarrow A_2, A_3, A_4, A_5$$

$$A_7 \rightarrow A_3, A_4, A_5$$

- En el algoritmo de Cheng si se permiten recurrencias, pero el orden o cronología de observación tampoco es importante.

$$A_1 \rightarrow A_{10}$$

$$A_2 \rightarrow A_1, A_4, A_9, A_7, A_3, \dots$$

$$A_3 \rightarrow A_4, A_5, A_9, A_5, A_3, A_7, \dots$$

$$A_4 \rightarrow A_7, A_6, A_8, A_7, A_5, A_7, \dots$$

$$A_5 \rightarrow A_4, A_4, A_5, A_4, A_6, A_6, \dots$$

$$A_6 \rightarrow A_9, A_5, A_7, A_3, A_3, A_7, \dots$$

$$A_7 \rightarrow A_4, A_5, A_5, A_5, A_4, A_2 \dots$$

$$A_8 \rightarrow A_9, A_6, A_3, A_8, A_5, A_8, \dots$$

$$A_9 \rightarrow A_6, A_4, A_2, A_7, A_8$$

$$A_{10} \rightarrow A_3, A_5, A_7$$

- En el algoritmo de Yu sí se permiten recurrencias y además el orden o cronología de observación sí es importante.

$$A_1 \rightarrow \emptyset$$

$$A_2 \rightarrow A_{13}$$

$$A_3 \rightarrow A_2, A_6, A_9, A_{14}$$

$$A_4 \rightarrow A_{12}, A_5, A_5, A_6, A_3$$

$$A_5 \rightarrow A_6, A_7, A_{11}, A_7, A_5, A_9 \dots$$

$$A_6 \rightarrow A_9, A_8, A_{10}, A_9, A_7, A_9, \dots$$

...

$$A_{12} \rightarrow A_9$$

$$A_{13} \rightarrow A_5, A_7$$

$$A_{14} \rightarrow A_9$$

Etapa 7. Se calcula el valor puntual de la predicción, a partir de unas reglas de predicción. Para ello se calculan los puntos medios de los intervalos que componen los conjuntos difusos:

Sea m_i el punto medio del intervalo u_i , representante del conjunto difuso A_i para $i = 1, 2, \dots, k_{max}$.

- En el algoritmo de Chen: la última observación $r_{152} = -2.24 \in u_3 = [-6.43, -2.14] \mapsto F(152) = A_3$, así la predicción puntual es la media aritmética de los puntos medios (m_i) asociados a los conjuntos difusos que componen la parte derecha de las relaciones lógicas difusas de grupos, el ejemplo está en la fila A_3 .

Se tiene $A_3 \rightarrow A_2, A_3, A_4, A_5, A_6, A_7$, el vector, ya normalizado, asociado a esta RLDG sería

$$w = [0, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6]$$

y el vector de puntos medios

$$m = [m_1, m_2, m_3, m_4, m_5, m_6, m_7]$$

$$m = [-12.86, -8.57, -4.29, 0.00, 4.29, 8.57, 12.86]$$

finalmente la predicción puntual sería:

$$\hat{r}_{153} = w'm = \frac{m_2 + m_3 + m_4 + m_5 + m_6 + m_7}{6} = 2.14$$

Ahora bien, la observación real es $r_{153} = -0.87$, luego el error cometido es $e = |-0.87 - 2.14| = 3.01$

- En el algoritmo de Cheng: la última observación $r_{152} = -2.24 \in u_4 = [-4.29, -2.14] \mapsto F(152) = A_4$, así la predicción puntual es la combinación lineal ponderada de los puntos medios que componen la parte derecha de las relaciones lógicas difusas de grupos de la fila A_4 .

En general, se puede representar matricialmente los pesos asociados a las RLDG, $W = (w_{i,j})$, correspondiendo la fila i -ésima al conjunto difuso A_i y la columna j -ésima al conjunto A_j para $i, j = 1, \dots, k_{max}$, calculándose de la siguiente forma:

sea p_i el tamaño de la parte derecha de las RLDG asociadas al conjunto A_i

$$w_{i,j} = \sum_{t=1}^{p_i} \delta_{i,j}^t \quad (2.11)$$

y

$$\delta_{i,j}^t = \begin{cases} 1 & \text{si } A_j \text{ está en la posición } t \text{ en la RLDG de } A_i \\ 0 & \text{en otro caso} \end{cases}$$

por lo tanto $w_{i,j}$ es la frecuencia de aparición del conjunto A_j en la RLDG asociada a A_i , finalmente se normaliza el vector quedando:

$$w_i = \left[\frac{w_{i,1}}{p_i}, \frac{w_{i,2}}{p_i}, \dots, \frac{w_{i,k_{max}}}{p_i} \right] \quad (2.12)$$

Siguiendo con el ejemplo

$$A_4 \rightarrow \{A_7, A_6, A_8, A_7, A_5, A_7, A_6, A_{10}, \\ A_6, A_6, A_4, A_7, A_4, A_7, A_7, A_8, A_6, A_5, A_7\}$$

por lo tanto $p_4 = 19$. Para ejemplificar el cálculo de los pesos, tómesese como ejemplo $w_{4,1}$ que en este caso es igual a 0 debido a que no hay ninguna relación lógica difusa entre A_4 y A_1 , en cambio $w_{4,4} = 2$ debido a que hay dos relaciones lógicas difusas entre A_4 y A_4 o $w_{4,6} = 5$ hay cinco relaciones lógicas difusas entre A_4 y A_6 , análogamente se procedería igual con el resto de conjuntos difusos. Resultando el vector de pesos asociados y normalizados como sigue:

$$w_4 = \left[0, 0, 0, \frac{2}{19}, \frac{2}{19}, \frac{5}{19}, \frac{7}{19}, \frac{2}{19}, 0, \frac{1}{19}\right]$$

el vector de puntos medios de los intervalos sería:

$$\begin{aligned} m &= [m_1, m_2, m_3, m_4, m_5 \dots, m_{10}] = \\ &= [-12.86, -8.57, -5.36, -3.21, -1.07 \dots, 12.86] \end{aligned}$$

Y la predicción puntual es:

$$\hat{r}_{153} = w_4' m = 2.26$$

La observación real es $r_{153} = -0.87$ luego el error cometido es $e = |-0.87 - 2.26| = 3.12$

- En el algoritmo de Yu: la última observación $r_{152} = -2.24 \in u_6 = [-4.29, -2.14] \mapsto F(152) = A_6$, así la predicción puntual es la combinación lineal ponderada de los puntos medios que componen la parte derecha de las relaciones lógicas difusas de grupos de la fila A_6 .

Al igual que el algoritmo de Cheng, en general, se puede representar matricialmente los pesos asociados a las RLDG, $W = (w_{i,j})$, correspondiendo la fila i -ésima al conjunto difuso A_i y la columna j -ésima a conjunto A_j para $i, j = 1, \dots, k_{max}$, calculándose de la siguiente forma:

sea p_i el tamaño de la parte derecha de las RLDG asociadas al conjunto A_i

$$[w_{i,1}, w_{i,2}, \dots, w_{i,k_{max}}]$$

donde

$$w_{i,j} = \sum_{t=1}^{p_i} \delta_{i,j}^t \quad (2.13)$$

y

$$\delta_{i,j}^t = \begin{cases} t & \text{si } A_j \text{ está en la posición } t \text{ en la RLDG de } A_i \\ 0 & \text{en otro caso} \end{cases}$$

por lo tanto $w_{i,j}$ es la suma de todos los pesos cronológicos asociados al conjunto difuso A_j en la RLDG de A_i , y finalmente normalizando quedaría:

$$w_i = \left[\frac{w_{i,1}}{\sum_{k=1}^{p_6} k}, \frac{w_{i,2}}{\sum_{k=1}^{p_6} k}, \dots, \frac{w_{i,k_{max}}}{\sum_{k=1}^{p_6} k} \right] \quad (2.14)$$

Siguiendo con el ejemplo

$$A_6 \rightarrow \{A_9, A_8, A_{10}, A_9, A_7, A_9, A_8, A_{13}, A_8, A_8, A_6, A_9, A_6, A_9, A_9, A_{10}, A_8, A_7, A_9\}$$

se tiene que $p_6 = 19$ y los pesos asociados a cada uno de los conjuntos según su posición son $\{1, 2, 3, \dots, 19\}$. Tómese como ejemplo $w_{6,1} = 0$ esto es debido a que no hay ninguna relación lógica difusa entre A_6 y A_1 , en cambio $w_{6,7} = 5 + 18 = 23$ porque hay dos relaciones lógicas difusas entre A_6 y A_7 con pesos asociados cinco y dieciocho, que corresponden a la posición relativa de A_7 dentro de la parte derecha de la RLDG asociada a A_6 . Análogamente se procedería igual con el resto de conjuntos difusos. Resultando el vector de pesos asociados y normalizados como sigue:

$$w_6 = [0, 0, 0, 0, 0, \frac{24}{190}, \frac{23}{190}, \frac{45}{190}, \frac{71}{190}, \frac{19}{190}, 0, 0, \frac{8}{190}, 0]$$

el vector de puntos medios de los intervalos sería:

$$m = [m_1, m_2, m_3, \dots, m_{14}] = [-13.93, -11.79, -9.64, \dots, 13.93]$$

Y la predicción puntual es:

$$\hat{r}_{153} = w'_6 m = 1.95$$

La observación real es $r_{153} = -0.87$ luego el error cometido es $e = |-0.87 - 1.95| = 2.82$

Nótese que en los tres algoritmos sólo se hacen predicciones puntuales, es por ello que a continuación se presenta un nuevo algoritmo de predicción difusa ya que es necesario tener un conjunto difuso como predicción porque se pueden extraer diferentes momentos posibilísticos, que en el problema de selección de carteras se pueden usar como medidas predictivas.

2.3. Predicción difusa del rendimiento de una cartera

Nuestro propósito es utilizar métodos de series temporales difusas (STD) para hacer predicciones del rendimiento futuro de carteras previamente definidas (en su composición). En este capítulo se aproxima la incertidumbre del rendimiento futuro de una cartera dada mediante un número difuso trapezoidal, siguiendo la misma estrategia de aproximación conjunta que en la sección 2.1.1. Para ello se estudian las series temporales difusas de los rendimientos de las carteras, no de cada uno de los activos financieros que componen la cartera.

La predicción mediante un número difuso hace posible la aproximación tanto del rendimiento esperado como del riesgo de la inversión a través del intervalo posibilístico de medias. Se introducen algunas modificaciones en los modelos clásicos de series temporales difusas, que permiten generar números trapezoidales difusos como predicción del rendimiento futuro de una cartera de valores.

A continuación véase a grandes rasgos cuáles son las modificaciones propuestas en esta memoria para el procedimiento de predicción de las series temporales difusas desarrollado por Chen (1996). Primero, para la partición inicial del universo de los datos en la etapa 2 considérense todos los intervalos de la misma longitud, dichas longitudes vienen determinadas en función de la desviación típica muestral de los datos. En cuanto a la etapa 3, se definen todas las funciones de pertenencia de cada uno de los conjuntos difusos A_i como números difusos trapezoidales. Finalmente, el método construye predicciones difusas mediante una combinación lineal convexa de números difusos trapezoidales. En dicha combinación lineal la asignación de pesos se hace teniendo en cuenta las relaciones lógicas difusas (RLD) de la serie temporal difusa y las relaciones lógicas difusas de grupos (RLDG), utilizando siempre los pesos ya normalizados.

2.3.1. Nuevo algoritmo propuesto

Primero véase cómo influyen las modificaciones propuestas en el proceso de construcción de los conjuntos difusos que representan la incertidumbre de los datos. Este proceso afecta a las tres primeras etapas del procedimiento desarrollado por Chen (1996).

- En la etapa 1, se define el universo U :
Sea $\{y_t\}_{t=1,\dots,N}$ el conjunto de datos observados donde N es el último instante observado, se tiene que $y_{min} \leq y_t \leq y_{max}$ donde $y_{min} = \min_{t=1,\dots,N} y_t$ es el valor mínimo del conjunto de datos observados (análogamente y_{max} el máximo).
Se construye el universo a partir de la desviación típica muestral de los datos de la serie, tal que $u_{inf} = y_{min} - \epsilon$ donde $\epsilon \in \mathbb{R}$ y por otro lado $u_{sup} = u_{inf} + m\sigma$ donde $m \in \mathbb{N}$ suficientemente grande para que $u_{sup} > y_{max}$ y σ es la desviación típica muestral de los datos.
- En cuanto a la etapa 2, la longitud de cada clase de la partición coincide con la desviación típica muestral de los datos σ y m corresponde al número de intervalos de la partición con lo que se tiene:

$$u_i = [u_{inf} + (i - 1)\sigma, u_{inf} + i\sigma] \text{ para } i = 1, \dots, m$$

donde los u_i son los intervalos que recubren el universo U .

Se propuso aumentar el número de intervalos en la partición del universo. Nótese que las estructuras temporales de las series de rendimientos tienen un comportamiento y unas propiedades distintas a las series temporales habituales. En los experimentos realizados, el excesivo refinamiento de la partición inicial provocó un empeoramiento notable de las medidas de ajuste, por lo que se desestimo la realización de un sobrepaticionamiento del universo de los datos.

- En la etapa 3, la información requerida para definir la función de pertenencia de cada uno de los números difusos trapezoidales $A_i, i = 1, \dots, m$, se obtiene a partir de los datos observados pertenecientes a los intervalos $\{u_{i-1}, u_i, u_{i+1}\}$ de tal forma que el núcleo del número difuso trapezoidal A_i sea u_i , y sus amplitudes izquierda y derecha dependen de una decisión. Se analizan dos alternativas:

Algoritmo 1: Ambas amplitudes tanto izquierda como derecha coinciden con el punto medio de los intervalos u_i , todos ellos tomados con un mismo tamaño. Así la amplitud del soporte de A_i es dos veces la amplitud de su núcleo. A lo largo de este capítulo a esta propuesta se llamará Algoritmo 1. Por lo tanto se tiene que:

$$A_i = (u_{inf} + (i - 1)\sigma, u_{inf} + i\sigma, \sigma/2, \sigma/2) \text{ con } i = 1, \dots, m$$

Algoritmo 2: Sea $med(u_i)$ la mediana del conjunto de datos pertenecientes al intervalo u_i . Entonces para esta propuesta la amplitud izquierda

cubre la distancia entre $med(u_{i-1})$ y el extremo inferior del intervalo u_i . Análogamente se procede sobre la amplitud derecha, que cubre la distancia desde el extremo superior de u_i hasta $med(u_{i+1})$. A lo largo de este capítulo a esta propuesta se llamará Algoritmo 2. Por lo tanto se tiene que:

$$A_i = (u_{inf} + (i-1)\sigma, u_{inf} + i\sigma, amp_1, amp_2) \text{ con } i = 1, \dots, m$$

donde $amp_1 = u_{inf} + (i-1)\sigma - med(u_{i-1})$ y $amp_2 = med(u_{i+1}) - (u_{inf} + i\sigma)$.

La principal consecuencia de las propuestas descritas anteriormente afectan a la etapa 3 en lo que concierne a la ambigüedad (definida en 1.19) de los conjuntos difusos A_i . Por construcción la ambigüedad está directamente fijada por el tamaño de los intervalos u_i . En el Algoritmo 1 para todo i los conjuntos difusos A_i son números difusos trapezoidales simétricos y por lo tanto tienen la misma ambigüedad; mientras en el Algoritmo 2 dependiendo de los datos observados la ambigüedad puede ser distinta para cada A_i . Nótese que el conjunto difuso A_i puede ser un número difuso trapezoidal truncado, ante la ausencia de observaciones en los intervalos u_{i-1} o u_{i+1} en el caso del Algoritmo 2, para el caso del Algoritmo 1 los únicos números difusos trapezoidales truncados son A_1 y A_m .

En cuanto a la etapa 7 se utiliza una nueva estrategia de ponderación, asignándose pesos a cada una de las relaciones lógicas difusas (RLD), haciendo posible que se considere su importancia. En nuestra aproximación, se considerará la recurrencia (número de veces que una relación lógica difusa aparece en las relaciones lógicas difusas de grupos ya propuesto por Cheng et al. (2008a)) y el orden cronológico (las observaciones recientes de las relaciones lógicas difusas tendrán más peso que las antiguas) en la secuencia original de la serie temporal difusa. Así dada una serie temporal difusa la matriz de pesos, $W = [w_{i,k}] \in M_{m \times m}$, almacenará los pesos cronológicos y los pesos que provienen de las recurrencias de las relaciones difusas como sigue:

$$w_{i,j} = \sum_{t=2}^{N-1} \delta_{i,j}^t \quad \text{donde } \delta_{i,j}^t = \begin{cases} t & \text{si } F(t-1) = A_i \text{ y } F(t) = A_j \\ 0 & \text{en otro caso} \end{cases} \quad (2.15)$$

Posteriormente, para calcular las predicciones difusas, los pesos serán normalizados por filas. Y se construye la predicción como un número difuso trapezoidal, aplicando aritmética difusa (definición 1.16), mediante una

combinación lineal convexa de números difusos trapezoidales.

Una vez comentadas todas las peculiaridades de nuestra propuesta, véase a continuación cómo calcular predicciones difusas aplicando las modificaciones sugeridas anteriormente.

Sean las relaciones lógicas difusas de grupos siguientes:

$$\begin{aligned} A_1 &\rightarrow A_{k_1^1}, \dots, A_{k_{n_1}^1} \\ A_2 &\rightarrow A_{k_1^2}, \dots, A_{k_{n_2}^2} \\ &\dots \\ A_m &\rightarrow A_{k_1^m}, \dots, A_{k_{n_m}^m} \end{aligned}$$

donde m es el número de conjuntos difusos y tal que si $k_r^i = j$ con $r = 1, \dots, n_i$ entonces $\exists A_i \rightarrow A_j$ relación lógica difusa, y donde n_i es el número total de las relaciones difusas existentes entre A_i y el resto de conjuntos difusos.

Sea $F(N) = A_i$ la observación en el instante N , para simplificar la notación y dado que se trata en todo momento del conjunto genérico A_i , se denotarán a los subíndices k_r^i como k_r .

A partir de las relaciones lógicas difusas de grupos (RLDG) y de los pesos calculados se construye una predicción difusa, la cual se llama $F(N+1) = \hat{A}_{N+1}^c$, como un número difuso trapezoidal. La letra 'c' que hay en el superíndice de la predicción difusa \hat{A}_{N+1}^c , hace referencia a que los pesos utilizados en la ponderación tienen origen cronológico, en capítulos posteriores se verán otros tipos de ponderaciones.

Se aplican las siguientes reglas heurísticas:

1. Si $n_i = 1$ e $k_1 = j$ entonces la RLDG de A_i va a A_j , nuestro procedimiento construye un número difuso trapezoidal con todos los valores pertenecientes a A_i : $\hat{A}_{N+1}^c = A_j$
2. Si $n_i = s$ entonces RLDG de A_i va a $A_{k_1}, A_{k_2}, \dots, A_{k_s}$, y utilizando la definición aritmética difusa (definición 1.16) se construye la predicción trapezoidal difusa como sigue:

$$\hat{A}_{N+1}^c = \frac{w_{i,1}}{\sum_{j=1}^m w_{i,j}} A_1 + \dots + \frac{w_{i,m}}{\sum_{j=1}^m w_{i,j}} A_m$$

Los pesos $w_{i,j}$ están definidos en (2.15).
Definiéndose el vector de pesos normalizados como:

$$w_i = \left[\frac{w_{i,1}}{\sum_{j=1}^m w_{i,j}}, \dots, \frac{w_{i,m}}{\sum_{k=1}^m w_{i,k}} \right] \quad (2.16)$$

3. Si $n_i = 0$ entonces la RLDG de A_i es vacía, nuestro procedimiento construye un número difuso trapezoidal con los valores pertenecientes a A_i : $\hat{A}_{N+1}^c = A_i$

□

Para representar los pesos calculados en las relaciones lógicas difusas de grupos se utiliza una matriz $W = [w_{i,j}]$ con $i = 1, \dots, m$ y $j = 1, \dots, m$ de tal forma que cada fila i de la matriz representa al vector w_i calculado en (2.16) y que corresponde a los pesos asociados al correspondiente conjunto difuso A_i en las RLDG.

Nótese que en el caso de Cheng, $w_{i,j}$ representa la frecuencia de aparición de la relación lógica difusa $A_i \rightarrow A_j$ en las relaciones lógicas difusas de grupos. Y a diferencia de Yu que representa una ponderación cronológica sobre cada fila de las relaciones lógicas difusas de grupos de forma independiente; en nuestro procedimiento aunque los pesos también son de naturaleza cronológica, la ponderación cronológica se realiza sobre serie temporal difusa.

Finalmente nuestro procedimiento calcula el punto medio del intervalo posibilístico de medias del número difuso trapezoidal \hat{A}_i^c , tomándolo como predicción puntual. Así se puede comparar nuestra aproximación ponderada difusa con las propuestas clásicas de aproximaciones ponderadas para la predicción: los métodos de Yu (2005b) y Cheng et al. (2008a), los cuales proporcionan predicciones puntuales. Sin embargo, nótese que nuestras predicciones difusas permiten calcular otros momentos posibilísticos de \hat{A}_{N+1}^c .

A continuación véase en que consiste el experimento que se plantea para este capítulo y también un ejemplo práctico de aplicación de nuestras propuestas.

2.4. Experimentos numéricos

Con tal de comprobar la precisión predictiva de los métodos ponderados de predicción de las series temporales difusas, se propone un experimento de simulación. En los experimentos numéricos se trabaja en tantos por cien ya que resulta más fácil y comprensible la presentación de los resultados.

El conjunto de datos se compone por los rendimientos de 153 semanas de $n = 23$ activos financieros que están en el IBEX35, que es el índice de referencia del mercado bursátil español, entre los años 2011 y 2013; tomando como representante de la semana el valor del activo financiero en el momento del cierre del miércoles. Se construyen una serie de carteras ($X = (x_1, \dots, x_n)$) de forma aleatoria, para ello se consideran los rendimientos de los veintitrés activos financieros. Aunque hay treinta y cinco activos en el índice IBEX35, tan sólo se toman veintitrés debido a que, durante el período estudiado, hubo muchos activos que se salieron del índice o entraron nuevos, o bien no se pudieron recopilar todos los datos necesarios.

Se introducen restricciones para definir el conjunto de las carteras factibles: para la obtención de una mayor diversificación de la inversión, se propone una cota superior u_j en la proporción de la inversión (x_j) para cada uno de los activos financieros, siendo $u_j = 0.2$ y se restringe $0 \leq x_j \leq 0.2 \quad \forall x_j \in X$, las proporciones x_j están expresadas en tantos por uno.

Para mejorar el control de las carteras, se introducen restricciones de cardinalidad, denotado como $K = 10$, siendo K el número de activos financieros incluidos en las carteras con una proporción de inversión positiva (tamaño de la cartera).

Las carteras factibles son un subconjunto del simplex del espacio real de dimensión 23 ($\sum_{j=1}^{23} x_j = 1$ y con $x_j \geq 0$), sujeto a las condiciones dadas por las restricciones anteriormente mencionadas y por la restricción presupuestaria.

Todas las carteras del experimento de simulación, $X = (x_1, \dots, x_{23})$ son obtenidas mediante el siguiente procedimiento, con lo que se asegura que son uniformemente seleccionadas dentro del conjunto factible:

1. Para generar una distribución uniforme en el simplex de dimensión K , se aplica el algoritmo propuesto en Bermúdez et al. (2012). Este proceso consiste en simular una distribución Dirichlet K -dimensional

con todos los parámetros iguales a 1, desechando aquellos puntos que no cumplan la restricción de diversificación. El resultado es un vector $p = (p_1, \dots, p_K)$ donde sus componentes pertenecen al intervalo $(0, 0.2]$ y verifican $\sum_{j=1}^K p_j = 1$.

2. Los K activos financieros son seleccionados mediante muestreo aleatorio simple entre los n posibles.
3. Una cartera factible X es obtenida por medio de la asignación de los porcentajes p_j a los K activos seleccionados, y asignando valor de cero al resto de activos.

El rendimiento de un activo j en la semana t se define como sigue:

$$r_{tj} = 100(P_{tj} - P_{(t-1)j})/P_{(t-1)j}$$

para $t = 2, \dots, 153$ y $j = 1, \dots, 23$, donde P_{tj} es el precio de cierre del activo j el miércoles de la semana t . Para toda cartera generada X se evalúa su rendimiento semanal en cada miércoles como sigue:

$$r_t(X) = \sum_{j=1}^{23} r_{tj} x_j$$

Se usarán los primeros 152 rendimientos semanales de una cartera X como conjunto de datos de entrenamiento y se aplicarán diversos métodos de series temporales difusas para la predicción del rendimiento en la semana 153 (predicciones de un paso).

2.4.1. Análisis de una cartera

Para ver cómo funcionan nuestras propuestas se toma como ejemplo la primera simulación del experimento descrito en el apartado anterior y se hace un resumen del funcionamiento del Algoritmo 2 aplicado sobre la cartera #1 que corresponde a dicha primera simulación. Esta cartera es también la utilizada anteriormente como ejemplo para explicar las predicciones puntuales en los algoritmos de series temporales difusas (Chen, 1996; Yu, 2005b; Cheng et al., 2008a), y cuya representación gráfica se puede ver en la figura 2.1.

En la serie temporal tomada como ejemplo, que corresponde a la cartera #1, es notoria la oscilación alrededor del cero del rendimiento (r_t), esto hace que las relaciones entre dos datos consecutivos sean débiles y por lo tanto que influya en el proceso de predicción. Se analiza la precisión predictiva de

nuestras propuestas para las series temporales difusas cuando se aplica a este tipo de series.

Primero se calcula la desviación típica muestral de los datos $\sigma = 4.47$ siendo $rango(r_t) = [-12.67, 12.95]$, a partir de la desviación típica muestral se definirá el universo de los datos como $U = [-14.00, 17.29]$ con $m = 7$ y $\epsilon = 1.33$, de este modo se divide dicho universo en siete intervalos u_i , todos ellos de longitud igual que la desviación típica muestral. La clasificación de los datos de entrenamiento entre esos intervalos permiten definir los números trapezoidales difusos.

Número Difuso	Valor esperado	RLDG	Predicción
A_1	-11.77	A_6, A_4, A_7	9.66
A_2	-7.30	$A_3, A_1, A_5, A_6, A_2, A_4$	-3.42
A_3	-2.83	A_4, A_5, A_3, A_2, A_6	0.51
A_4	1.64	A_5, A_3, A_4, A_2, A_1	0.30
A_5	6.10	A_6, A_3, A_4, A_2, A_5	1.64
A_6	10.57	A_4, A_2, A_3	-2.02
A_7	15.04	A_5	6.04

Cuadro 2.2: Predicciones puntuales de nuestro método para las series temporales difusas: Algoritmo 2.

En el cuadro 2.2 se presentan los resultados obtenidos cuando se usa la mediana (Algoritmo 2) de cada uno de los subconjuntos de los datos que están en los intervalos u_i para la construcción de los números difusos trapezoidales $\{A_1, \dots, A_7\}$.

Las dos primeras columnas del cuadro 2.2 muestran los conjuntos difusos A_i y los valores esperados de los conjuntos trapezoidales difusos no simétricos. En la tercera columna vienen recogidas las relaciones lógicas difusas de grupos observadas en la serie temporal difusa de la cartera #1, y la última columna se muestra la predicción puntual de una etapa.

Para calcular las predicciones difusas \hat{A}_t^c , y las predicciones puntuales (mostradas en la última columna del cuadro 2.2), los métodos propuestos usan matrices de pesos $W \in M_{7 \times 7}$, en las cuales vienen determinadas las frecuencias cronológicas de las relaciones lógicas difusas sobre el conjunto de los datos de entrenamiento. El cuadro 2.3 muestra esos pesos $w_{i,k}$ calculados utilizando la ecuación (2.15), los cuales son posteriormente normalizados por

filas, tal y como se define en las reglas heurísticas de predicción (punto 2 en la sección 2.3.1. de este capítulo).

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	0	0	0	49	0	31	81
A_2	110	338	239	170	33	46	0
A_3	0	182	940	1668	373	39	0
A_4	48	256	1769	2281	792	0	0
A_5	0	112	320	867	528	1	0
A_6	0	32	40	49	0	0	0
A_7	0	0	0	0	82	0	0

Cuadro 2.3: Representación matricial de los pesos W de las relaciones lógicas difusas de grupos correspondiente a la cartera #1.

El último valor observado en el conjunto de los datos de entrenamiento correspondientes a la cartera #1 es $r_{152} = -2.24$, que pertenece al tercer intervalo, donde el número difuso trapezoidal asociado es A_3 .

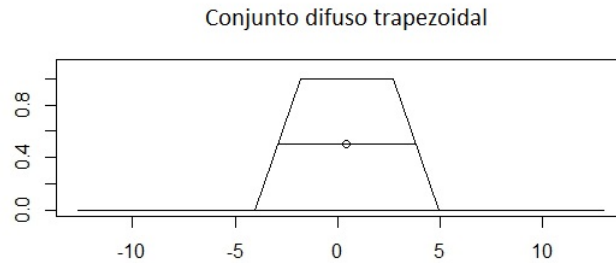


Figura 2.2: Función de pertenencia de la predicción difusa \hat{A}_{153} para la serie de rendimientos de la cartera #1.

La predicción difusa es $F(153) = \hat{A}_{153}^c$. La figura 2.2 muestra la función de pertenencia de la predicción difusa obtenida para la serie de rendimientos de la cartera #1, que corresponde al número difuso trapezoidal $\hat{A}_{153}^c = (-1.79, 2.68, 1.68, 1.91)$, con intervalo de medias $I_m(\hat{A}_{153}^c) = [-2.63, 3.64]$ (ecuación 2.5), y valor esperado 0.51 como predicción puntual que aparece en el cuadro 2.2.

Nótese que \hat{A}_{153}^c se obtiene como una agregación ponderada de números difusos trapezoidales que representan los cinco valores difusos que componen

la parte derecha de la correspondiente relación lógica difusa de grupos, aplicando aritmética difusa (definición 1.16).

Esta predicción puntual es coincidente con el valor esperado del intervalo posibilístico de medias de \hat{A}_{153} , el punto medio del α -corte para $\alpha = 0.5$, el cual está representado por la figura 2.2 mediante un punto circular. Finalmente el valor observado es $r_{153} = -0.87$ y el error de predicción de un paso es 1.37, siendo la ambigüedad de la predicción difusa de 3.13.

Análogamente, se aplicará nuestro modelo de predicción pero utilizando números trapezoidales simétricos, Algoritmo 1, así la predicción puntual es 0.47, con un error de predicción de 1.31 y la ambigüedad es de 3.35.

Son claras las diferencias numéricas con respecto a los pesos que se muestran en el cuadro 2.3 y los vectores de pesos asociados de los modelos de Yu y Cheng. Sin embargo, la diferencia fundamental entre los métodos de series temporales difusas ponderadas expuestos anteriormente y nuestra propuesta radica en el hecho de que se puede proporcionar un valor aproximado para la ambigüedad de la predicción difusa del rendimiento de la cartera, a través del tamaño del intervalo posibilístico de medias, no sólo generar predicciones puntuales. Así se puede aproximar el riesgo de la cartera mediante la ambigüedad, ya que es una medida relacionada con la variabilidad de la predicción.

2.4.2. Análisis comparativo

Para comparar la calidad de los métodos de predicción de series temporales difusas ponderadas, se genera un conjunto de $N = 2000$ carteras aleatorias, obteniendo la serie temporal de sus rendimientos. Luego, utilizando los métodos de predicción de series temporales difusas ponderadas descritas en la sección anterior, se obtienen las predicciones puntuales de los rendimientos de una etapa.

Para cada método $M_j = \{Yu, Cheng, Algoritmo 1, Algoritmo 2\}$ y cada cartera X_i , se calculan en primer lugar los errores de ajuste. El ajuste se calcula de forma análoga a una predicción, para ello utilizando el conjunto de datos de entrenamiento se observa el valor anterior (cronológico) al que se quiere ajustar y se observa a qué conjunto difuso pertenece; finalmente se aplican las reglas de predicción descritas en la sección 2.3.1, y se calculan los errores:

$$e_{i,M_j}(t) = \hat{r}_{M_j(t,i)} - r_{(t,i)}$$

siendo $\hat{r}_{M_j(t,i)}$ el ajuste y $r_{(t,i)}$ el rendimiento real observado de esa cartera para $t = 2, \dots, 152$.

Para comparar la calidad de las series ajustadas, se calculan tres de las medidas de bondad de ajuste más comunes para cada cartera:

- la raíz cuadrada del error cuadrático medio,

$$\text{RMSE}_{i,M_j} = \sqrt{\sum_{t=1}^{152} e_{i,M_j}^2(t)/152}$$
;
- la desviación media absoluta,

$$\text{MAD}_{i,M_j} = \sum_{t=1}^{152} |e_{i,M_j}(t)|/152$$
;
- el sesgo,

$$\text{BIAS}_{i,M_j} = \sum_{t=1}^N e_{i,M_j}(t)/152$$
;

La tercera medida es una estimación del sesgo del método; lo ideal es que debería ser cero, mientras que los otros miden el error cuadrático medio y la desviación media absoluta, respectivamente.

Finalmente, se calcula la media aritmética de estas tres medidas de error para las 2000 carteras, tal y como se muestra en el cuadro 2.4 donde vienen recogidos los errores medios de ajuste de cada modelo.

Métodos	RMSE	MAD	BIAS
Yu	3.65	2.78	-0.17
Cheng	3.69	2.83	-0.002
Algoritmo 1	3.76	2.87	-0.21
Algoritmo 2	4.16	2.12	-0.09

Cuadro 2.4: Errores de ajuste obtenidos en el experimento de simulación para los métodos de STD ponderadas.

Obsérvese que en el cuadro 2.4 el método con un menor RMSE es el de Yu (2005b), mientras que es el Algoritmo 2 quien obtiene un menor MAD, en cuanto al sesgo corresponde al método de Cheng et al. (2008a) el que más se aproxima al valor cero.

Sin embargo, es más interesante conocer la calidad de la predicción de una etapa de estos procedimientos. Para cada cartera X_i y cada método M_j , se calcula el error de predicción de una etapa:

$$e_{i,M_j} = \hat{r}_{M_j(153,i)} - r_{(153,i)}$$

siendo $\hat{r}_{M_j(153,i)}$ la predicción y $r_{(153,i)}$ el rendimiento observado de esa cartera.

Para hacer la comparación de los resultados de predicción del experimento de simulación, se mide la precisión de la predicción usando el RMSE, MAD y el sesgo de los errores observados:

- la raíz cuadrada del error cuadrático medio,

$$\text{RMSE}_{M_j} = \sum_{i=1}^{2000} \sqrt{e_{i,M_j}^2} / 2000$$

- la desviación media absoluta,

$$\text{MAD}_{M_j} = \sum_{i=1}^{2000} |e_{i,M_j}| / 2000;$$

- Sesgo,

$$\text{BIAS}_{M_j} = \sum_{i=1}^{2000} e_{i,M_j} / 2000;$$

Nótese que sólo se realiza una predicción de cada cartera y que por lo tanto el RMSE coincide con el MAD, por ello no se utiliza el RMSE en este experimento numérico, aunque sí se utiliza en capítulos posteriores.

Para comparar el rendimiento de los métodos STD anteriores con otros modelos no difusos, se utiliza el software de predicción automático incluido en el paquete `forecast` para R (Hyndman y Khandakar, 2008) que está disponible en CRAN (<http://cran.r-project.org/>), se utilizan las opciones de *suavizado exponencial simple* y *ARIMA*. Se decide aplicar el suavizado exponencial simple (SES) comando `'ets'` restringido al modelo aditivo ('ANN'), para todas las 2000 series temporales de rendimientos de carteras, mientras para los modelos *ARIMA* el paquete `forecast` selecciona automáticamente el mejor, utilizando el criterio AIC y el comando `'auto.arima'` (véase el manual Hyndman et al. (2015)).

También se implementa un conjunto de modelos de media móvil, pero el mejor rendimiento lo proporciona el MM1 que toma la última observación de la serie como predicción a una etapa, cuyas predicciones también se incluyen en las comparaciones de las predicciones a una etapa. El cuadro 2.5 también muestra los errores de predicción promediados usando estos enfoques no difusos.

Los resultados numéricos obtenidos vienen recogidos en el cuadro 2.5. Tal y como se muestra en dicho cuadro, los errores de predicción de una etapa MAD son menores que los errores de ajuste asociados, mientras que los

errores del sesgo son mayores que el sesgo para el ajuste. Los resultados obtenidos por nuestras propuestas son mejores que los proporcionados por los otros métodos ponderados de series temporales difusas (STD).

Métodos	MAD	BIAS
Yu	1.034	-0.509
Cheng	1.114	-0.490
Algoritmo 1	0.985	-0.687
Algoritmo 2	0.982	-0.694
MM1	1.067	0.766
SES	1.158	1.015
ARIMA	1.132	0.989

Cuadro 2.5: Precisión de las predicciones a una etapa

El diagrama de cajas de la figura 2.3 ilustra las diferencias entre los errores de predicción de una etapa del rendimiento logrados por cada método de predicción. Obsérvese que los errores de predicción proporcionados por nuestras propuestas son bastante simétricos y menos dispersos que las de los métodos de Yu (2005b); Cheng et al. (2008a) y MM1. Sin embargo, se comportan un poco más dispersos pero con menos sesgo que los errores de predicción proporcionados por el *SES* y el *ARIMA*.

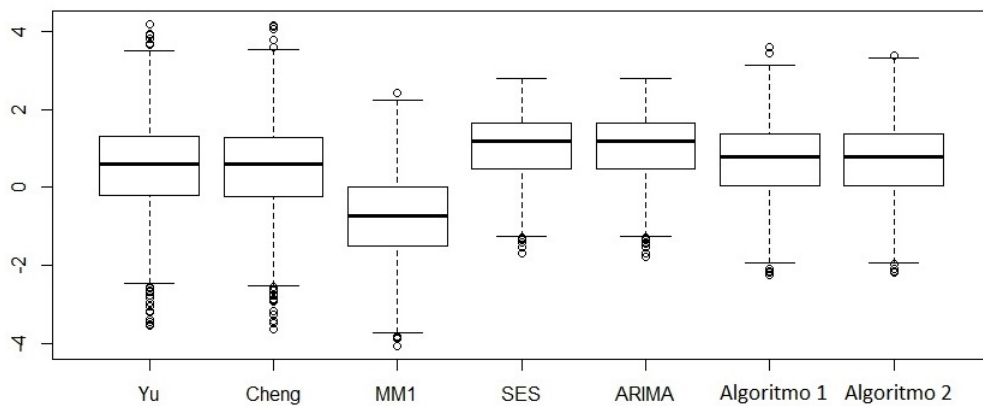


Figura 2.3: Diagrama de cajas de los errores de predicción de una etapa obtenidos en el experimento de simulación de 2000 carteras.

A continuación, se realiza un análisis estadístico para comprobar si las diferencias observadas, en el análisis descriptivo anterior, son estadísticamente

significativas. Para ello, se realiza una comparación estadística por pares de los errores de predicción y el sesgo a través del test t, ajustando los p-valores utilizando el método Holm (para tener en cuenta el problema de comparación múltiple). Los p-valores ajustados están presentados en los cuadros 2.6 y 2.7, respectivamente.

MAD	Yu	Cheng	<i>MM1</i>	<i>SES</i>	<i>ARIMA</i>	Algoritmo 1
Cheng	1.0e-06					
<i>MM1</i>	7.4e-01	3.4e-01				
<i>SES</i>	3.2e-12	1.3e-01	6.9e-03			
<i>ARIMA</i>	8.4e-08	7.4e-01	1.1e-01	2.4e-22		
Algoritmo 1	7.3e-03	1.9e-15	1.1e-02	2.8e-48	2.8e-34	
Algoritmo 2	3.7e-03	2.8e-16	9.0e-03	8.3e-55	4.6e-39	4.5e-01

Cuadro 2.6: P-valores ajustados para comparaciones dos a dos de los errores de predicción puntual (MAD).

La primera columna del cuadro 2.6 muestra los p-valores ajustados de la comparación por pares del MAD de Yu y los otros métodos de predicción, y así sucesivamente para las otras columnas.

Estas pruebas estadísticas muestran que existen diferencias estadísticamente significativas entre la media del MAD alcanzado por nuestras propuestas y los otros métodos de predicción (p-valores ajustados siempre menores a 0.011), a favor de los nuevos métodos STD ponderados que incorporan nuestras propuestas 1 y 2; sin embargo, no hay diferencia estadística entre la media de los errores MAD de la Algoritmo 1 y Algoritmo 2 (valor de p ajustado = 0,45).

BIAS	Yu	Cheng	<i>MM1</i>	<i>SES</i>	<i>ARIMA</i>	Algoritmo 1
Yu						
Cheng	5.4e-01					
<i>MM1</i>	5.7e-297	1.2e-249				
<i>SES</i>	5.0e-161	1.2e-159	0.0e+00			
<i>ARIMA</i>	4.3e-142	1.4e-140	0.0e+00	7.2e-23		
Algoritmo 1	4.0e-27	1.2e-27	0.0e+00	4.0e-152	1.1e-121	
Algoritmo 2	9.0e-30	4.6e-30	0.0e+00	3.1e-163	1.8e-129	6.4e-04

Cuadro 2.7: P-valores ajustados de la comparación por pares del sesgo de los errores de predicción puntual de una etapa.

Además, también se analizan los resultados obtenidos para el sesgo, que se muestran en el cuadro 2.7. Cada comparación por parejas es estadísticamente significativa, excepto la alcanzada por los métodos STD de Yu (2005b) y Cheng et al. (2008a) (p-valor ajustado = 0,54).

2.5. Selección de carteras

En este apartado se toman como medidas predictivas difusas, en el problema de selección de carteras bi-objetivo, el rendimiento esperado (como el punto medio del intervalo posibilístico de medias) y el riesgo lateral (como la semi-amplitud del intervalo posibilístico de medias).

Para aproximar las soluciones a dicho problema se utiliza un algoritmo genético descrito en Bermúdez et al. (2012), introduciendo las medidas predictivas difusas como parámetros de referencia del problema de selección de carteras.

Al igual que en el experimento numérico de este capítulo, el conjunto de datos se compone por los rendimientos de 153 semanas de $n = 23$ activos financieros que están en el IBEX35, que es el índice de referencia del mercado bursátil español, entre los años 2011 y 2013; tomando como representante de la semana el valor del activo financiero en el momento del cierre del miércoles.

Para ver el comportamiento de los rendimientos para distintos niveles de riesgo se decide utilizar un algoritmo genético para problemas de optimización bi-objetivo que permite aproximar la frontera de Pareto. Partiendo de las funciones de rendimiento y riesgo de la formulación presentada en (2.10) del apartado 2.1.1, planteamos un problema bi-objetivo. Sea $D_X = (a, b, \gamma, \beta)$ la predicción difusa (número difuso trapezoidal) de la serie temporal de rendimientos de una cierta cartera X . El núcleo y las amplitudes (parámetros a, b, γ, β) de la predicción difusa D_X vienen determinados por la partición inicial y por el Algoritmo 1, respectivamente.

Se introducen restricciones para definir el conjunto de las carteras factibles: para la obtención de una mayor diversificación de la inversión, se propone una cota inferior $l_i = 0$ y cota superior $u_i = 0.3 \quad \forall i = 1, \dots, 23$ para cada uno de los activos financieros y se restringe $0 \leq x_i \leq 0.3 \quad \forall x_i \in X$. También se introducen restricciones de cardinalidad, la cartera está compuesta entre 6 y 9 activos ($K_1 = 6$ y $K_2 = 9$).

Por lo que el problema de optimización quedaría como sigue:

$$\begin{aligned}
 & \text{Max}_x \quad \frac{1}{2}(a + b) + \frac{1}{4}(\beta - \gamma) \\
 & \text{Min}_x \quad b - a + \frac{1}{2}(\gamma + \beta) \\
 & \text{s.a.} \\
 & \sum_{i=1}^{23} x_i = 1 \\
 & 0 \leq x_i \leq 0.3y_i \quad \forall i = 1, \dots, 23 \\
 & y_i \in \{0, 1\} \\
 & 6 \leq \sum_{i=1}^{23} y_i \leq 9
 \end{aligned} \tag{2.17}$$

La población inicial de la que parte el algoritmo genético está generada aleatoriamente, tal y como viene descrito en el apartado 2.4. Para tener la suficiente variedad de carteras se realizan cincuenta ejecuciones del algoritmo genético y así poder explorar la frontera de Pareto con suficiente detalle. Se proponen los siguientes parámetros a utilizar en el algoritmo genético:

- Población inicial: 200 carteras.
- Proporción de carteras seleccionadas como élite: 20 %.
- Número de generaciones: 25.

En la figura 2.4 aparecen representados el rendimiento esperado y el riesgo lateral de las cincuenta ejecuciones del algoritmo genético, a saber, en cada ejecución se generan aleatoriamente doscientas carteras, también se representa la última población de cada una de las ejecuciones (coloreado de puntos azules) y finalmente se aproxima la envoltura convexa mediante una iteración más a partir de la familia de todas las carteras que conforman las envolturas convexas de cada ejecución (coloreado con líneas y puntos verdes).

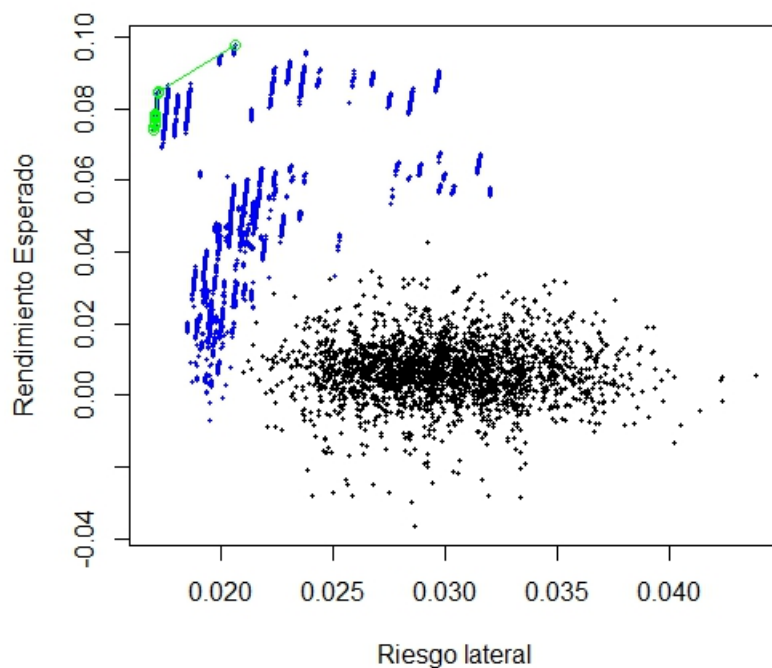


Figura 2.4: Gráfica del rendimiento esperado y riesgo lateral de carteras generadas aleatoriamente, la última población obtenida por la aplicación del algoritmo genético y la envoltura convexa.

2.6. Conclusiones

En este capítulo se ha introducido el problema de selección de carteras, se han dado diferentes medidas del rendimiento esperado y del riesgo de una cartera, siendo una de esas medidas una aproximación difusa desde un punto de vista posibilístico.

Siguiendo uno de los objetivos de esta tesis, se ha dado una aproximación alternativa mediante medidas predictivas difusas, desde un punto de vista posibilístico. Para ello se han presentado las series temporales difusas, siguiendo el esquema clásico de series temporales difusas ponderadas propuesta por Yu (2005b) e introduciendo algunas modificaciones, que han dado como resultado dos algoritmos (Algoritmo 1 y Algoritmo 2), diferenciándose en el cálculo del parámetro amplitud de los números difusos trapezoidales utilizados en la STD. Entre dichas modificaciones se encuentra una nueva forma de partir el

universo, de ponderar y nuevas reglas heurísticas para la predicción difusa.

En este capítulo se ha aproximado la incertidumbre del rendimiento futuro de una cartera dada, mediante un número difuso trapezoidal; la utilización de esta clase de conjunto difuso es debido a que la interpretación de un número difuso trapezoidal, para modelizar una cartera de valores, es más sencilla. Para ejemplificar el funcionamiento de las series temporales difusas, se han estudiado distintos métodos de predicción de series temporales difusas clásicas (Chen, 1996; Yu, 2005b; Cheng et al., 2008a), aplicándolos a una cartera en concreto.

Aunque no es un objetivo principal sí que lo considerábamos un objetivo secundario de la tesis y por ello se ha querido comprobar la calidad de la predicción. Se han realizado predicciones puntuales tomando el rendimiento esperado de la predicción difusa (el punto medio del intervalo posibilístico de medias). Para la comparación de los algoritmos propuestos con los métodos de series temporales difusas y con otros métodos de predicción convencionales, se han generado, de forma aleatoria, dos mil carteras donde cada cartera está compuesta por los rendimientos de diez activos de entre veintitrés posibles activos extraídos del IBEX35.

Se ha observado la mejora en media del error de predicción MAD, y pese a tener un sesgo ligeramente superior en media al de los métodos de series temporales difusas, se ha concluido que hay una mejor calidad de la predicción en los algoritmos propuestos, encontrando diferencias estadísticamente significativas.

Se ha finalizado el capítulo utilizando las medidas predictivas difusas en el problema de selección de carteras, donde se ha aplicado un algoritmo genético para aproximar la frontera de Pareto. No se han comparado los resultados de aplicar las medidas predictivas difusas, al problema de selección de carteras, con otras aproximaciones de dicho problema. Sin embargo, los resultados obtenidos en este apartado han sido, a tenor de la falta de comparaciones, satisfactorios.

Dada la mejora en la calidad de la predicción y siendo una primera aproximación de la predicción difusa, en los siguientes capítulos se siguen investigando las posibles mejoras en cuanto a la calidad de la predicción puntual y a las predicciones difusas, estudiando distintos métodos de ponderación teniendo en cuenta la tendencia de la serie temporal e introduciendo la minería de datos en el proceso de predicción.

Series temporales con tendencia

En este capítulo se propone utilizar un nuevo modelo de ponderación basado en lógica difusa para series temporales y hacer predicciones del rendimiento futuro. En el capítulo anterior ya se estudió cómo basándose en la secuencia cronológica de las relaciones lógicas difusas originales se puede definir una ponderación de orden cronológico, ahora, se introduce una nueva ponderación de orden de tendencia y se incorporan a nuestras propuestas para la predicción difusa de una etapa dando un nuevo enfoque del modelado clásico de series temporales difusas y representando un cambio con respecto al capítulo anterior.

Estas modificaciones para la asignación y cálculo de ponderaciones afectan al proceso de predicción. Se utiliza como indicador técnico para predecir la tendencia lo que se denomina como tamaño de salto (se definirá más adelante). Entonces lo que se propone es un nuevo algoritmo donde la predicción difusa se calcula como la combinación lineal convexa de las ponderaciones cronológicas (vistas en el capítulo 2) y las ponderaciones que provienen de los tamaños de saltos. También se propone un procedimiento para estimar el valor del parámetro de dicha combinación lineal, que se fundamenta en el error de ajuste.

Ya se ha estudiado, en el capítulo anterior, que trabajar con números difusos trapezoidales permite analizar tanto el valor esperado como la ambigüedad del comportamiento futuro de una cartera determinada, a partir del intervalo posibilístico de medias. Por lo tanto en este capítulo se sigue trabajando con números difusos trapezoidales y tomando sus distintos mo-

Parte del contenido de este capítulo viene recogido en el artículo Rubio et al. (2017a) de la revista *Expert Systems With Applications (ESWA)*.

mentos posibilísticos para contrastar la calidad de las predicciones. Las series de rendimientos (ya sean de activos o de carteras) no suelen tener tendencia, es por ello que ahora se estudian series con tendencia, por lo que se utilizan índices bursátiles de referencia en los experimentos numéricos.

Se analiza la efectividad de nuestro nuevo enfoque con respecto a otros métodos ponderados de series temporales difusas (STD) y con métodos estadísticos de predicción clásicos, utilizando bases de datos del índice de Taiwán (TAIEX), el índice japonés (NIKKEI), el índice bursátil alemán (DAX30) y el índice bursátil español (IBEX35).

Este capítulo se estructura como sigue: primero se ven los principales cambios y definiciones en la algoritmia ya descrita en el capítulo anterior, y en las reglas heurísticas de predicción, continuando con un ejemplo de aplicación del nuevo algoritmo. En una segunda parte se analizan los experimentos numéricos y se finaliza el capítulo con las conclusiones.

3.1. Métodos de ponderación de series temporales con tendencia

Se proponen varias modificaciones de los enfoques de STD estándar siguiendo el mismo esquema que en el capítulo 2, a este nuevo procedimiento se denota como Algoritmo 3.

Ahora, en nuestra propuesta la longitud de los intervalos l es una función que se define a partir de la desviación típica muestral σ de los datos históricos observados. Y a partir de la longitud l se puede definir el universo $U = [\underline{u}, \bar{u}]$ donde $\underline{u} = y_{min} - d_1$ con $d_1 \in \mathbb{R}^+$ y $\bar{u} = y_{max} + d_2$ con $d_2 \in \mathbb{R}^+$ | $\bar{u} = \underline{u} + lm$ para $m \in \mathbb{N}$ suficientemente grande. Así m corresponderá al número de intervalos que recubren el universo U .

También al igual que en el capítulo anterior, se define la función de pertenencia de cada variable A_i como un número difuso trapezoidal. Resumiendo, hay m intervalos de longitud l y m números difusos trapezoidales A_i , cuyo núcleo coincide con u_i y con soporte $[\underline{u}_i - \frac{l}{2}, \bar{u}_i + \frac{l}{2}]$. Finalmente, se aplican diferentes reglas para construir las predicciones difusas en la etapa 7.

Ahora se va a presentar un nuevo indicador para medir la relación entre

dos datos consecutivos de la serie temporal difusa.

Definición 3.1.

Sea $F(t - 1) = A_i$ y $F(t) = A_j$, para $i, j = 1, 2, \dots, m$, por lo tanto se tiene la relación lógica difusa $A_i \rightarrow A_j$.

Se define $J(t) = j - i$ como el *tamaño de salto* observado asociado a dicha relación lógica difusa en el instante t .

□

A partir de la definición anterior el tamaño del salto estaría entre $-m + 1 \leq J(t) \leq m - 1$. La notación de "J" viene del inglés "jump".

Una vez que se evalúa el tamaño de salto asociado a cada relación lógica difusa (que proporciona un movimiento relativo positivo, negativo o nulo), se puede utilizar esta información para medir la tendencia entre conjuntos difusos consecutivos. Véase a continuación las principales modificaciones del proceso de ponderación.

3.1.1. Nueva propuesta de ponderación

Sean los datos observados y_1, \dots, y_N , donde y_t es la cotización de un índice de referencia bursátil en el instante t , con $t = 1, \dots, N$ y sea el universo $U = \{u_1, u_2, \dots, u_m\}$. Sin pérdida de generalidad tómesese $y_{t_0} \in u_i$ y $y_{t_0+1} \in u_j$, entonces $F(t_0) = A_i$ y $F(t_0 + 1) = A_j$, por lo tanto se tiene que $A_i \rightarrow A_j$ es la relación lógica difusa asociada, la cual tendrá un peso cronológico asociado $t_{i,j} = t_0$ (véase la definición en el capítulo anterior, ecuación 2.15).

Por un lado se tienen los pesos que provienen del modelo cronológico introducido en el capítulo anterior, pesos asignados $w_{i,j}$ (ecuación 2.15) a los conjuntos difusos de la parte derecha de RLDG que incorporan el tiempo en que se observan las relaciones, siendo la representación matricial de la ponderación cronológica de la forma $W = [w_{i,j}] \in \mathbb{R}^{m \times m}$ y donde el proceso de normalización de la matriz se realiza por filas.

Ahora se introduce una nueva ponderación para tener en cuenta la tendencia en la serie temporal difusa, de la siguiente forma: sea $J(t_0 + 1) = k = j - i$ el tamaño de salto observado entre los conjuntos difusos temporalmente consecutivos A_i y A_j , donde $-m + 1 \leq k \leq m - 1$. Entonces, se puede calcular un vector de pesos S , que mide la frecuencia de cada tamaño de salto en la serie temporal difusa observada.

Sea s_k el número de veces que se observa un salto de magnitud k desde $t = 2, \dots, N$.

$$s_k = \sum_{t=2}^N \delta_k^t \quad \text{donde} \quad \delta_k^t = \begin{cases} 1 & \text{si } J(t) = k \\ 0 & \text{en otro caso} \end{cases} \quad (3.1)$$

Los pesos normalizados, se denotan como s'_k , calculándose como sigue:

$$s'_k = \frac{s_k}{\sum_{h=-m+1}^{m-1} s_h}, \quad -m+1 \leq k \leq m-1 \quad (3.2)$$

siendo S la representación vectorial de dichos pesos

$$S = (s'_{-m+1}, \dots, s'_{-1}, s'_0, s'_1, \dots, s'_{m-1})$$

Para hacer la predicción difusa se pueden utilizar alternativamente estos dos tipos de pesos, W y S , aunque en nuestro procedimiento se propone utilizar una combinación lineal de ellos. A continuación se proponen algunas modificaciones a las reglas heurísticas introducidas en el capítulo anterior para este nuevo procedimiento.

3.1.2. Reglas heurísticas para la predicción difusa

Sean $F(N) = A_i$ el último conjunto difuso observado, el vector de pesos S (vector de ponderación asociado a los tamaños de saltos) se usa para construir el número difuso trapezoidal \hat{A}_{N+1}^s , de la siguiente manera:

$$\hat{A}_{N+1}^s = s'_{i+a} A_{i+a} \oplus s'_{i+(a+1)} A_{i+(a+1)} \oplus \dots \oplus s'_{i+b} A_{i+b} \quad (3.3)$$

con $a = \max\{-m+1, 1-i\}$, $b = \min\{m-1, m-i\}$.

Y la predicción difusa $F(N+1) = \hat{A}_{N+1}$ se calcula usando las siguientes reglas heurísticas:

1. Si el RLDG de A_i va a A_k , esto significa que sólo se observa una RLD desde A_i , entonces $\hat{A}_{N+1}^c = A_k$. La predicción difusa resultante sería:

$$\hat{A}_{N+1} = \gamma \hat{A}_{N+1}^s \oplus (1-\gamma) A_k$$

donde $\gamma \in (0, 1)$ y \hat{A}_{N+1}^s viene determinada en la ecuación 3.3.

2. Si la RLDG de A_i va a $A_{k_1}, A_{k_2}, \dots, A_{k_r}$. En las RLDG asociadas a A_i se observan en la parte derecha $A_{k_1}, A_{k_2}, \dots, A_{k_r}$; entonces la predicción difusa sería un número difuso trapezoidal calculado como resultado de una combinación ponderada adecuada de $A_{k_1}, A_{k_2}, \dots, A_{k_r}$ y la tendencia ponderada alrededor de A_i , de la siguiente manera:

$$\hat{A}_{N+1} = \gamma \hat{A}_{N+1}^s \oplus (1 - \gamma) \hat{A}_{N+1}^c$$

donde \hat{A}_{N+1}^c es la predicción cronológica difusa que viene definida en las reglas heurísticas del capítulo anterior sección 2.3.1, \hat{A}_{N+1}^s se calcula como en la ecuación 3.3 y $\gamma \in (0, 1)$.

3. Si la RLDG de A_i es vacía, i.e., no se observan RLD desde A_i a otro conjunto difuso. Entonces se usa el vector de tendencia ponderada S para construir el número difuso trapezoidal \hat{A}_{N+1}^s , y por lo tanto la predicción difusa sería $\hat{A}_{N+1} = \hat{A}_{N+1}^s$.

Nótese que para los casos donde $i + a < 1$ o $i + b > m$, todos los conjuntos difusos que no estuvieren previamente definidos, se construirían según establece el procedimiento y se procede como sigue:

Sea $k \in \mathbb{Z}$ tal que $i + a \leq k < 1$, entonces se define el número difuso trapezoidal

$$A_k = (u_{inf} + (k - 1)l, u_{inf} + kl, l/2, l/2)$$

análogamente se construyen para $m < k \leq i + b$. De esta forma se permiten generar predicciones fuera del universo definido.

En este capítulo al nuevo procedimiento descrito se llama Algoritmo 3 y para facilitar la explicación de su funcionamiento se definen los conjuntos difusos cómo números trapezoidales simétricos ¹, de la misma forma que viene recogido en el Algoritmo 1 del capítulo anterior.

Véase a continuación como se determina el parámetro γ de la combinación lineal convexa descrita en las reglas heurísticas.

¹Nótese que si se usaran números difusos trapezoidales no simétricos para representar A_i , $i = 1, \dots, m$ (Algoritmo 2), o siguiendo otros enfoques sugeridos en Wang et al. (2013, 2014) donde se usan técnicas de *clustering* para partir el universo lo que devenga en una asimetría de los conjuntos difusos; se verían afectados, por extensión, el núcleo y las amplitudes de la predicción difusa \hat{A}_{N+1} .

3.1.3. Aproximación del parámetro γ

El parámetro γ mide la influencia de la tendencia sobre la serie temporal. El valor "óptimo" para γ se fundamenta en el error de ajuste, por lo tanto se analizan los resultados del Algoritmo 3 aplicado sobre el conjunto de datos de entrenamiento de la serie temporal, siguiendo las reglas heurísticas siguientes:

1. Se aplica repetidamente el Algoritmo 3 para diferentes valores de $\gamma = \{0, 0.1, 0.2, \dots, 1\}$ y se calcula el RMSE promediado de los errores de ajuste para cada γ ;
2. Se selecciona el valor de γ asociado al RMSE mínimo y es este valor el que se usa para las predicciones de una etapa.

En resumen, se ha calculado un número trapezoidal \hat{A}_{N+1} como predicción difusa, y así tanto el valor esperado como la ambigüedad o cualquier otra característica de interés pueden obtenerse a partir de dicha predicción difusa. Por construcción, la predicción difusa \hat{A}_{N+1} conserva las características de los números difusos trapezoidales originales (amplitudes centrales y amplitudes), ya que se ha obtenido como una combinación lineal convexa de la tendencia ponderada \hat{A}_{N+1}^s , y de la ponderación cronológica \hat{A}_{N+1}^c . También se ha establecido un procedimiento para aproximar el parámetro γ de la combinación lineal. Véase a continuación un ejemplo de aplicación del Algoritmo 3 descrito en el diagrama de flujo (figura 3.1).

3.1.4. Predicción de una etapa: un ejemplo

Ahora se analiza cómo funciona el procedimiento descrito del Algoritmo 3, analizando un ejemplo que consta de quinientas once cotizaciones diarias del índice del IBEX35 entre los años 2013 y 2015. En la figura 3.2 se muestran las cotizaciones observadas del IBEX35, la cual está dividida en dos partes por un lado están los datos de entrenamiento y por otro los datos de contraste.

Para estimar la longitud del intervalo se propone $l = \frac{\sigma}{10}$, esta función se aproxima a través múltiples de experimentos; se tiene claro que la desviación típica muestral debería estar relacionada con el tamaño del intervalo, sólo faltaba ajustar dicho parámetro para que diera la máxima información posible (información granulada). En nuestra experiencia dicho tamaño es el que mejor resultado proporciona en los distintos experimentos numéricos realizados para datos de índices bursátiles.

En este ejemplo, $y_{min} = 7553$ y $d_1 = 953$ luego $u_{inf} = 6600$ también se tiene que $y_{max} = 11188$ y $\sigma = 1029.4$ luego la longitud de los intervalos es

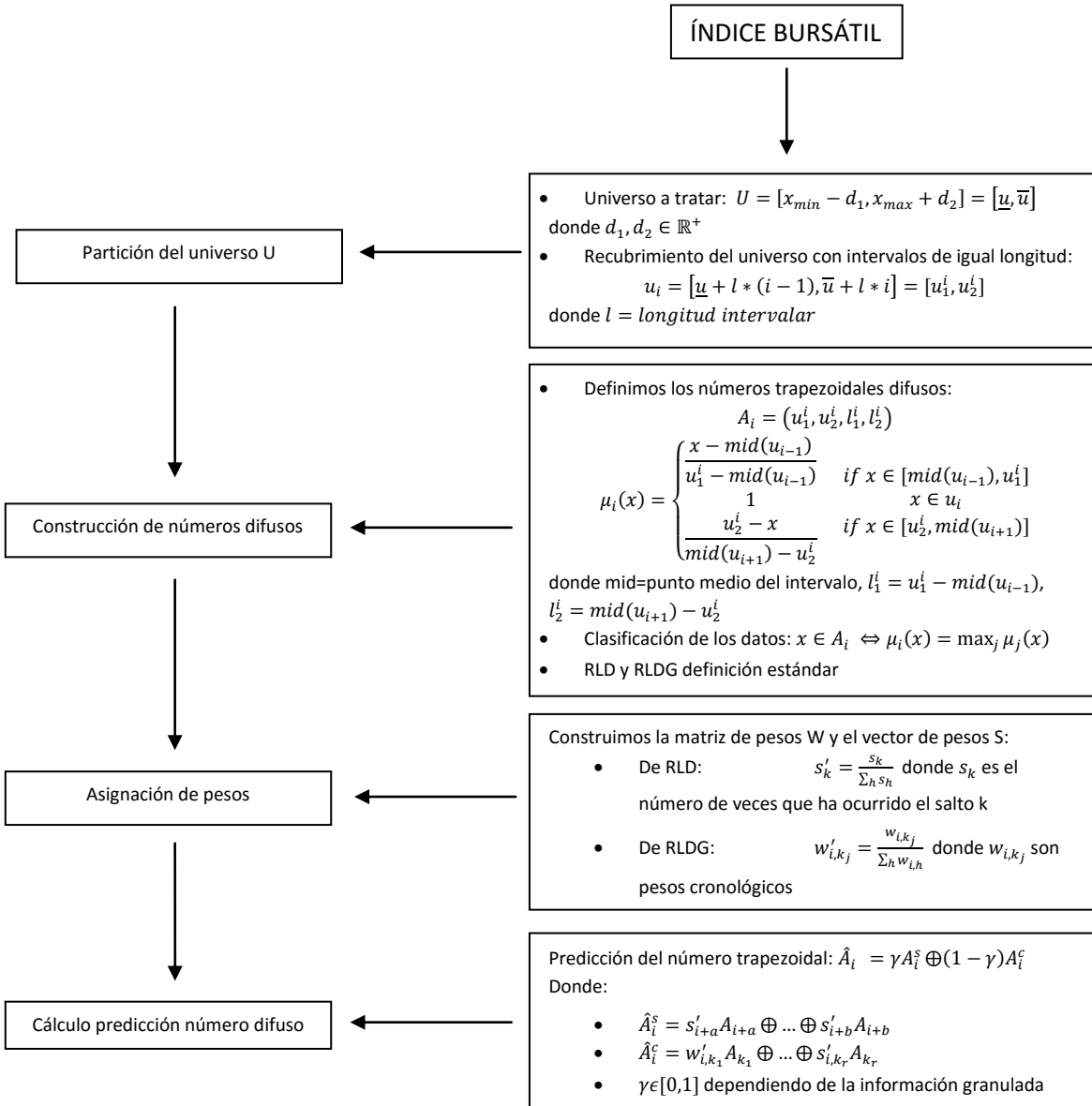


Figura 3.1: Esquema de flujo del Algoritmo 3

$l = 103$, así pues sea $m = 52$ número de intervalos y $u_{sup} = 11956$, por lo tanto el universo U es aproximado por $U = [6600, 11956]$. Al igual que

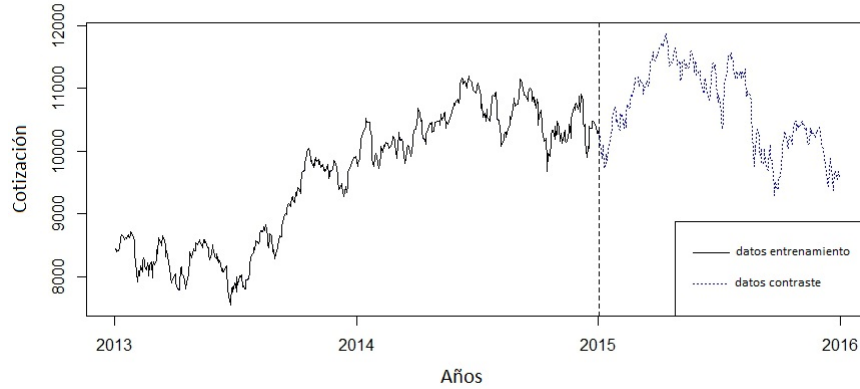


Figura 3.2: Serie temporal de la cotización diaria del IBEX35 desde Enero del 2013 hasta Diciembre del 2015.

en el Algoritmo 1, se construyen números difusos trapezoidales simétricos A_i , $i = 1, \dots, 52$.

El cuadro 3.1 muestra como utilizando el conjunto de entrenamiento y asignando distintos valores a γ se calculan los errores de ajuste, donde se ve la influencia creciente de la tendencia. Aplicando las reglas heurísticas para aproximar γ se observa que se alcanza el ajuste óptimo para $\gamma = 0.9$ (RMSE = 29.39), de este modo dicho valor de γ se utiliza para construir la predicción difusa a una etapa con los datos de contraste.

	$\gamma=0$	$\gamma=0.1$	$\gamma=0.3$	$\gamma=0.5$	$\gamma=0.7$	$\gamma=0.8$	$\gamma=0.9$	$\gamma=1$
RMSE	40.95	38.89	35.20	32.24	31.11	29.66	29.39	29.43

Cuadro 3.1: Errores de ajuste de los datos de entrenamiento usando diferentes γ -valores.

Ahora véase explícitamente cómo construir la primera predicción difusa, siendo $y_{511} = 10350.8$ la observación final en el conjunto de entrenamiento y el último conjunto difuso observado es A_{37} , cuya parte derecha asociada de las RLDG es $A_{35}, A_{36}, A_{37}, A_{38}, A_{39}$, con un vector de pesos cronológicos ya normalizados

$$w_{37} = [0.14, 0.28, 0.27, 0.28, 0.03]$$

nótese que los pesos mostrados son los que están asociados a cada uno de los conjuntos difusos de la parte derecha de RLDG, ya que en este caso el resto de pesos asociados al resto de conjuntos difusos son cero.

Entonces, el procedimiento devuelve el número difuso

$$\hat{A}_{512}^c = (10283, 10386, 51.5, 51.5)$$

Por otro lado, el vector de pesos positivos asociados con los saltos es

$$S = [s'_{-4}, \dots, s'_3] = [.005, .016, .056, .216, .376, .250, .067, .014]$$

sólo se muestran los pesos asociados a los tamaños de salto de -4 a 3 ya que los pesos asociados al resto de tamaños de salto son cero o muy próximos a cero. Luego el procedimiento calcula el número difuso trapezoidal, \hat{A}_{512}^s , basado en el vector de pesos S , como

$$\hat{A}_{512}^s = \bigoplus_{h=-4}^3 s'_h A_{37+h} = (10309.5, 10412.5, 51.5, 51.5)$$

Finalmente, la predicción difusa de una etapa es

$$\hat{A}_{512} = 0.9\hat{A}_{512}^s \oplus 0.1\hat{A}_{512}^c = (10306.85, 10409.85, 51.5, 51.5)$$

La predicción puntual, $\hat{y}_{512} = 10358.35$, será el punto medio del intervalo posibilístico de medias de \hat{A}_{512} , el valor observado es $y_{512} = 9993.3$, con error de predicción de un paso $|\hat{y}_{512} - y_{512}| = 365.05$

La figura 3.3 muestra en el lado izquierdo la serie temporal difusa de los últimos datos diarios observados del período de entrenamiento de este ejemplo. En el lado derecho, el gráfico corresponde a la predicción difusa \hat{A}_{512} y la predicción puntual \hat{y}_{512} (punto verde). En esta sección se ha estudiado el funcionamiento del Algoritmo 3 en un ejemplo concreto, véase a continuación un experimento numérico en el que se analizará la calidad de las predicciones puntuales y se comparará con los resultados obtenidos por otros procedimientos.

3.2. Experimentos numéricos

Para verificar la precisión predictiva del método STD descrito en la sección 3.1, denominado como Algoritmo 3, se utilizan conjuntos de datos históricos de cuatro índices bursátiles de referencia (IBEX35, NIKKEI225, DAX30 y TAIEX) obtenidos en <https://www.google.com/finance> y en la website que tiene el índice de Taiwan. Se utiliza el valor de cotización diaria como dato

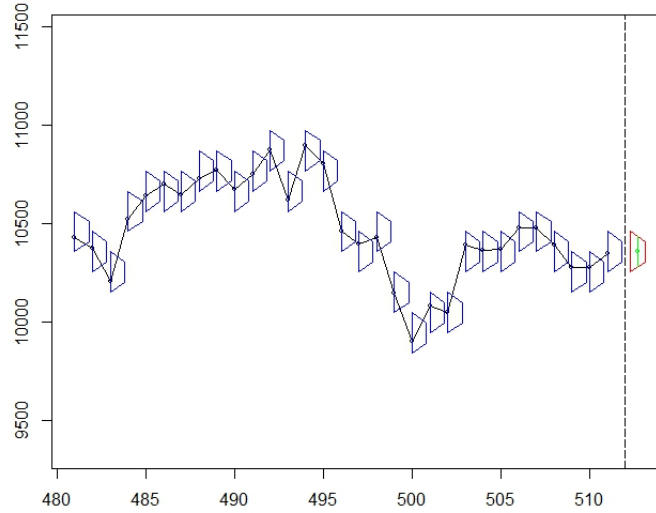


Figura 3.3: Representación parcial de una serie temporal difusa de cotizaciones diarias del IBEX35 y la predicción difusa, y puntual de una etapa.

en estudio, desde enero de 2013 hasta diciembre de 2015. Los datos de los primeros dos años se utilizan como entrenamiento y los datos del año 2015 para el contraste. Hay que tener en cuenta que los tamaños de las muestras pueden ser diferentes dependiendo de los días de apertura de cada mercado bursátil y país.

A continuación se expone el experimento numérico diseñado para comprobar el funcionamiento de la metodología propuesta: se aplica una estrategia de horizonte móvil semanal para cinco datos diarios consecutivos del conjunto de datos de entrenamiento (como representación de una semana estándar), de tal manera que los pesos proporcionados en el período de entrenamiento se mantengan en cada ciclo. Sin embargo, la predicción difusa diaria de una etapa depende del último dato observado antes de la predicción. La estrategia de horizonte móvil conserva el tamaño de cada conjunto de entrenamiento eliminando sus primeras cinco observaciones e incluyendo los últimos cinco datos semanales observados. Las relaciones lógicas difusas para el nuevo conjunto de entrenamiento y el RLDG se recalculan cada vez que se añade un período y, también, la matriz de pesos W y el vector de pesos S . Este experimento se repite con cada ejecución para los períodos semanales corres-

pondientes del conjunto de datos de 2015, respectivamente.

	IBEX35	NIKKEI225	DAX30	TAIEX
Datos de entrenamiento	511	502	508	494
Datos de contraste	255	245	250	240
Ejecuciones	51	49	50	48
Tiempo total (segundos)	45.41	31.05	39.06	26.97
Tiempo promedio	0.89	0.63	0.78	0.56

Cuadro 3.2: Tamaños de datos experimentales y tiempos de computación.

El cuadro 3.2 muestra los tamaños de las muestras y el tiempo computacional dedicado por nuestro método de series temporales difusas con tendencia para resolver los experimentos numéricos. En nuestras pruebas computacionales, una vez que se proporciona la base de datos, nuestro procedimiento proporciona la predicción difusa con un promedio de 0.72 segundos para cada ejecución (de los cuatro índices bursátiles).

El rendimiento del Algoritmo 3 se compara con los métodos ponderados de referencia en STD: Yu (2005b), Cheng et al. (2008a) y Wang et al. (2013). Posteriormente también se hace un análisis comparativo con modelos estadísticos de predicción clásicos (*MM1*, *SE* y *ARIMA*). A lo largo de este capítulo en los experimentos numéricos, con tal de reducir el número de citas a los diferentes autores y que no se creen excesivas repeticiones de las citas, se denota: Yu (2005b) como Yu, Cheng et al. (2008a) como Cheng y Wang et al. (2013) como Wang.

Para analizar la precisión del rendimiento de los diferentes métodos STD, se utilizan el RMSE y el MAD de los errores de predicción de una etapa, considerando bloques de cinco días, aunque no coincidan necesariamente con los días de una semana estándar. Este experimento numérico está separado en cuatro apartados independientes, un apartado para cada índice bursátil. Véase a continuación cada uno de los apartados.

3.2.1. Predicción del índice IBEX35

En el primer experimento, se consideran los valores diarios del índice bursátil español IBEX35 entre los años 2013 y 2015, tomando como datos de entrenamiento de enero del año 2013 a diciembre del año 2014 con un total

de quinientas once cotizaciones diarias, y los datos de contraste de enero a diciembre del año 2015 con un total de doscientas cincuenta y cinco cotizaciones diarias. Nótese que este conjunto de entrenamiento coincide con el presentado en el ejemplo de predicción a una etapa de la sección 3.1.4.

Siguiendo una estrategia de horizonte móvil, el experimento se repite cincuenta y una veces tomando períodos semanales de cinco cotizaciones. La figura 3.4 muestra la serie temporal de las observaciones diarias y las predicciones diarias obtenidas por nuestro procedimiento de tendencia para el año 2015, usando $\gamma = 0.9$.

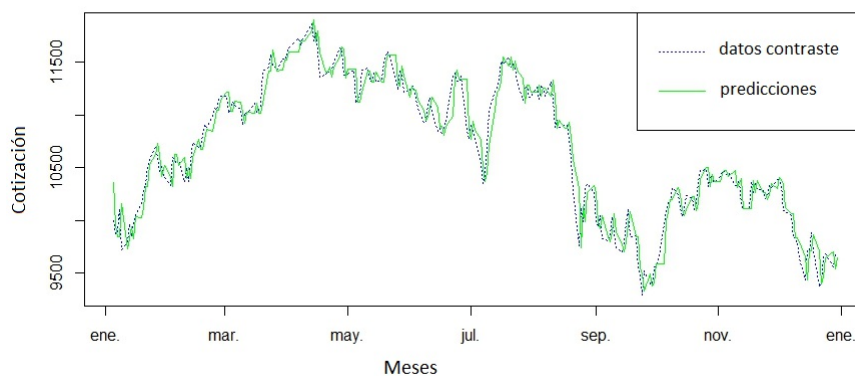


Figura 3.4: Trazado conjunto de la serie temporal de las cotizaciones diarias de IBEX35 y de las predicciones proporcionadas por nuestro método STD de tendencia Algoritmo 3, de enero de 2015 a diciembre de 2015.

En la primera semana de predicción con los datos de contraste (esto es $i = 512, \dots, 516$) se calculan los errores de predicción para cada día de dicha semana, $e_i = \hat{y}_i - y_i$ donde \hat{y}_i corresponde a la predicción puntual. La media de sus valores absolutos, $MAD = \sum_i |e_i|/5$, y la raíz cuadrada de la media de los errores al cuadrado, $RMSE = \sqrt{\sum_i e_i^2/5}$. De forma análoga a este período se evalúan los errores de predicción durante los siguientes 50 períodos semanales. Finalmente, se obtiene el promedio de los errores RMSE y MAD de cincuenta y una semanas.

La precisión de nuestro método de predicción para STD se compara con los obtenidos aplicando los métodos de STD ponderadas de Yu, Cheng y Wang; bajo las mismas condiciones experimentales. El cuadro 3.3 muestra los resultados obtenidos, donde se puede observar que nuestra propuesta para

STD obtiene mejor precisión de predicción que los otros métodos ponderados.

	Yu	Cheng	Wang	Algoritmo 3
MAD	116.35	119.29	394.19	110.40
RMSE	145.49	144.55	429.27	134.27

Cuadro 3.3: Promedio de errores de predicción para diferentes modelos de STD, con el conjunto de datos IBEX35.

El método propuesto en Wang et al. (2013) es el que proporciona los peores resultados, esto puede deberse a que proponen una partición de U de tan sólo siete intervalos. Esta partición podría ser apropiada en series pequeñas o series temporales con baja correlación, pero series largas altamente autocorreladas (según la definición clásica), como las que se consideran en este capítulo de la tesis, parecen necesitar una partición más fina.

Para analizar el comportamiento de los errores de predicción logrados por cada método, se realizó un análisis estadístico. El cuadro 3.4 muestra los p-valores ajustados de la comparación estadística por pares de los errores de predicciones RMSE (submatriz inferior) y MAD (submatriz superior), que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (tiene en cuenta el problema de comparación múltiple).

	Yu	Cheng	Wang	Algoritmo 3
Yu		0.453	6.6e-11	0.223
Cheng	0.840		7.0e-11	0.028
Wang	3.3e-11	2.3e-11		6.6e-11
Algoritmo 3	0.020	0.020	2.1e-11	

Cuadro 3.4: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos de STD, con el conjunto de datos IBEX35.

El análisis estadístico revela que existen diferencias significativas entre la media del RMSE alcanzado por nuestra propuesta y los otros métodos de predicción, a favor de nuestra propuesta. Sin embargo, no hay diferencia estadística entre la media de los errores de predicción entre RMSE de Yu y Cheng et al., mientras que el Wang et al. presenta diferencias significativas

con todos los otros métodos STD ponderadas. El análisis estadístico de los errores de predicción MAD de Cheng y Wang muestran un esquema similar al estudio de los errores RMSE, sin embargo con Yu se observa que no hay diferencias significativas pese a que en media el Algoritmo 3 tenga un error MAD inferior a la propuesta de Yu.

Al igual que en el capítulo anterior, también se analizan los resultados comparándolos con modelos estadísticos de predicción clásicos, aunque no se han podido aplicar exactamente las mismas condiciones experimentales descritas en este apartado. Se realiza una estrategia de horizonte móvil diario, i.e., cada día observado de los datos de contraste se agrega a los datos de entrenamiento, desprendiéndose de los datos de entrenamiento también un día (el primero de todos) y ejecutándose de nuevo el modelo estadístico para realizar una nueva predicción.

Posteriormente para el análisis y comparación con el Algoritmo 3, se agrupan las predicciones en conjuntos de cinco, calculándose de cada conjunto de cinco predicciones los errores MAD y RMSE obteniendo cincuenta y un errores de predicción, finalmente se calculan las medias de los errores del MAD y del RMSE que se muestran en el cuadro 3.5.

	MM1	SE	ARIMA	Algoritmo 3
MAD	107.41	107.43	107.04	110.40
RMSE	131.49	131.55	131.63	134.27

Cuadro 3.5: Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del IBEX35.

Análogamente al análisis estadístico anterior, el cuadro 3.6 muestra los p-valores ajustados de la comparación estadística por pares de los errores de predicciones RMSE (submatriz inferior) y MAD (submatriz superior), que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (tiene en cuenta el problema de comparación múltiple).

Véase el cuadro 3.5 cómo todos los modelos estadísticos de predicción mejoran en media de los errores de predicción, tanto del MAD como del RMSE, a las medias de los errores de predicción del Algoritmo 3, aunque el análisis estadístico del cuadro 3.6 revela que no existen diferencias significativas entre los errores medios del RMSE y del MAD obtenidos por el Algoritmo 3 y los otros modelos estadísticos de predicción. A raíz del análisis estadístico, tampoco se observan diferencias significativas entre los distintos modelos

	<i>MM1</i>	<i>SE</i>	<i>ARIMA</i>	Algoritmo 3
<i>MM1</i>		0.23	0.25	0.25
<i>SE</i>	0.55		1.00	1.00
<i>ARIMA</i>	0.55	1.00		1.00
Algoritmo 3	0.55	1.00	0.55	

Cuadro 3.6: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos IBEX35.

estadísticos clásicos de predicción, hecho que concuerda con lo observado en los errores de predicción del cuadro 3.5.

3.2.2. Predicción del índice NIKKEI225

Ahora considérese el conjunto de datos del índice de referencia bursátil NIKKEI225 entre los años 2013 a 2015, donde los datos de entrenamiento comprenden el período que va desde enero del año 2013 a diciembre del año 2014 obteniendo un total de quinientas dos cotizaciones diarias y los datos de contraste comprenden el período que va desde enero hasta diciembre del año 2015 con un total de 245 cotizaciones diarias. La figura 3.5 muestra el trazo de la serie temporal del NIKKEI225. Tener en cuenta que en algunas ocasiones la última observación podría situarse en alguno de los extremos del universo de los datos o muy cerca, por lo que es muy importante considerar una medida de tendencia para construir las predicciones de una etapa.

Para el conjunto de datos de entrenamiento inicial el procedimiento establece el universo definido como $U = [9500, 19016]$ que se parte en $m = 61$ intervalos iguales de longitud $l = 156$. Utilizando las reglas heurísticas para determinar la importancia de la tendencia a partir del conjunto de entrenamiento, permite obtener el valor γ de tal forma que minimiza el RMSE de los errores de ajuste de una etapa, en este caso para $\gamma = 1$ se tiene un mínimo de RMSE= 46.90.

Se intenta tomar particiones más finas de la unidad, particiones de tamaño 0.01, donde se obtuvo un valor óptimo diferente para $\gamma = 0.98$, con RMSE = 46.88. Estos resultados no son ni estadísticamente ni numéricamente significativos; además, se analiza este sobreajuste para el valor γ , y se comprueba que no proporciona mejores resultados en la fase de predicción de una etapa (RMSE es de 220.24 en lugar de 220.13 para $\gamma = 1$), por lo que se desestima este sobreajuste de γ .

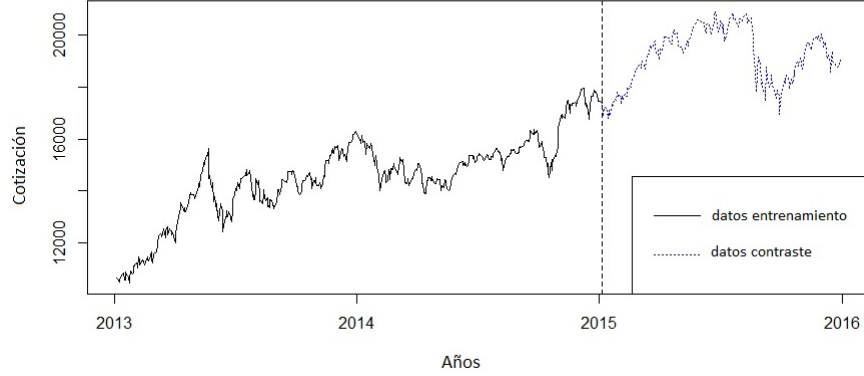


Figura 3.5: Trazo de la serie temporal de cotizaciones diarias del NIKKEI225 japonés, de enero de 2013 a diciembre de 2015.

Por lo tanto, dado que $\gamma = 1$ nuestro procedimiento STD solo funcionará con la información de la tendencia proporcionada por los pesos asociados a los saltos. Esta aproximación de γ se debe principalmente a que son series temporales altamente autocorreladas, así pues existe una mayor influencia de la tendencia sobre la serie temporal.

Véase ahora el funcionamiento del Algoritmo 3 en este experimento, la última observación es $y_{502} = 17409$ y el último conjunto difuso observado A_{51} . El procedimiento obtiene el vector de pesos $S = [s'_{-8}, \dots, s'_5]$, y proporciona la predicción difusa $\hat{A}_{503}^s = (17311, 17467, 78, 78)$. Finalmente, la predicción puntual es $\hat{y}_{503} = 17389$, y el error de predicción puntual de una etapa es $|\hat{y}_{503} - y_{503}| = 505.7$.

En este experimento, el horizonte móvil trata con 49 períodos, donde cada período se compone 5 datos diarios consecutivos. Los cuadros 3.7 y 3.8 muestran los errores promediados de predicción MAD y RMSE para estos 49 períodos, y los p-valores ajustados de la comparación por pares de los errores RMSE de predicción a través del test-t utilizando el método Holm para ajustar los p-valores.

Nuestra propuesta nuevamente logra una mejor precisión de predicción puntual de una etapa que los otros modelos de STD ponderadas, para ambos errores de predicción. El análisis estadístico reveló que existen diferencias

	Yu	Cheng	Wang	Algoritmo 3
MAD	199.22	221.05	445.84	182.33
RMSE	245.16	277.02	510.44	220.13

Cuadro 3.7: Promedio de errores de predicción del índice NIKKEI225, con conjunto de datos del año 2015 y para diferentes modelos de STD.

significativas entre la media del RMSE alcanzado por nuestra propuesta y los otros métodos de predicción, a favor de nuestra propuesta, y la misma evidencia significativa con respecto de los métodos de Cheng y Wang con el error MAD, en cuanto al método de Yu y el análisis comparativo del error MAD se observa que no hay diferencias significativas.

	Yu	Cheng	Wang	Algoritmo 3
Yu		0.170	4.0e-08	0.170
Cheng	0.063		9.6e-07	5.6e-05
Wang	1.0e-07	5.2e-06		1.0e-08
Algoritmo 3	0.043	2.9e-05	8.1e-09	

Cuadro 3.8: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos STD, con el conjunto de datos NIKKEI225 del año 2015.

En el cuadro 3.8 muestra los p-valores ajustados (por el método de Holm) de los errores de predicción, donde la submatriz inferior corresponde a los errores RMSE y la submatriz superior corresponde a los errores MAD.

Para el índice NIKKEI225 también se analizan los resultados comparándolos con modelos estadísticos de predicción clásicos.

	MM1	SE	ARIMA	Algoritmo 3
MAD	177.40	177.70	179.58	182.33
RMSE	213.89	214.05	215.95	220.13

Cuadro 3.9: Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del NIKKEI225.

Al igual que en el experimento con el índice IBEX35, no se pueden aplicar exactamente las mismas condiciones experimentales descritas en este apartado. Se realiza una estrategia de horizonte móvil diario análogo al experimento

del IBEX35. Posteriormente para el análisis y comparación con el Algoritmo 3, se agrupan las predicciones en conjuntos de cinco, calculándose de cada conjunto de cinco predicciones los errores MAD y RMSE obteniendo cincuenta errores de predicción, y finalmente se calculan las medias de los errores del MAD y del RMSE obteniendo el cuadro 3.9.

	MM1	<i>SE</i>	<i>ARIMA</i>	Algoritmo 3
MM1		0.97	0.93	0.93
SE	0.80		0.93	0.93
ARIMA	0.32	0.80		0.97
Algoritmo 3	0.32	0.80	0.80	

Cuadro 3.10: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos NIKKEI225.

Análogamente al análisis estadístico anterior, el cuadro 3.10 muestra los p-valores ajustados de la comparación estadística por pares de los errores de predicciones RMSE (submatriz inferior) y MAD (submatriz superior), que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (tiene en cuenta el problema de comparación múltiple).

Obsérvese que todos los modelos estadísticos de predicción vuelven a mejorar en media los errores de predicción tanto del MAD como del RMSE a los obtenidos por el Algoritmo 3, aunque también el análisis estadístico del cuadro 3.10, se muestra nuevamente que no existen diferencias significativas. Tampoco se observan diferencias significativas entre los distintos modelos estadísticos clásicos de predicción.

3.2.3. Predicción del índice DAX30

En este experimento, los procedimientos de STD se aplican a los conjuntos de datos de la bolsa de valores de Alemania. La figura 3.6 muestra las cotizaciones diarias para los años 2013, 2014 y 2015. El conjunto de datos de entrenamiento se compone de las cotizaciones diarias del período que va desde enero del año 2013 a diciembre del año 2014 con un total de 508 observaciones y el conjunto de datos de contraste se compone por las cotizaciones diarias del período que comprende desde enero a diciembre del año 2015 con un total de 250 observaciones.

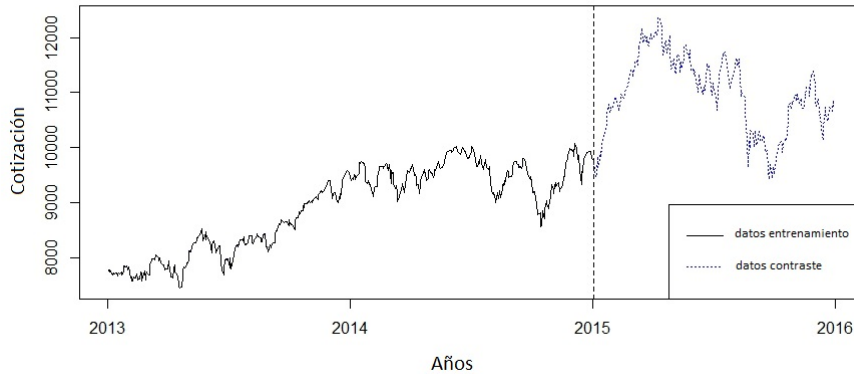


Figura 3.6: Trazo de la serie temporal de cotizaciones diarias del DAX30.

En primer lugar, nuestro procedimiento calcula el mejor valor de γ para las 508 cotizaciones del conjunto de datos de entrenamiento DAX30, $\gamma = 1$. Luego, realiza el mismo experimento numérico, anteriormente aplicado al IBEX35 y al NIKKEI225, para el conjuntos de datos del índice DAX30, en el cual las RLD y las RLDG son redefinidas cada vez que se observa un nuevo conjunto de datos semanales, lo que permite incorporar información granular reciente. El horizonte rodante se aplica para 50 períodos semanales, respectivamente.

Finalmente, el cuadro 3.11 muestra los errores de predicción semanal promediados (MAD y RMSE) logrados para los métodos de STD ponderadas. Estos resultados comparativos también son consistentes con el mejor rendimiento de nuestro método de STD de tendencia, al menos con respecto a la precisión de la predicción.

	Yu	Cheng	Wang	Algoritmo 3
MAD	145.87	139.98	343.25	126.34
RMSE	172.69	170.56	376.80	153.15

Cuadro 3.11: Promedio de errores de predicción del índice alemán DAX30, con conjunto de datos del año 2015 y para diferentes modelos de STD.

En el cuadro 3.12 la submatriz inferior contiene los p-valores para las comparaciones por pares de RMSE para el conjunto de datos DAX30 y la

submatriz superior contiene los p-valores para las comparaciones por pares de MAD para el conjunto de datos DAX30. Se encuentran resultados análogos cuando se compararon los errores de predicción de una etapa promediados usando MAD. Las diferencias estadísticas y las homogeneidades siguen el mismo esquema que RMSE promediado.

	Yu	Cheng	Wang	Algoritmo 3
Yu		0.364	1.5e-07	0.002
Cheng	0.790		3.5e-07	0.007
Wang	1.1e-07	4.0e-07		2.2e-08
Algoritmo 3	0.003	0.01	1.0e-08	

Cuadro 3.12: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos STD, con el conjunto de datos DAX30 2015.

Con respecto al análisis estadístico de la comparación por pares del RMSE y MAD promedios, para el Algoritmo 3 los p-valores más bajos están asociados al método de Wang, mientras que el mayor p-valor obtenido para nuestra propuesta en comparación con los otros dos métodos STD ponderadas es de 0.01 y que por lo tanto sí hay diferencias estadísticamente significativas entre el Algoritmo 3 y el resto de métodos de STD. Por otro lado obsérvese que no hay diferencias significativas entre los métodos de Yu y Cheng.

Para el índice DAX30 también se analizan los resultados comparándolos con modelos estadísticos de predicción clásicos.

	MM1	SE	ARIMA	Algoritmo 3
MAD	123.64	123.64	123.37	126.34
RMSE	149.49	149.75	149.50	153.15

Cuadro 3.13: Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del DAX30.

Al igual que en experimentos anteriores, no se pueden aplicar exactamente las mismas condiciones experimentales descritas en este apartado. Se realiza una estrategia de horizonte móvil diario siguiendo el mismo procedimiento descrito en los experimentos anteriores. Posteriormente para el análisis y comparación con el Algoritmo 3, se agrupan las predicciones en conjuntos de cinco, calculándose de cada conjunto de cinco predicciones los errores MAD

y RMSE obteniendo cincuenta errores de predicción, y finalmente se calculan las medias de los errores del MAD y del RMSE obteniendo el cuadro 3.13.

	MM1	<i>SE</i>	<i>ARIMA</i>	Algoritmo 3
MM1		0.38	0.46	0.46
SE	0.19		1.00	1.00
ARIMA	0.20	1.00		1.00
Algoritmo 3	0.25	1.00	0.40	

Cuadro 3.14: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos DAX30.

Análogamente al análisis estadístico anterior, el cuadro 3.14 muestra los p-valores ajustados de la comparación estadística por pares de los errores de predicciones RMSE (submatriz inferior) y MAD (submatriz superior), que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (tiene en cuenta el problema de comparación múltiple).

A tenor de los resultados recogidos en el cuadro 3.13 se observa que todos los modelos estadísticos de predicción mejoraran en media los errores de predicción a los obtenidos por el Algoritmo 3, sin embargo del análisis estadístico se desprende que no existen diferencias significativas entre los resultados obtenidos de los distintos modelos y el Algoritmo 3. Tampoco se observan diferencias significativas entre los distintos modelos estadísticos clásicos de predicción.

3.2.4. Predicción del índice TAIEX

Finalmente los procedimientos de STD se aplican al conjunto de datos de la bolsa de valores de Taiwán, cuyo comportamiento temporal durante el período considerado difiere de los anteriores experimentos. La figura 3.7 muestra las cotizaciones diarias para los años 2013, 2014 y 2015, donde se refleja la presencia de tendencias positivas y negativas.

En primer lugar, nuestro procedimiento calcula el mejor valor de γ para las 494 cotizaciones del TAIEX que es $\gamma = 0.9$. Luego, se aplican las mismas condiciones experimentales de horizonte rodante semanal que a los anteriores índices bursátiles (IBEX35, NIKKEI225 y DAX30). El horizonte rodante se aplica para un total de cuarenta y ocho períodos semanales.

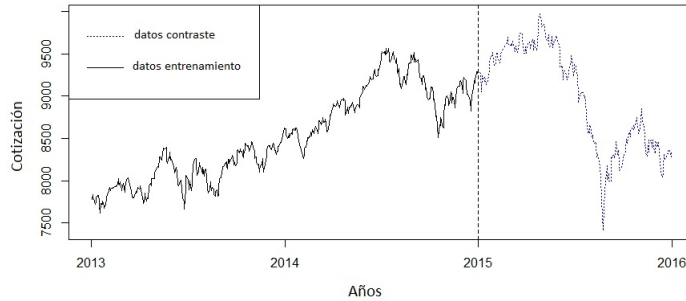


Figura 3.7: Trazo de la serie temporal de cotizaciones diarias del TAIEX.

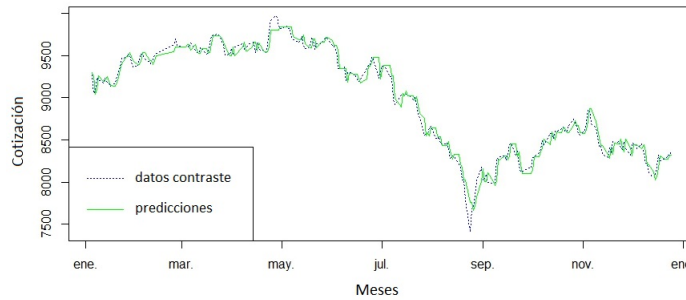


Figura 3.8: Trazo de la serie temporal de cotizaciones diarias del TAIEX y predicciones, desde Enero del 2015 hasta Diciembre del 2015.

La figura 3.8 muestra el trazo de las predicciones diarias obtenidas utilizando nuestro método ponderado de tendencia con las cotizaciones diarias observadas del TAIEX durante el año 2015.

Finalmente, el cuadro 3.15 muestra los errores de predicción semanal promediados (MAD y RMSE) logrados para los métodos de STD ponderadas. Estos resultados comparativos también son consistentes con el mejor rendimiento de nuestro método de STD de tendencia, al menos con respecto a la precisión de la predicción.

En el cuadro 3.16 la submatriz inferior presenta los p-valores ajustados del RMSE y la submatriz superior presenta los p-valores ajustados del MAD. Se encuentran resultados paralelos comparando los errores de predicción de una etapa promediados del RMSE y del MAD. Las diferencias estadísticas y

	Yu	Cheng	Wang	Algoritmo 3
MAD	71.84	70.59	198.93	65.36
RMSE	87.56	87.39	220.83	80.58

Cuadro 3.15: Promedio de errores de predicción del índice TAIEX, con conjunto de datos del año 2015.

las homogeneidades siguen el mismo esquema que RMSE promediado.

TAIEX	Yu	Cheng	Wang	Algoritmo 3
Yu		0.685	7.0e-09	0.037
Cheng	0.959		4.4e-09	0.037
Wang	1.5e-09	8.3e-10		1.4e-09
Algoritmo 3	0.027	0.011	8.3e-10	

Cuadro 3.16: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos STD.

Con respecto a la significación estadística de la comparación por pares del error promedio de predicción puntual de una etapa con el test-t para el RMSE y el MAD para el Algoritmo 3, los p-valores más bajos están asociados al método STD de Wang, para cada conjunto de datos. Además, el mayor p-valor obtenido para nuestra propuesta en comparación con los otros dos métodos STD ponderadas es de 0.027 para el RMSE y de 0.037 para el MAD; por lo que se deduce que sí hay diferencias significativas entre el Algoritmo 3 y el resto de métodos de STD, mientras que también se observa que no hay diferencias significativas entre los métodos STD de Yu y Cheng.

En este último experimento del índice TAIEX también se analizan los resultados comparándolos con modelos estadísticos de predicción clásicos.

	MM1	SE	ARIMA	Algoritmo 3
MAD	63.56	63.56	64.04	65.36
RMSE	78.02	78.02	78.40	80.58

Cuadro 3.17: Promedio de errores de predicción para modelos estadísticos de predicción clásicos y el Algoritmo 3, con el conjunto de datos del DAX30.

Al igual que en experimentos anteriores, no se pueden aplicar exactamente las mismas condiciones experimentales descritas en este apartado. Se realiza

una estrategia de horizonte móvil diario siguiendo el mismo procedimiento descrito en los experimentos anteriores. Posteriormente para el análisis y comparación con el Algoritmo 3, se agrupan las predicciones en conjuntos de cinco, calculándose de cada conjunto de cinco predicciones los errores MAD y RMSE obteniendo cincuenta errores de predicción, y finalmente se calculan las medias de los errores del MAD y del RMSE obteniendo el cuadro 3.17.

	MM1	SE	ARIMA	Algoritmo 3
MM1		1.00	1.00	1.00
SE	0.55		1.00	1.00
ARIMA	0.55	0.55		0.83
Algoritmo.3	0.55	0.55	0.55	

Cuadro 3.18: P-valores ajustados de la comparación por pares de los errores de predicción obtenidos con diferentes modelos estadísticos junto con el Algoritmo 3, con el conjunto de datos TAIEX.

Análogo al análisis estadístico anterior, el cuadro 3.18 muestra los p-valores ajustados de la comparación estadística por pares de los errores de predicciones RMSE (submatriz inferior) y MAD (submatriz superior), que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (tiene en cuenta el problema de comparación múltiple).

En el cuadro 3.17 se observa que todos los modelos estadísticos de predicción mejoraran en media los errores de predicción a los obtenidos por el Algoritmo 3, aunque del cuadro 3.18 se observa que no existen diferencias significativas entre los resultados obtenidos de los distintos modelos y el Algoritmo 3. Tampoco se aprecian diferencias significativas entre los distintos modelos estadísticos clásicos de predicción.

3.3. Conclusiones

En este capítulo se ha propuesto un nuevo algoritmo, el Algoritmo 3. En el Algoritmo 3 se ha descrito un método de predicción basado en una combinación lineal convexa (con parámetro γ) de dos modelos de ponderación para predicción de series temporales difusas, por un lado se ha utilizado el modelo cronológico introducido en el capítulo 2 (Algoritmo 1) y por otro se ha incorporado un nuevo modelo (de tendencia) basado en lo que se ha definido como tamaño de salto. También se ha descrito un procedimiento donde se ha estimado el parámetro γ a partir de los errores de ajuste. Adicionalmente el

algoritmo ha proporcionado números difusos trapezoidales como predicciones difusas a una etapa.

La principal diferenciación entre los modelos (cronológico y de tendencia) ha sido el origen de los pesos utilizados en las reglas heurísticas para la predicción difusa. En el modelo cronológico se ha propuesto utilizar un conjunto de pesos, obtenidos de la secuencia cronológica en la serie temporal difusa original, utilizando la información proporcionada por las relaciones lógicas difusas una a una y agrupándose dichos pesos a partir de las relaciones lógicas difusas de grupos. En el modelo de tendencia, que considera la posible tendencia de las series temporales difusas, se ha introducido un conjunto de pesos que miden la frecuencia relativa y la magnitud de los tamaños de saltos observados en las series temporales difusas.

Para analizar la calidad de las predicciones puntuales, se han propuesto varios experimentos numéricos, donde se han utilizado datos históricos de series temporales de cuatro índices bursátiles: IBEX35, NIKKEI225, DAX30 y el TAIEX. Los resultados de los algoritmos propuestos se han comparado en cada uno de los experimentos numéricos con tres métodos conocidos de series temporales difusas ponderadas (Yu, 2005b; Cheng et al., 2008a; Wang et al., 2013). Se ha observado una mejora en la calidad de la predicción en todos los experimentos y que en un análisis estadístico posterior se ha mostrado que dichas diferencias son estadísticamente significativas, tanto en los errores de predicción del RMSE como del MAD. Salvo en los dos primeros experimentos (con el IBEX35 y NIKKEI225) que no se han apreciado diferencias significativas con el método de Yu, en el análisis estadístico del error MAD, pese a que el Algoritmo 3 ha tenido en media un error inferior.

Obsérvese, en el Algoritmo 3, que ante las sucesivas aproximaciones del parámetro γ a través del error de ajuste, en todos los experimentos siempre γ ha sido 1 o un valor muy cercano. Este hecho por un lado ha denotado una tendencia subyacente de la serie temporal a estudiar, lo que ha devengado en un mejor funcionamiento o precisión del modelo de tendencia. En todos los experimentos propuestos, los análisis estadísticos realizados de los errores de predicción RMSE y MAD, entre los modelos estadísticos de predicción (*SE*, *ARIMA* y *MM1*) y el Algoritmo 3, han concluido que no existen diferencias significativas.

La calidad de la información que ha proporcionado la tendencia descrita en este capítulo, a través del tamaño de salto, está relacionada con el número de intervalos que componen el universo de los datos U ; si hay un pequeño número de intervalos, esta medida podría ser menos informativa. Sin embargo, una contrapartida interesante de usar saltos en STD es que ha permitido introducir predicciones puntuales de una etapa incluso cuando la última observación se encuentra en los extremos del universo, i.e., podría producirse el caso de que nuestra predicción saliera fuera del universo definido. Esta casuística generalmente no se contempla en la mayoría de los métodos STD, que sólo construyen predicciones de una etapa dentro del universo definido. Nótese que el conjunto de ponderaciones cronológicas basadas en RLDG ha intentado incorporar información de patrones idénticos observados en la STD, por lo que podrían ser útiles cuando la serie temporal no muestra una tendencia clara. Por otro lado, el conjunto de pesos basados en saltos intenta aprender de patrones afines -concretamente, patrones idénticos después de una traslación en la escala del universo U -, y podrían ser muy útiles cuando la serie temporal tenga una tendencia clara porque, en ese caso, la serie temporal difusa presenta pocos patrones idénticos pero muchos patrones afines.

Por lo tanto en el siguiente capítulo se incorpora la minería de datos, estableciendo reglas heurísticas basadas en el tamaño de la muestra N del conjunto de datos, la amplitud de U y la variabilidad de los datos observados.

Minería de datos en series temporales

En este capítulo se propone, a partir de los tamaños de saltos introducidos en el capítulo anterior, un modelo de patrones utilizando lo que se definirá más adelante como rachas. También se introducen ciertos cambios en la algoritmia de predicción que favorecen la introducción de dichos conceptos. Se propone utilizar técnicas de suavizado que mejoran la información extraída del procedimiento de ponderación, ya que se observan algunos problemas derivados de la falta de frecuencias de ciertos sucesos entre las relaciones de los datos, devengando un notable empeoramiento del proceso de ponderación.

Ante la variedad de algoritmos propuestos (cronológico, tendencia y patrones) y debido a la influencia del tamaño del conjunto de datos en el estudio de la eficiencia en cada uno de los modelos, se propone un SAD (sistema de ayuda a la decisión), utilizando la minería de datos y el tamaño del conjunto de datos disponible.

Finalmente, para comprobar la calidad de las predicciones ante los cambios propuestos y tomando los mismos conjuntos de datos de los índices bursátiles (DAX30 y TAIEX) introducidos en el capítulo anterior, se comparan con métodos clásicos de predicción de series temporales y con los métodos de referencia en series temporales difusas ponderadas.

Este capítulo se estructura como sigue: primero se introducen las principales definiciones (rachas y patrones), luego se ven algunos de los cambios pro-

Parte del contenido de este capítulo viene recogido en un capítulo del libro en que aparece el trabajo de Rubio et al. (2017b) de la editorial Springer.

puestos en la algoritmia ya descrita en el capítulo 2, y en las reglas heurísticas de predicción, continuando con un ejemplo de aplicación del nuevo algoritmo; después se explican cómo y dónde utilizar las técnicas de suavizado de patrones, y en la segunda sección se propone una aproximación alternativa para la predicción difusa. Se continúa con la tercera sección con la proposición del SAD. En una cuarta sección se analizan los experimentos numéricos y se finaliza el capítulo con las conclusiones.

4.1. Rachas y patrones

Hasta ahora todos los modelos propuestos se han basado en dos observaciones consecutivas en las series temporales difusas. En esta sección se presenta un *modelo de patrón ponderado* que se define a partir de patrones de más de dos observaciones consecutivas basándose en el concepto de *racha*, la cual es utilizada ampliamente en otros contextos.

Definición 4.1.

Una *racha* de longitud k , con $k \geq 1$ es una secuencia de k tamaños de salto observados consecutivamente, en los que se permiten repeticiones.

□

La representación vectorial de una racha de longitud k denotada como r_k , es como sigue: $r_k = (p_1, p_2, \dots, p_k)$ siendo p_i para $i = 1, 2, \dots, k$ tamaños de salto observados cronológicamente consecutivos según el subíndice.

Definición 4.2.

Se define *patrón* de orden k a las frecuencias de rachas de longitud k sobre la serie de tamaños de salto.

□

A continuación véase como calcular el patrón de orden k de la frecuencia relacionada con un tamaño de salto q . Sea $q \in [-m + 1, m - 1]$ el tamaño de salto a estudiar y $r_k = (p_1, p_2, \dots, p_k)$ una racha de longitud k tal que $p_k = q$ luego $r_k = (p_1, p_2, \dots, p_{k-1}, q)$, entonces

$$\alpha_{r_k, q} = \sum_t \beta_{r_k, q}^t \tag{4.1}$$

esta ecuación mide la frecuencia con la que se observa el salto q dentro de una racha concreta r_k a lo largo de la serie temporal de saltos, donde

$$\beta_{r_k,q}^t = \begin{cases} 1 & \text{si } J(t) = p_1, J(t+1) = p_2, \dots, J(t+k-1) = p_{k-1}, J(t+k) = q \\ 0 & \text{en otro caso} \end{cases}$$

siendo $J(t)$ el tamaño de salto observado en el instante t . En el caso que el tamaño del conjunto de los datos observados sea N y m el número de conjuntos difusos construidos, por un lado se tiene que $1 \leq t \leq N - k$ y por otro $-m + 1 \leq p_i \leq m - 1 \forall i$; pudiendo ser representado el patrón de orden k mediante una matriz de dimensión $(2m - 1)^k$ y se denota como $S^* = [\alpha_{r_k,p}]$ con $-m + 1 \leq p \leq m - 1$. El proceso de normalización de la matriz S^* se realiza por filas, tomando como referencia la fila relacionada con cada racha r_k .

Se han introducido las definiciones teóricas a utilizar para este nuevo procedimiento de predicción difusa, a continuación se verán algunos cambios propuestos en la algoritmia de predicción para series temporales difusas.

4.1.1. Predicción difusa con patrones

Se llama Algoritmo 4 al procedimiento descrito a continuación. El esquema que sigue es igual que el del Algoritmo 3.

Primero, en cuanto a la partición del universo dependerá del número de datos observados y también del rango del conjunto de datos. Por otra parte como la amplitud de la partición determina la ambigüedad, entonces se decide que dicha amplitud también dependa de la desviación típica muestral. Así cuantos más datos se tienen, mayor es la información del sistema y menor debería ser su incertidumbre, por lo tanto menor debe ser la longitud de los intervalos que componen el universo de los datos.

Por ello se propone utilizar la medida del error estándar como longitud de los intervalos para hacer la partición del universo.

$$l = \sigma / \sqrt{N}$$

donde σ es la desviación típica muestral y N número de datos. Entonces sea el universo $U = [u_{inf}, u_{sup}]$ y sea el número de intervalos $m \in \mathbb{N}$ lo suficientemente grande como para que $u_{sup} \leq u_{inf} + ml$ donde l es el error estándar. Se construye cada uno de los intervalos que componen el universo, como sigue:

$$u_j = [u_{inf} + (j - 1)l, u_{inf} + jl] \quad j = 1, \dots, m$$

A partir de los intervalos se definen los números difusos trapezoidales A_i con $i = 1, \dots, m$ y se construye la serie temporal difusa.

Para el análisis de las rachas se debe tener una serie temporal de saltos y así se buscan patrones mediante la identificación de recurrencias en las rachas. Véase a continuación cómo se obtiene dicha serie de saltos.

Dados $F(t_0) = A_i, F(t_0 + 1) = A_j, F(t_0 + 2) = A_k$, se tiene que: $A_i \rightarrow A_j \rightarrow A_k$; así se tiene que $p = j - i$ y $q = k - j$ son los tamaños de saltos asociados a cada una de las relaciones con $-m + 1 \leq p, q \leq m - 1$, así se induce a la relación $p \leftrightarrow q$.

Por lo tanto las relaciones de la serie temporal difusa inducen relaciones entre saltos, obteniéndose una serie temporal de saltos y a partir de las secuencias de saltos (sobre la serie temporal) se construyen las rachas. Sea $\mathbb{S} = \{p \in \mathbb{Z} / -m + 1 \leq p \leq m - 1\}$ el rango de saltos de la serie de saltos, nótese que $|\mathbb{S}| = 2m - 1$.

Un patrón de orden k se representa matricialmente como S^* , esta matriz se calcula mediante la frecuencia relativa de la ocurrencia de rachas de longitud k y por lo tanto:

$$S^* \in \mathcal{M}_{(2m-1) \times \binom{k-2}{\cdot, \cdot} \times (2m-1)}(\mathbb{N})$$

así la dimensión de S^* sería $(2m - 1) \times \binom{k-2}{\cdot, \cdot} \times (2m - 1) = (2m - 1)^k$.

Finalmente, véase a continuación cómo se modifican las reglas heurísticas de predicción y cómo se calculan los pesos que se usan en la ponderación:

Sea $F(N) = A_i$ y sea $r_k = (p_{N-k}, \dots, p_N)$ la última racha observada de longitud k , donde p_j es el tamaño de salto observado para cada instante j con $j = N - k, \dots, N$.

Sea $v = (n_{-m+1}, \dots, n_0, \dots, n_{m-1})$ con $n_q = \alpha_{r_k, q}$ (ecuación 4.1) un vector fila de la matriz S^* asociado a la racha r_k . Sea $v' = (v'_1, \dots, v'_{2m-1})$ la normalización del vector v donde:

$$v'_j = \frac{n_j}{\sum_{q=-m+1}^{m-1} n_q}$$

la predicción difusa queda como sigue

$$\hat{A}_{N+1} = v'_{i+a} A_{i+a} \oplus v'_{i+(a+1)} A_{i+(a+1)} \oplus \dots \oplus v'_{i+b} A_{i+b} \quad (4.2)$$

con $a = \max\{-m+1, 1-i\}$, $b = \min\{m-1, m-i\}$. Nótese que para los casos donde $i + a < 1$ o $i + b > m$, todos los conjuntos difusos que no estuvieren previamente definidos, se construirían según establece el procedimiento y se procede como sigue: sea $k \in \mathbb{Z}$ tal que $i + a \leq k < 1$, entonces se define el conjunto difuso

$$A_k = (u_{inf} + (k - 1)l, u_{inf} + kl, l/2, l/2)$$

análogamente se construyen para $m < k \leq i + b$. De esta forma se permiten generar predicciones fuera del universo definido.

Este nuevo modelo de patrón es una generalización del modelo de tendencia (si se toma $k = 1$), introducido en el capítulo anterior y será útil cuando haya una tendencia, no necesariamente lineal. A continuación véase un ejemplo práctico de utilización de este nuevo procedimiento propuesto.

4.1.2. Ejemplo de aplicación del modelo de patrones

Para ejemplificar el funcionamiento del modelo de patrones, se toman rachas de longitud $k = 2$ y se aplica a la base de datos de matrículas de la Universidad de Alabama entre los años 70 y 90 (Song y Chissom, 1993a; Huarng, 2001a; Duru y Bulut, 2014), éste es un ejemplo clásico muy utilizado entre los investigadores en el campo de las STD para probar sus modelos.

años	matriculados	años	matriculados
1971	13055	1982	15433
1972	13563	1983	15497
1973	13867	1984	15145
1974	14696	1985	15163
1975	15460	1986	15984
1976	15311	1987	16859
1977	15603	1988	18150
1978	15861	1989	18970
1979	16807	1990	19328
1980	16919	1991	19337
1981	16388	1992	18876

Cuadro 4.1: Matriculados de la Universidad de Alabama

Se tiene que $\min(Y(t)) = 13055$ y $\max(Y(t)) = 19337$, el error estándar es 387.28 así la longitud del intervalo $l = 387.28$ y el número de intervalos

$m = 21$. Entonces el universo sería $U = [12000, 20132.28]$ y los intervalos que recubren el universo $u_1 = [12000, 12387.28], \dots, u_{21} = [19745.60, 20132.28]$

Se definen los números difusos trapezoidales a partir de los intervalos definidos tal y como viene descrito en el capítulo anterior figura 3.1, obteniendo: los números difusos trapezoidales $A_i = (12000 + (i - 1) * 387.28, 12000 + i * 387.28, 193.64, 193.64)$ para $i = 2, \dots, 20$ y los números trapezoidales truncados $A_1 = (12000, 12387.28, 0, 193.64), A_{21} = (19745.60, 20132.28, 193.64, 0)$ que corresponden a los conjuntos extremos del universo.

Véase a continuación cómo se obtiene la serie temporal difusa. Dadas las dos primeras observaciones correspondientes a los años 1971 y 1972, $F(1971) = A_3$ y $F(1972) = A_5$ inducen una relación lógica difusa $A_3 \rightarrow A_5$. Aplicado recurrentemente al conjunto de datos observados desde 1973 a 1992, la serie temporal difusa resulta como sigue:

$$A_3 \rightarrow A_5 \rightarrow A_5 \rightarrow A_7 \rightarrow A_9 \rightarrow A_9 \rightarrow A_{10} \rightarrow A_{10} \rightarrow A_{13} \rightarrow A_{13} \rightarrow A_{12} \rightarrow A_9 \rightarrow A_{10} \rightarrow A_9 \rightarrow A_9 \rightarrow A_{11} \rightarrow A_{13} \rightarrow A_{16} \rightarrow A_{18} \rightarrow A_{19} \rightarrow A_{19} \rightarrow A_{18}$$

El tamaño del salto se calcula como la diferencia de los subíndices de las RLD, el primer caso la RLD es $A_3 \rightarrow A_5$ teniendo un salto de tamaño 2 ($=5-3$). Por otro lado como el número de intervalos es $m = 21$ se tiene que el conjunto de todos los saltos sería $\mathbb{S} = \{k \in \mathbb{Z} / -20 \leq k \leq 20\}$ y $|\mathbb{S}| = 41$.

Tómese como ejemplo la observación en el instante $t_0 = 1981$ y las dos consecutivas, entonces $F(1981) = A_{12}, F(1982) = A_9, F(1983) = A_{10}$, con RLD $A_{12} \rightarrow A_9 \rightarrow A_{10}$, siendo las relaciones entre los tamaños de salto $9 - 12 = -3 \leftrightarrow 1 = 10 - 9$. Aplicado a toda la STD se induce la siguiente serie temporal de tamaños de saltos:

$$2 \leftrightarrow 0 \leftrightarrow 2 \leftrightarrow 2 \leftrightarrow 0 \leftrightarrow 1 \leftrightarrow 0 \leftrightarrow 3 \leftrightarrow 0 \leftrightarrow -1 \leftrightarrow -3 \leftrightarrow 1 \leftrightarrow -1 \leftrightarrow 0 \leftrightarrow 2 \leftrightarrow 2 \leftrightarrow 3 \leftrightarrow 2 \leftrightarrow 1 \leftrightarrow 0 \leftrightarrow -1$$

Obsérvese que los saltos varían entre -3 y 3, así se puede restringir $\mathbb{S} = \{k \in \mathbb{Z} / -3 \leq k \leq 3\}$ y $|\mathbb{S}| = 7$, de esta manera las matrices donde se almacenan las frecuencias serían de 7x7 mucho más manejables que matrices de 41x41, obteniéndose una matriz de frecuencias de saltos (cuadro 4.2).

En el ejemplo a tratar el último salto observado es menos uno, luego para hacer la predicción difusa se usa el vector $v'_{-1} = (0.5, 0, 0, 0.5, 0, 0, 0)$ y como la penúltima observación es $F(1991) = A_{19}$, se obtiene $F(1992) = \hat{A}_{1992} = 0.5 * A_{16} \oplus 0.5 * A_{19}$, obteniendo una estimación puntual de 18680.58, el valor observado en 1992 es 18876, obteniendo un error de 195.42 .

	-3	-2	-1	0	1	2	3
-3	0.00	0.00	0.00	0.00	1.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-1	1.00	0.00	0.00	1.00	0.00	0.00	0.00
0	0.00	0.00	2.00	0.00	1.00	2.00	1.00
1	0.00	0.00	1.00	2.00	0.00	0.00	0.00
2	0.00	0.00	0.00	2.00	1.00	2.00	1.00
3	0.00	0.00	0.00	1.00	0.00	1.00	0.00

Cuadro 4.2: Matriz de frecuencias de saltos asociada a rachas de longitud 2

A tenor del ejemplo obsérvese que el vector de pesos asociado a una racha preestablecida no es homogéneo debido a la falta de datos de entrenamiento. La debilidad de este modelo es que el conjunto de posibles ejecuciones podría ser muy grande, por lo que sólo es recomendable cuando N es grande. Para subsanar este inconveniente sin aumentar el número de observaciones, a continuación se propone utilizar una técnica de suavizado, que simularía el resultado de tener los suficientes datos de entrenamiento como para tener un vector de ponderación gradualmente homogéneo.

4.1.3. Suavizado de la matriz de patrones

Muchas veces cuando el número de observaciones para una racha dada no es muy alto, da como resultado vectores asociados a las rachas con frecuencias muy dispares. La utilización del suavizado sobre la matriz de patrones permite que dada una racha y un vector de frecuencias de salto asociado a dicha racha, la distribución de frecuencias de cada salto este gradualmente homogeneizada.

La regresión Kernel es una técnica no paramétrica en estadística para estimar la esperanza condicional de una variable aleatoria. El objetivo es encontrar una relación no lineal entre un par de variables aleatorias X y Z .

En una regresión no paramétrica, la esperanza condicional de una variable Z condicionada con respecto a X , se puede escribir como:

$$E(Z|X) = m(X)$$

donde m es una función desconocida.

En los trabajos de Nadaraya (1964) y Watson (1964) propusieron estimar la función $m(\cdot)$ con una media ponderada local, usando el kernel como función de ponderación. El estimador de Nadaraya-Watson es:

$$M^h(x) = \frac{\sum K_h(x - x_i) z_i}{\sum K_h(x - x_i)} \quad \text{para } i = 1, \dots, l$$

donde $K_h = K\left(\frac{x-x_i}{h}\right)$ es el kernel con una anchura h . El denominador es un término ponderado con suma igual a 1.

Se toma X como los tamaños de salto y Z la frecuencia absoluta en cada uno, con función de densidad conjunta $f(x, z)$. Entonces la función de regresión estimada sería:

$$M(X) = E(Z|X = x) = \sum_{z \in Z} z f(z|x) = \sum_{z \in Z} z \frac{f(x, z)}{f(x)} \quad \text{si } f(x) > 0$$

donde $f(x) = \sum_{z \in Z} f(x, z)$ es la densidad marginal, entonces para aproximar la función $m(\cdot)$ se estimará f .

Usando la estimación de densidad del kernel para la densidad conjunta $f(x, z)$ y $f(x)$ con un kernel K ,

$$\hat{f}(x, z) = l^{-1} h^{-2} \sum K\left(\frac{x-x_i}{h}\right) K\left(\frac{z-z_i}{h}\right)$$

$$\hat{f}(x) = l^{-1} h^{-1} \sum K\left(\frac{x-x_i}{h}\right)$$

donde h = ancho (podría depender de l), l = número de elementos y K = kernel es una función real elegida a priori.

La función K debe cumplir:

$$\int_{-\infty}^{+\infty} |K(x)| dx < \infty \quad \text{y que} \quad \int_{-\infty}^{+\infty} K(x) dx = 1$$

Se debe tener en cuenta que cuanto mayor sea h mayor será el suavizado y viceversa.

En nuestro caso $h = 5$ y $l = |\mathbb{S}| = 2m - 1$ (rango de saltos) y K = función de densidad Gaussiana (Normal). Sean $R_X = \mathbb{S}$ y $W = R_{Z|X} = \{v_{r_k}\}$ donde $X = \{r_k\}$ son las rachas de tamaños de saltos de longitud k y $v_{r_k} = (\alpha_{r_k, p_1}, \dots, \alpha_{r_k, p_{2m-1}})$ vector de frecuencias de la matriz S^* asociado

a la racha r_k , y donde $\alpha_{r_k,i}$ viene determinado en la ecuación 4.1.

Con el estimador de Nadaraya-Watson se hace una primera aproximación de esta función, tomando como ancho de banda 5, obteniendo los valores suavizados.

$$\hat{v}_{r_k} = (s_1, \dots, s_{2m-1})$$

y normalizando el vector resulta:

$$\hat{v}'_{r_k} = \left(\frac{s_1}{\sum s_j}, \dots, \frac{s_{2m-1}}{\sum s_j} \right)$$

Estos son los pesos suavizados y normalizados. Notar que tanto el suavizado como el normalizado de la matriz S^* se hace tomando como referencia las rachas. Para el suavizado de la matriz S^* se utiliza el paquete `KernSmooth` para R (Wand, 1995) donde viene ya implementada la técnica Kernel descrita en este apartado. Véase a continuación un ejemplo práctico de aplicación de la técnica de suavizado descrita en este apartado.

Ejemplo de aplicación del suavizado

Siguiendo con el ejemplo de los matriculados de Alabama, donde se calcula la matriz de frecuencias de saltos (cuadro 4.2). Considérese, como ejemplo, las rachas de longitud dos $r_2 = (0, \cdot)$, que corresponde a la cuarta fila de la matriz, con vector de frecuencias $v_0 = (0, 0, 2, 0, 1, 2, 1)$ y que tiene por gráfico (figura 4.1).

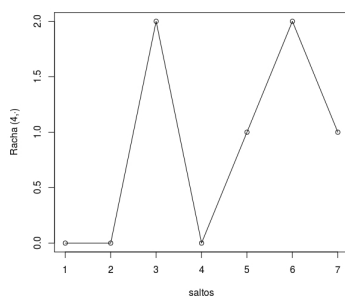


Figura 4.1: Gráfica de v_0

Suavizando la matriz de frecuencias (cuadro 4.2) aplicando la técnica de suavizado de la regresión Kernel, explicada anteriormente, da como resultado una matriz de frecuencias suavizadas cuadro 4.3.

	-3	-2	-1	0	1	2	3
-3	0.03	0.07	0.13	0.20	0.24	0.23	0.20
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-1	0.45	0.39	0.34	0.29	0.23	0.16	0.10
0	0.45	0.60	0.76	0.91	1.05	1.16	1.23
1	0.39	0.54	0.65	0.65	0.54	0.38	0.23
2	0.25	0.44	0.69	0.97	1.19	1.32	1.36
3	0.10	0.18	0.27	0.36	0.41	0.42	0.40

Cuadro 4.3: Matriz suavizada de frecuencias de saltos asociada a rachas de longitud 1

Y el vector asociado a la racha de longitud dos con un salto cero queda $\hat{v}_0 = (0.45, 0.60, 0.76, 0.91, 1.05, 1.16, 1.23)$ con gráfico figura 4.2.

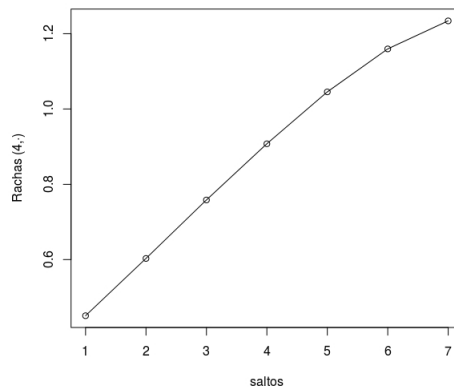


Figura 4.2: Gráfica de v_0 suavizado

En la figura 4.2 se puede observar el vector suavizado asociado a las rachas de longitud dos asociado al salto cero que normalizado resulta $\hat{v}_0 = (0.07, 0.10, 0.12, 0.15, 0.17, 0.19, 0.20)$.

Por lo tanto se utiliza la matriz de frecuencias suavizada y normalizada como ponderación en el procedimiento de predicción difusa con patrones.

Siguiendo con el ejemplo, el último salto observado es -1 , por lo tanto para hacer la predicción difusa se usa el vector ya normalizado $\hat{v}'_{-1} = (0.23, 0.20, 0.17, 0.15, 0.12, 0.08, 0.05)$

y como la penúltima observación es $F(1991) = A_{19}$, se obtiene $F(1992) = \hat{A}_{1992} = 0.23 * A_{16} \oplus 0.20 * A_{17} \oplus 0.17 * A_{18} \oplus 0.15 * A_{19} \oplus 0.12 * A_{20} \oplus 0.08 * A_{21} \oplus 0.05 * A_{22}$, obteniendo una estimación puntual de 18844.58, el valor observado en 1992 es 18876, obteniendo un error de 31.41

Obsérvese que el conjunto A_{22} no está definido, pero el procedimiento permite generar predicciones fuera del universo, en este ejemplo quedaría $A_{22} = (20132.88, 20520.16, 193.64, 193.64)$. Por lo tanto el algoritmo de predicción difusa para patrones quedaría como se muestra en la figura 4.3.

Ante la variedad de algoritmos propuestos y utilizando la minería de datos, a continuación se propone un sistema de ayuda a la decisión para discriminar entre los distintos algoritmos dependiendo del tamaño del conjunto de entrenamiento.

4.2. Sistema de ayuda a la decisión (SAD)

Tanto en el ejemplo de los matriculados de la universidad de Alabama como en diversos experimentos numéricos, se analizan las carencias de cada una de las propuestas presentadas en esta memoria, donde se aprecia que el número de datos observados influye en gran manera en la precisión de la predicción. Es por ello que se plantea discriminar entre los diferentes procedimientos propuestos tomando como referencia el número de observaciones.

Resumiendo las aportaciones propuestas en esta memoria, se presentan los siguientes modelos de ponderación:

- Cronológico: se pondera sobre las relaciones difusas de la serie temporal y se utilizan dichos pesos en las relaciones difusas de grupos.
- Tendencia: se definen a partir de la serie temporal difusa, como la diferencia de los subíndices de dos conjuntos difusos consecutivos, llamado tamaño de salto. Y se construye la matriz de pesos a partir de las recurrencias de los tamaños de salto.
- Patrones: a partir de la serie temporal de saltos inducida por la STD y sobre el análisis de rachas (secuencias de saltos) en la serie temporal de saltos, se construye la matriz de pesos a partir de las recurrencias de las rachas.

Los tres modelos de predicción introducidos en esta tesis permiten proponer una regla de predicción general basada en todos ellos. Denótese O_C , O_T y O_P

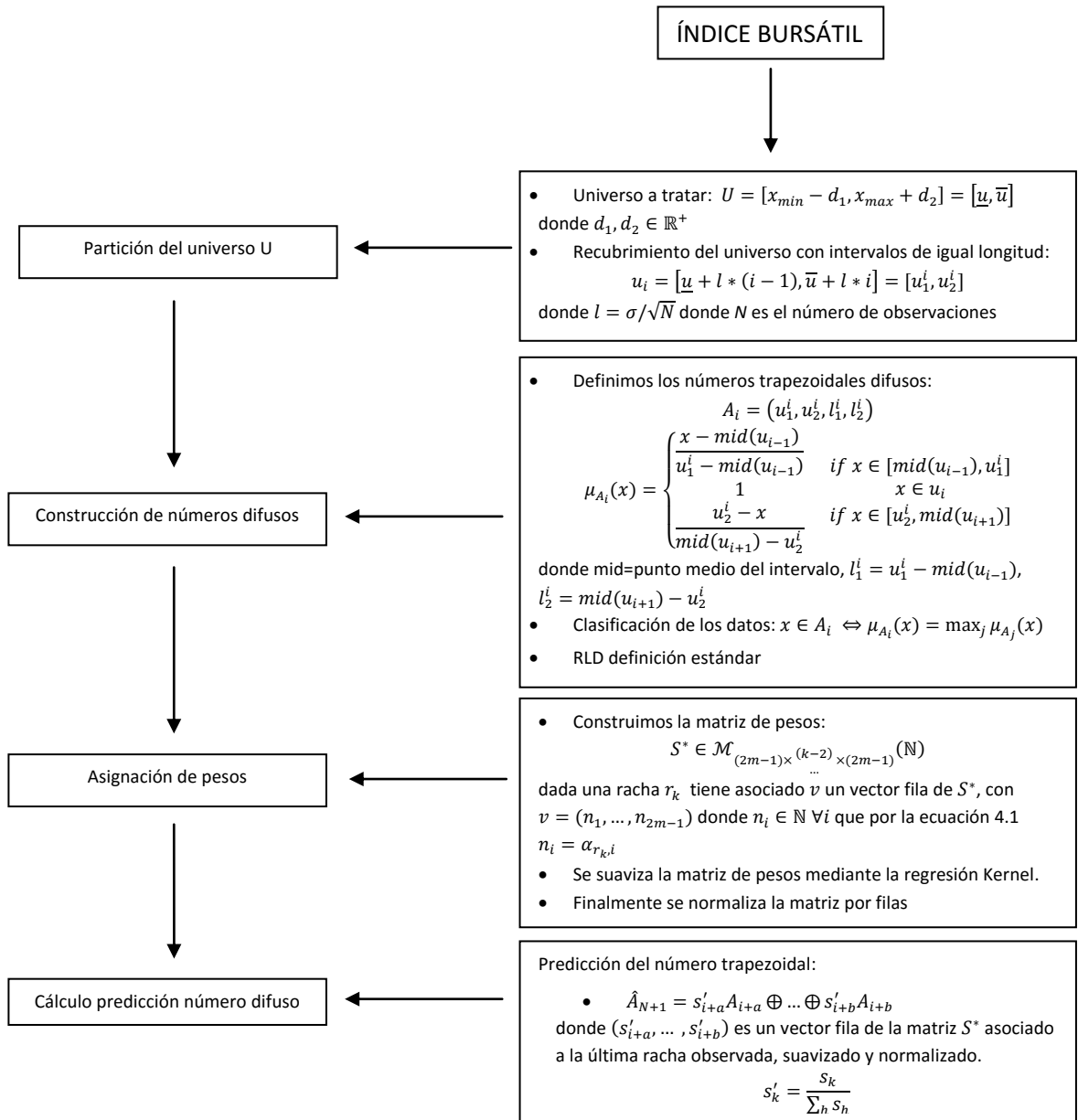


Figura 4.3: Esquema de flujo del Algoritmo 4

como los modelos cronológicos, de tendencia y de patrón, respectivamente. Dada una serie temporal observada, y una vez que se obtuvieron las series

temporales difusas asociadas, se propone utilizar el siguiente predictor:

$$\hat{F}(N+1) = \gamma_1 O_C(F(1), \dots, F(N)) + \gamma_2 O_T(F(1), \dots, F(N)) + \gamma_3 O_P(F(1), \dots, F(N))$$

donde $\{\gamma_1, \gamma_2, \gamma_3\}$ son coeficientes no negativos tales que $\gamma_1 + \gamma_2 + \gamma_3 = 1$, y $\hat{F}(N+1)$ es un predictor difuso, que da como resultado una predicción que es un número difuso trapezoidal del que también es posible obtener una predicción puntual, así como otras medidas de la incertidumbre. Específicamente, ese número difuso trapezoidal permite calcular el valor esperado y la ambigüedad de la predicción a partir del concepto del intervalo posibilístico de medias.

En capítulos anteriores, se han estudiado y analizado los dos primeros modelos (O_C y O_T) y una combinación lineal de ambos, cuyo peso se evaluó utilizando los errores de ajuste obtenidos a partir del conjunto de datos de entrenamiento. Ahora, se está interesado en el rendimiento del modelo de patrones y la precisión de las predicciones obtenidas al trabajar con él; sin embargo, este modelo necesita una gran cantidad de observaciones para aprender y construir estructuras de patrones adecuadas. Esta es la principal limitación de la aplicabilidad del nuevo modelo, aunque el tamaño del conjunto de datos también juega un papel importante cuando se utilizan los otros modelos.

Por lo tanto basándose en los resultados proporcionados por los experimentos numéricos desarrollados utilizando series temporales de diferentes tamaños, se propone tener en cuenta el tamaño N de la serie en el esquema SAD, de tal manera que:

- Si $N < 30$, se usa el modelo cronológico ($\gamma_1 = 1$).
- Si $30 \leq N < 250$, se añade el modelo de tendencia ($\gamma_1 + \gamma_2 = 1$).
- Si $250 \leq N < 500$, se añade el modelo de patrones con rachas de longitud $k = 2$ ($\gamma_1 + \gamma_2 + \gamma_3 = 1$).

Se podrían usar otras rachas de mayor longitud si N es mayor que 500.

4.3. Experimentos numéricos

Para analizar el rendimiento predictivo del sistema de apoyo a la decisión, se utilizan dos índices de referencia, el stock alemán (DAX30) y el índice taiwanés (TAIEX), entre los años 2013 y 2015. Toda la información sobre las cotizaciones diarias de estos índices se obtienen de www.finance.yahoo.com. Las bases de datos utilizadas para estos experimentos son las mismas que se usan para los experimentos numéricos del DAX30 y del TAIEX del capítulo anterior.

El rendimiento de nuestro procedimiento SAD se compara con los métodos ponderados de Yu (2005b) y Cheng et al. (2008a) y Wang et al. (2013), y con los modelos estadísticos de predicción clásicos medias móviles (MM1), suavizado exponencial (*SE*) y *ARIMA*. Se debe tener en cuenta que los métodos de STD ponderadas mencionados utilizan diferentes estrategias para dividir el universo de los datos. Ya se ha comprobado en capítulos anteriores los rendimientos de los modelos cronológicos (capítulo 2) y de tendencia (capítulo 3), es por ello que se tomará un número de datos lo suficientemente grande para trabajar con el modelo de patrón, adicionalmente se tomará $\gamma_1 = \gamma_2 = 0$ y $\gamma_3 = 1$ para analizar la calidad de las predicciones de dicho modelo.

Los conjuntos de datos de las cotizaciones diarias se dividieron dinámicamente, en un esquema de horizonte móvil a una etapa, en dos partes: el conjunto de datos de entrenamiento y el conjunto de datos de contraste (para las predicciones de una etapa). Se utilizan las observaciones del conjunto de datos de entrenamiento para calcular la predicción a una etapa, luego se añade la primera observación de los datos de contraste a los datos de entrenamiento y de éste se desprenderá del primer dato que se observó, debiéndose calcular los nuevos pesos para otra nueva predicción. La calidad de la predicción se mide por medio de la desviación media absoluta (MAD).

Este esquema de horizonte móvil a una etapa, se utiliza por igual para todos los métodos y modelos de predicción empleados en el experimento numérico, bien sean difusos (propios y ajenos) o estadísticos.

A lo largo de este capítulo en los experimentos numéricos, con tal de reducir el número de citas a los diferentes autores y que no se creen excesivas repeticiones de las citas, se denota: Yu (2005b) como Yu, Cheng et al. (2008a) como Cheng y Wang et al. (2013) como Wang.

4.3.1. Predicción del índice DAX30

Para el primer experimento, se usan las cotizaciones diarias del índice del stock alemán DAX30 durante los años 2013 y 2014.

Las primeras trescientas cotizaciones sirven como conjunto de datos de entrenamiento inicial, después se obtiene una predicción de la siguiente forma: los procedimientos (difusos o no) construyen una predicción a una etapa para el primer día de los datos de contraste e incorpora posteriormente la cotización diaria observada al conjunto de datos de entrenamiento, así el procedimiento funciona en un esquema de horizonte móvil a una etapa, de tal manera que la primera observación de enero de 2013 se elimina. Para este nuevo conjunto de datos de entrenamiento, los procedimientos generan nuevas predicciones y aplican la misma estrategia para la próxima predicción posterior. Este esquema de predicción diario se repite cien veces, obteniendo una cantidad total de cien predicciones diarias.

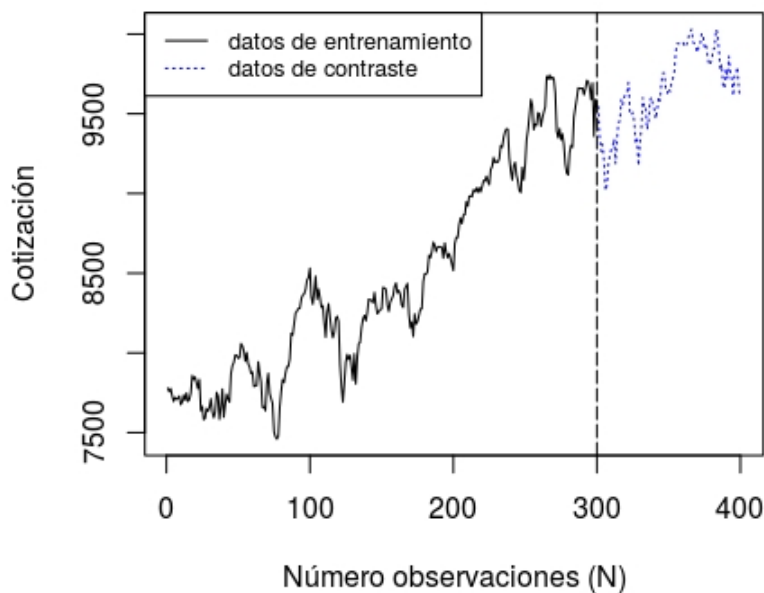


Figura 4.4: Trazo de las 400 primeras observaciones de la serie temporal de las cotizaciones diarias del índice alemán de DAX30, entre los años 2011 y 2013.

La figura 4.4 muestra en el lado izquierdo las trescientas primeras cotizaciones que corresponde al conjunto de datos de entrenamiento inicial y en el lado derecho está el trazo de los datos de contraste.

Notar que nuestro enfoque supone que cada valor observado pertenece a un número difuso trapezoidal, y cada predicción de una etapa es también un número difuso trapezoidal, para el cual su valor esperado se considera la predicción puntual. Esto se muestra en la figura 4.5, donde se incluye una representación parcial de la serie temporal difusa de 20 cotizaciones del conjunto de entrenamiento y la primera predicción difusa.

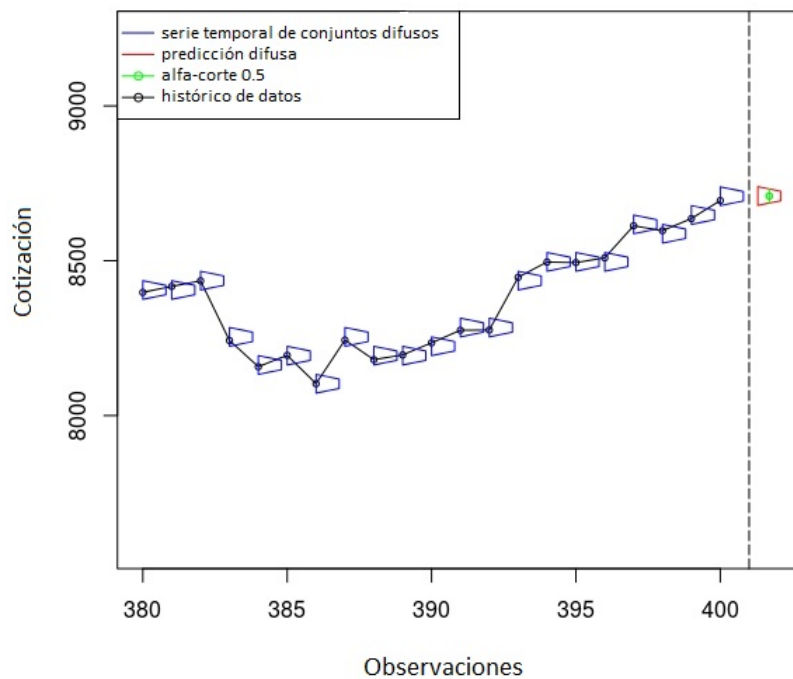


Figura 4.5: Representación parcial de la serie temporal difusa de 20 cotizaciones diarias del índice DAX30 alemán y la predicción difusa a una etapa.

Para este experimento, como el tamaño de la muestra es 300, el método de predicción seleccionado que usa el esquema SAD es un modelo de patrón ($\gamma_1, \gamma_2 = 0, \gamma_3 = 1$) con rachas de longitud 2, este corresponde al Algoritmo 4. Por una parte la precisión del Algoritmo 4 propuesto se compara con las ob-

tenidas mediante la aplicación de Yu, Cheng, Wang. Posteriormente también se compara el procedimientos con métodos estadísticos de predicción clásicos.

El cuadro 4.4 muestra los resultados obtenidos de los métodos de series temporales difusas, obsérvese que el Algoritmo 4 propuesto consigue una mejor precisión en la predicción del MAD.

	Yu	Cheng	Wang	Algoritmo 4
MAD	78.92	122.21	195.46	65.70

Cuadro 4.4: Errores de predicción del índice DAX30, con conjuntos de datos de tamaño 300.

En el análisis estadístico, el cuadro 4.5 muestra los p-valores ajustados de la comparación estadística por pares del error de predicción al cuadrado, que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (teniendo en cuenta el problema de comparación múltiple).

	Yu	Cheng	Wang
Cheng	0.0002		
Wang	9.5e-08	6.8e-08	
Algoritmo.4	0.0193	7.7e-07	1.1e-08

Cuadro 4.5: P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos STD.

De análisis de los resultados de los p-valores del cuadro 4.5, se concluye que existen diferencias estadísticamente significativas del error de predicción al cuadrado, utilizando la comparación por pares de los errores de predicción puntual de una etapa del Algoritmo 4 con respecto a las predicciones obtenidas por los modelos de STD ponderadas.

Al igual que en el capítulo anterior, también se analizan los resultados comparándolos con modelos estadísticos de predicción clásicos.

	MM1	SE	ARIMA	Algoritmo 4
MAD	63.17	63.27	62.58	65.70

Cuadro 4.6: Errores de predicción para modelos estadísticos y para el Algoritmo 4, con conjuntos de datos del DAX30 de tamaño 300.

Análogamente al análisis estadístico anterior, el cuadro 4.7, muestra los p-valores ajustados de la comparación estadística por pares de los errores de predicción al cuadrado, que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (tiene en cuenta el problema de comparación múltiple).

	MM1	SE	ARIMA
SE	0.57		
ARIMA	1.00	1.00	
Algoritmo 4	0.45	0.45	0.42

Cuadro 4.7: P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos estadísticos, junto con el Algoritmo 4.

Obsérvese el cuadro 4.6 y cómo todos los modelos estadísticos de predicción mejoran los errores de predicción, MAD, a los errores de predicción del Algoritmo 4. Aunque el análisis estadístico del cuadro 4.7 revela que no existen diferencias significativas entre los errores de predicción al cuadrado obtenidos por el Algoritmo 4 y el resto de modelos estadísticos. A raíz del análisis estadístico, tampoco se observan diferencias significativas entre los distintos modelos estadísticos clásicos de predicción.

4.3.2. Predicción del índice TAIEX

El segundo experimento trata con la cotización diaria de TAIEX entre los años 2013 a 2015. Se aplica el mismo esquema de horizonte móvil diario que en el primer experimento, aunque ahora el conjunto de datos de entrenamiento contiene 350 datos.

La figura 4.6 muestra en el lado izquierdo las primeras 350 cotizaciones, que corresponde al conjunto de entrenamiento inicial, y en el lado derecho está el trazado del conjunto de datos de contraste.

En el algoritmo que sigue el esquema SAD para este conjunto de datos es el modelo de patrón (usando $\gamma_3 = 1$) que trabaja con rachas de longitud 2, que corresponde al Algoritmo 4.

El cuadro 4.8 muestra los resultados obtenidos de los métodos de series temporales difusas, obsérvese que el Algoritmo 4 propuesto consigue una me-

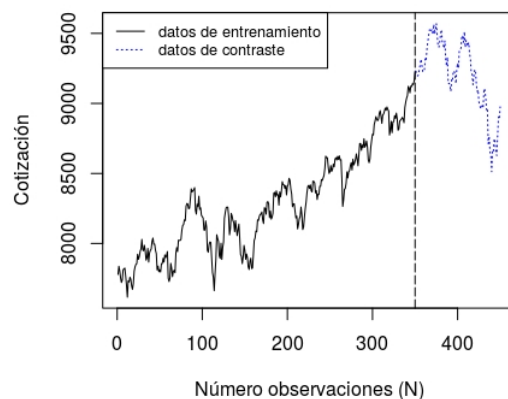


Figura 4.6: Trazo de la serie temporal de cotizaciones diarias del TAIEX.

por precisión en la predicción del MAD.

	Yu	Cheng	Wang	Algoritmo 4
MAD	59.31	88.33	124.20	52.21

Cuadro 4.8: Errores de predicción del índice TAIEX, con conjuntos de datos de tamaño 350.

El cuadro 4.9 muestra los p-valores ajustados de la comparación estadística por pares del error de predicción al cuadrado, que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (teniendo en cuenta el problema de comparación múltiple).

	Yu	Cheng	Wang
Cheng	0.00016		
Wang	3.7e-10	0.00020	
Algoritmo 4	0.00977	1.7e-06	1.4e-12

Cuadro 4.9: P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos STD.

Del análisis estadístico del cuadro 4.9 se desprende que existen diferencias estadísticamente significativas del error de predicción al cuadrado, utilizando la comparación por pares de los errores de predicción puntual de una etapa

con respecto a las predicciones obtenidas por el Algoritmo 4 y por los modelos de STD ponderadas.

Al igual que en el experimento anterior, también se analizan los resultados comparándolos con modelos estadísticos de predicción clásicos.

	MM1	SE	ARIMA	Algoritmo 4
MAD	52.60	52.59	52.40	52.21

Cuadro 4.10: Errores de predicción para modelos estadísticos y para el Algoritmo 4, con conjuntos de datos del TAIEX de tamaño 350.

Análogamente al análisis estadístico anterior, el cuadro 4.11 muestra los p-valores ajustados de la comparación estadística por pares de los errores de predicción al cuadrado, que se realiza a través del test-t por pares, ajustando los p-valores usando el método Holm (tiene en cuenta el problema de comparación múltiple).

	MM1	SE	ARIMA
SE	1		
ARIMA	1	1	
Algoritmo 4	1	1	1

Cuadro 4.11: P-valores ajustados de la comparación por pares de los errores de predicción al cuadrado obtenidos con diferentes modelos estadísticos, junto con el Algoritmo 4.

Según el cuadro 4.10 pese a que el error de predicción cometido del Algoritmo 4 es menor con respecto a los modelos estadísticos de predicción, del análisis estadístico del cuadro 4.11 revela que no existen diferencias significativas entre los errores de predicción al cuadrado obtenidos por los distintos modelos de predicción estudiados. Tampoco se observan diferencias significativas entre los distintos modelos estadísticos clásicos de predicción.

4.4. Conclusiones

En este capítulo, se propone un nuevo algoritmo, Algoritmo 4. Tomando como referencia el Algoritmo 3, el Algoritmo 4 ha añadido un nuevo procedimiento de minería de datos (patrones) basado en el concepto de rachas de tamaños de saltos. Por lo tanto en el modelo de patrones los pesos miden la frecuencia relativa de las rachas de tamaño de salto, entre saltos observados cronológicamente consecutivos. El Algoritmo 4 ha incorporado cambios en la algoritmia con respecto al Algoritmo 3, a saber, se ha utilizado el error estándar como medida para dividir el universo de los datos y se ha utilizado también técnicas de suavizado para mejorar los pesos obtenidos por la falta de observaciones en determinados patrones, adicionalmente también devuelve como resultado números difusos trapezoidales.

Finalmente en este capítulo también se ha definido un predictor (con parámetros $\gamma_1, \gamma_2, \gamma_3$), donde los parámetros del predictor se han determinado mediante un sistema de ayuda a la decisión (SAD); basado en el tamaño de la muestra, y que ayuda a discriminar en la elección de qué algoritmo utilizar en cada caso, de entre los distintos algoritmos propuestos a lo largo de esta memoria.

De los experimentos numéricos y las comparaciones con los métodos de predicción para series temporales difusas (Yu, Chen y Wang), con la nueva propuesta se continúa mejorando la calidad de las predicciones, siendo éstas estadísticamente significativas. En cuanto a las comparaciones con los modelos estadísticos clásicos de predicción, los resultados no han sido satisfactorios, bien sea por que no se mejora la calidad de la predicción de forma clara y notoria, o por que no han existido diferencias significativas en los análisis estadísticos.

Dado que parece que ha aumentado la complejidad de las formulaciones sin obtener resultados satisfactorios y como tampoco se ha podido proponer modelos de predicción difusa a varias etapas, ni se ha podido obtener predicciones por intervalos, en el siguiente capítulo se decide cambiar el enfoque de la tesis. Se verá el estado de la cuestión dentro de la teoría de conjuntos difusos desde un punto de vista credibilista, que devengará un modelo más robusto (en sentido estadístico), con una simplificación algorítmica notoria; manteniendo las predicciones difusas y la calidad de las predicciones puntuales, además se añaden predicciones difusas a varias etapas y predicciones por intervalos, todo con un coste computacional visiblemente reducido.

Predicción difusa credibilista

En este capítulo se propone un nuevo esquema de predicción para series temporales utilizando la lógica difusa. Obsérvese que normalmente un conjunto de datos determinado, como realización de una parte de una serie temporal, está bien definido y no hay incertidumbre en él. Es por ello que puede entenderse que donde se encuentra la incertidumbre es en los procesos subyacentes que hay entre las relaciones de los datos de la serie; así mediante la aplicación de la lógica difusa se pretende explicar dichos procesos ocultos y las relaciones existentes entre los datos de la serie temporal.

Hasta este capítulo se habían visto, dentro de la teoría de conjuntos difusos, aproximaciones desde un punto de vista posibilístico, ahora se verán aproximaciones desde un punto de vista credibilista, para ello se introducirán algunas definiciones nuevas como variable difusa y distribución de credibilidad (Liu y Liu, 2002).

En la práctica se construirán variables difusas que son aproximadas a partir de unos determinados cuantiles de las diferencias de los datos. Así se representará la incertidumbre sobre el comportamiento futuro de la serie temporal a través de dichas variables difusas. Trabajar con variables difusas permitirá hacer predicciones puntuales (valor esperado credibilista de la variable difusa), predicciones por intervalos (se construirán a partir de la distribución de credibilidad) y proponer modelos de predicción difusa a varias etapas. Obsérvese que en capítulos anteriores si que se han obtenido predicciones puntuales (aunque a una sola etapa), no obstante nunca se propuso ninguna estimación de predicción por intervalos.

Se analizarán los resultados obtenidos utilizando esta propuesta comparándolos con dos métodos estadísticos clásicos de predicción suavizado ex-

ponencial (SE) y *ARIMA*. No se hacen comparaciones de los resultados con otros métodos de predicción en series temporales difusas debido a que la mayor parte de métodos de este campo no se preocupan por dar predicciones por intervalos, se limitan a hacer predicciones puntuales. Adicionalmente, también se realizarán predicciones para varias etapas, aunque en capítulos anteriores sólo se habían propuesto procedimientos para predicciones a una etapa.

Por un lado, como datos de entrenamiento se utilizarán series temporales sin tendencia extraídas de la Competición **M4** (Makridakis et al., 2019) con estacionalidad igual a uno (en particular series temporales de datos anuales). En el análisis comparativo de los resultados no sólo se compararán las predicciones a una, dos, tres, cuatro, cinco y seis etapas sino también la predicción por intervalos y los tiempos de computación.

Este capítulo está estructurado como sigue: en primer lugar se introducirán las definiciones de variable difusa y distribución de credibilidad; en la segunda sección se mostrarán los modelos difusos de predicción; en la sección tres se continuará con un ejemplo explicativo del funcionamiento de dichos modelos, en la sección cuatro se mostrarán los resultados numéricos y el contraste estadístico, y finalmente en la última sección vendrán las conclusiones del capítulo.

5.1. Variable difusa y distribución de credibilidad

A continuación se introducirán definiciones que permitan describir los modelos de predicción propuestos en este capítulo. La posibilidad y la necesidad de cualquier evento difuso $\{T \geq x\}$, donde T es una variable difusa y $x \in \mathbb{R}$, se puede calcular mediante la correspondiente función de pertenencia (Zadeh, 1968):

$$Pos\{T \geq x\} = \sup_{t \geq x} \mu_T(t) \quad (5.1)$$

$$Nec\{T \geq x\} = 1 - \sup_{t < x} \mu_T(t) \quad (5.2)$$

Las siguientes definiciones han sido propuestas por Liu y Liu (2002) y Liu (2006):

Definición 5.1.

Se define credibilidad de un evento difuso $\{T \geq x\}$ como la siguiente semi-

suma:

$$Cr\{T \geq x\} = \frac{1}{2}Pos\{T \geq x\} + \frac{1}{2}Nec\{T \geq x\} \quad (5.3)$$

Proposición 5.2.

La credibilidad tiene la propiedad de ser auto-dual:

$$Cr\{T \geq x\} = 1 - Cr\{T < x\}$$

Definición 5.3.

Sea Θ un conjunto no vacío, y $\mathcal{P}(\Theta)$ el conjunto potencia de Θ (todos los subconjuntos de Θ).

La función Cr es una medida credibilista si satisface los siguientes axiomas:

1. $Cr\{\Theta\} = 1$
2. Cr es creciente, es decir, $Cr\{A\} \leq Cr\{B\}$ siempre que $A \subset B$

Definición 5.4.

Sea Θ un conjunto no vacío, $\mathcal{P}(\Theta)$ el conjunto potencia de Θ , y Cr una medida de credibilidad.

Entonces la terna $(\Theta, \mathcal{P}(\Theta), Cr)$ se llama espacio credibilístico.

Definición 5.5.

Una variable difusa ξ es una función que va de un espacio credibilístico $(\Theta, \mathcal{P}(\Theta), Cr)$ al conjunto de los números reales.

Definición 5.6.

La distribución de credibilidad $\Phi_\xi : \mathcal{R} \rightarrow [0, 1]$ de una variable difusa ξ se define como $\Phi_\xi(x) = Cr\{\theta \in \Theta | \xi(\theta) \leq x\}$

□

A lo largo de este capítulo, las variables difusas son del tipo LR potencia, donde las funciones L y R son funciones potencia genéricas, siendo estimadas mediante los cuantiles de las diferencias de datos cronológicamente consecutivos.

A continuación véase primero cómo, en esta memoria, se estiman las variables difusas del tipo LR potencia y segundo cómo, a partir de una variable difusa, se calcula un momento credibilístico y un intervalo de cubrimiento.

5.1.1. Estimación de la variable difusa

Este procedimiento para la construcción de un número difuso de tipo LR potencia se presentó en Vercher y Bermúdez (2013), y se usará para estimar la variable difusa.

Sean $\{y_t\}_{t=1}^N$ un conjunto de datos observados, de cardinal N y sea ξ la variable difusa a estimar.

1. Se calculan los cuantiles 0, 0.25, 0.5, 0.75, 1 de los elementos del conjunto de datos (denotados como q_{cuantil}).

Estos cuantiles se usan para construir el soporte de la variable difusa. En este capítulo el núcleo se reducirá a un solo punto que corresponde al cuantil 0.5, esto facilitará los cálculos y la comprensión de los modelos de predicción.

Alternativamente, se puede considerar que las observaciones más extremas son sucesos anómalos y se pueden excluir, así en vez de utilizar los cuantiles 0 y 1 (mínimo y máximo) también cabe la posibilidad de coger los cuantiles 0.03 y 0.97 como extremos de nuestra variable difusa, aunque esto podría afectar al tamaño de los intervalos de predicción.

2. Las funciones potencia teóricas se calculan de manera que:

$$L(x) = \left(\frac{q_{.50} - x}{q_{.50} - q_{.00}} \right)^\gamma \quad R(x) = \left(\frac{x - q_{.50}}{q_{1.0} - q_{.50}} \right)^\beta \quad (5.4)$$

exigiendo que γ y β sean tales que $L(q_{.25}) = 0.5$ y $R(q_{.75}) = 0.5$.

Nótese que dados los intervalos abiertos $I_1 =]0, q_{0.5}[$ y $I_2 =]q_{0.5}, 1[$ las funciones L y R son de clase \mathcal{C}^1 en I_1 y I_2 respectivamente, y además como para cualquier $a \in I_1$ se tiene $|L'(a)| \neq 0$ y para cualquier $b \in I_2$ se tiene $|R'(b)| \neq 0$, entonces se sabe por el teorema de la función inversa que L y R son localmente invertibles, es decir que las funciones L^{-1} y R^{-1} existen y que son de clase \mathcal{C}^1 .

Por lo tanto se tiene que:

$$\gamma = \frac{\log(0.5)}{\log\left(\frac{q_{.50} - q_{.25}}{q_{.50} - q_{.00}}\right)} \quad \beta = \frac{\log(0.5)}{\log\left(\frac{q_{.75} - q_{.50}}{q_{1.0} - q_{.50}}\right)} \quad (5.5)$$

A partir de las aproximaciones de las funciones L y R , se deduce que:

- La función de pertenencia de la variable difusa ξ es:

$$\mu_{\xi}(x) = \begin{cases} 0 & \text{si } x \leq q_{.00} \\ L(x) & \text{si } q_{.00} \leq x \leq q_{.50} \\ R(x) & \text{si } q_{.50} \leq x \leq q_{1.0} \\ 0 & \text{si } x \geq q_{1.0} \end{cases} \quad (5.6)$$

- La distribución de credibilidad de ξ es:

$$\Phi_{\xi}(x) = \begin{cases} 0 & \text{si } x \leq q_{.00} \\ 0.5\mu_{\xi}(x) & \text{si } q_{.00} \leq x \leq q_{.50} \\ 1 - 0.5\mu_{\xi}(x) & \text{si } q_{.50} \leq x \leq q_{1.0} \\ 1 & \text{si } x \geq q_{1.0} \end{cases} = \begin{cases} 0 & \text{si } x \leq q_{.00} \\ 0.5L(x) & \text{si } q_{.00} \leq x \leq q_{.50} \\ 1 - 0.5R(x) & \text{si } q_{.50} \leq x \leq q_{1.0} \\ 1 & \text{si } x \geq q_{1.0} \end{cases} \quad (5.7)$$

La variable difusa se denota como $\xi = (A, B, l_i, l_d, \gamma, \beta)$ donde $A = B = q_{.50}$, $l_i = q_{.50} - q_{.00}$, $l_d = q_{1.0} - q_{.50}$, donde l_i y l_d son las amplitudes izquierda y derecha de la variable difusa de tipo LR potencia, y γ, β parámetros de forma.

□

5.1.2. Media credibilista e intervalo de cubrimiento

La estimación puntual de una variable difusa se aproxima como el primer momento credibilístico que corresponde al valor esperado credibilístico de la variable difusa. En Liu y Liu (2002) se define el valor esperado de una variable difusa ξ como:

$$E(\xi) = \int_{-\infty}^{\infty} x\phi(x)dx = \int_0^{\infty} Cr\{\xi \geq x\}dx - \int_{-\infty}^0 Cr\{\xi \leq x\}dx$$

Para el caso a tratar de variables difusas obtenidas según el procedimiento propuesto del apartado 5.1.1, se tiene que $\xi = (q_{0.5}, q_{0.5}, l_i, l_d, \gamma, \beta)$ con

funciones potencia $L(x)$ y $R(x)$ definidas en (5.4) y donde $l_i = q_{0.5} - q_{0.0}$, $l_d = q_{1.0} - q_{0.5}$, $\int_0^1 R(t)dt = \frac{\beta}{1-\beta}$ y $\int_0^1 L(t)dt = \frac{\gamma}{1-\gamma}$.

En Vercher y Bermúdez (2015) prueban que el valor esperado de una variable difusa del tipo LR potencia, $\xi = (A, B, l_i, l_d, \gamma, \beta)$ es:

$$E(\xi) = 1 + \frac{(q_{1.0} - q_{0.5})\beta}{2(\beta + 1)} + \frac{(q_{0.5} - q_{0.0})\gamma}{2(\gamma + 1)} \quad (5.8)$$

donde γ y β vienen determinados en (5.5).

Por otra parte, los intervalos de cubrimiento se construirán a partir de la distribución de credibilidad (5.7) utilizando los cuantiles.

Así pues se calcula $I = [\underline{u}, \bar{u}]$ para un cubrimiento por ejemplo del 80%, esto es tomando los cuantiles 0.1 y 0.9, donde $\underline{u} = \Phi_\xi^{-1}(0.1)$ y $\bar{u} = \Phi_\xi^{-1}(0.9)$.

$$\underline{u} = -\left(\left(-\frac{0.1 - 0.5}{0.5}\right)^{1/\gamma}\right)(q_{.50} - q_{0.0}) - q_{.50} \quad (5.9)$$

$$\bar{u} = \left(\left(\frac{0.1 - 0.5}{0.5}\right)^{1/\beta}\right)(q_{1.0} - q_{.50}) + q_{1.0} \quad (5.10)$$

Hasta ahora se han introducido las principales definiciones de la teoría de conjuntos difusos desde un punto de vista credibilista y se propone un procedimiento para aproximar una variable difusa del tipo LR potencia, a continuación véase los modelos de predicción propuestos a partir de los conceptos y procedimientos previamente introducidos.

5.2. Modelos difusos de predicción

Dado que el objetivo principal en este capítulo es proponer una nueva modelización de la incertidumbre del comportamiento futuro de las series temporales, se introduce un nuevo esquema de predicción de series temporales usando la lógica difusa, este nuevo enfoque rompe claramente con todas las propuestas clásicas de series temporales difusa.

Se parte de la premisa de que los datos a estudiar están bien definidos y no se encuentra incertidumbre en ellos. Donde se plantea que se encuentra la incertidumbre es en los procesos subyacentes existentes entre las relaciones de los datos de la serie, expresándose estas relaciones mediante las diferencias de los datos. Y siendo dichas diferencias las que se consideran como realizaciones

de una variable difusa, y las que se utilizan para estimar la variable difusa. A continuación, se definen las diferencias de los datos de distintos órdenes y el conjunto de las diferencias:

Definición 5.7.

Sean $\{y_t\}_{t=1}^N$ los datos observados de una serie temporal $Y(t)$

Se define la k -diferencia o la diferencia entre las observaciones en el momento i de orden k como

$$dif_i^k = y_{i+k} - y_i \quad \forall i = 1, \dots, N - 1 \text{ y } k = 1, \dots, N - i \quad (5.11)$$

Definición 5.8.

Se define el conjunto k -diferencia como el conjunto de todas las diferencias de orden k , por lo tanto

$$d^k = \{dif_i^k \text{ para } i = 1, \dots, N - 1\} \quad (5.12)$$

Definición 5.9.

Se define D^k como la estimación de la variable difusa a partir del conjunto de las diferencias de orden k (d^k), aplicando el procedimiento descrito en el apartado 5.1.1.

De esta forma se representa la incertidumbre sobre el comportamiento futuro de las series temporales utilizando variables difusas que permiten construir tanto predicciones puntuales como por intervalos.

A continuación se presentan los modelos de predicción utilizando variables difusas:

Modelo PD1 (Predicción Difusa de 1 paso):

Sea $Y(t) = \{y_1, \dots, y_N\}$ el conjunto de datos de entrenamiento de tamaño N de la serie temporal donde y_N es la última observación de una serie y sea D^1 la variable difusa de las diferencias de orden 1.

Entonces la predicción difusa sería:

$$F(N + k) = \hat{y}_{N+k-1} + D^1 \quad (5.13)$$

donde

$$\hat{y}_{N+k-1} = \begin{cases} y_N & \text{si } k = 1 \\ E(F(N + k - 1)) & \text{si } k \geq 2 \end{cases}$$

y $E(\cdot)$ es el valor esperado credibilístico de la variable difusa $F(N + k - 1)$.

Modelo $PDxp$ (Predicción Difusa por pasos):

Sea D^k la variable difusa de las diferencias de orden k .
La predicción difusa para k pasos sería:

$$F(N + k) = y_N + D^k \quad \text{para } k \geq 1 \quad (5.14)$$

Modelos multiplicativos

Como se puede ver en (5.13) y (5.14) ambos modelos son aditivos. Adicionalmente, se presentan también las versiones multiplicativas de los modelos, que se obtienen aplicando los modelos aditivos a los logaritmos de los datos. El modelo $LPD1$ se obtiene aplicando el modelo $PD1$ sobre el logaritmo de los datos, donde d^1 será las primeras diferencias de logaritmos y a partir de estas primeras diferencias se estima la variable difusa D^1 (empleando el algoritmo de estimación descrito en 5.1.1); finalmente después de realizar las predicciones puntuales y por intervalos, se recuperan las unidades originales, mediante la función exponencial ' $exp(\cdot)$ '.

También se implementa el modelo $LPDxp$, aplicando modelo $PDxp$ sobre los logaritmos a los datos, de forma análoga a como se obtiene el modelo $LPD1$.

Otros modelos propuestos:

Finalmente se implementan las combinaciones convexas de los modelos aditivos y multiplicativos propuestos, con la semisuma de modelos, calculados como sigue:

$$SPD1 = \frac{PD1 + LPD1}{2}$$

$$SPDxp = \frac{PDxp + LPDxp}{2}$$

siendo $SPD1$ la semisuma del modelo aditivo y multiplicativo a un paso y $SPDxp$ la semisuma del modelo aditivo y multiplicativo a varios pasos.

En estudios anteriores (Singh, 2008; Stevenson y Porter, 2009) ya utilizaron las diferencias de datos cronológicamente consecutivos para crear modelos de predicción en series temporales difusas siguiendo el esquema básico propuesto por Chen (1996), aunque cabe resaltar que los modelos que se plantean en este capítulo rompen con dicho esquema.

Hasta ahora se han introducido los principales modelos de predicción difusa, véase a continuación un ejemplo de aplicación de cada uno de los modelos.

5.2.1. Predicción de una serie: un ejemplo

A continuación se verá cómo funcionan cada uno de los modelos principales expuestos, $PD1$ y $PDxp$, mediante un ejemplo, para ello se utilizará una serie temporal de datos anuales. Para los modelos $LPD1$, $SPD1$, $LPDxp$ y $SPDxp$ se procedería de forma análoga.

El cuadro 5.1 muestra un conjunto de datos extraídos de una serie temporal de datos anuales, que corresponde a la serie $Y1$ de la Competición **M4**, esta colección de series se comentará y utilizará en el apartado siguiente. El tamaño del conjunto de datos es de 31 observaciones, se tomarán como datos de entrenamiento las 28 primeras observaciones y como datos de contraste las 3 últimas observaciones, por tanto $k = 1, 2, 3$.

t	y_t	d^1	d^2	d^3	t	y_t	d^1	d^2	d^3
1	5172.10	-38.60	14.80	-87.50	15	6427.40	227.50	408.00	498.10
2	5133.50	53.40	-48.90	48.50	16	6654.90	180.50	270.60	418.60
3	5186.90	-102.30	-4.90	227.40	17	6835.40	90.10	238.10	308.60
4	5084.60	97.40	329.70	491.60	18	6925.50	148.00	218.50	305.10
5	5182.00	232.30	394.20	570.90	19	7073.50	70.50	157.10	276.10
6	5414.30	161.90	338.60	540.90	20	7144.00	86.60	205.60	195.20
7	5576.20	176.70	379.00	511.60	21	7230.60	119.00	108.60	20.20
8	5752.90	202.30	334.90	486.00	22	7349.60	-10.40	-98.80	-55.00
9	5955.20	132.60	283.70	362.00	23	7339.20	-88.40	-44.60	54.70
10	6087.80	151.10	229.40	174.90	24	7250.80	43.80	143.10	310.10
11	6238.90	78.30	23.80	122.10	25	7294.60	99.30	266.30	356.80
12	6317.20	-54.50	43.80	110.20	26	7393.90	167.00	257.50	
13	6262.70	98.30	164.70	392.20	27	7560.90	90.50		
14	6361.00	66.40	293.90	474.40	28	7651.40			

Cuadro 5.1: Serie de datos anuales y diferencias de distintos órdenes

El cuadro 5.1 consta de cinco columnas, en la columna ' t ' viene recogida la cronología de las observaciones, en la columna ' y_t ' están las observaciones propiamente dichas, y en las columnas ' d^1 ', ' d^2 ' y ' d^3 ' son los conjuntos de k -diferencias (5.12), cuyos elementos se calculan como las k -diferencias definidas en (5.11) para $k = 1, 2, 3$ respectivamente.

La estimación de las variables difusas asociadas a k -diferencias de distintos órdenes, se realiza como sigue:

A partir del conjunto de las 1-diferencias $d^1 = \{dif_i^1 \text{ para } i = 1, \dots, 27\}$ (columna d^1 del cuadro 5.1) se construye la variable difusa D^1 siguiendo los pasos presentados anteriormente 5.1.1.

<i>nucleo</i>	l_i	l_d	γ	β
97.40	199.70	134.90	0.41	0.84

Cuadro 5.2: Parámetros de la variable difusa D^1

Donde el núcleo en nuestro caso es la mediana de los datos (el percentil 50), γ y β son las potencias calculadas de las funciones L y R a partir de las fórmulas (5.5), finalmente l_i y l_d son las amplitudes a izquierda y derecha de los cuantiles 0 y 1 con respecto del núcleo que, coincide con la mediana:

$$l_i = q_{.50} - q_{.00}$$

$$l_d = q_{1.0} - q_{.50}$$

La variable difusa es $D^1 = (nucleo, l_i, l_d, \gamma, \beta)$, quedando como:

$$D^1 = (97.40, 199.70, 134.90, 0.41, 0.84)$$

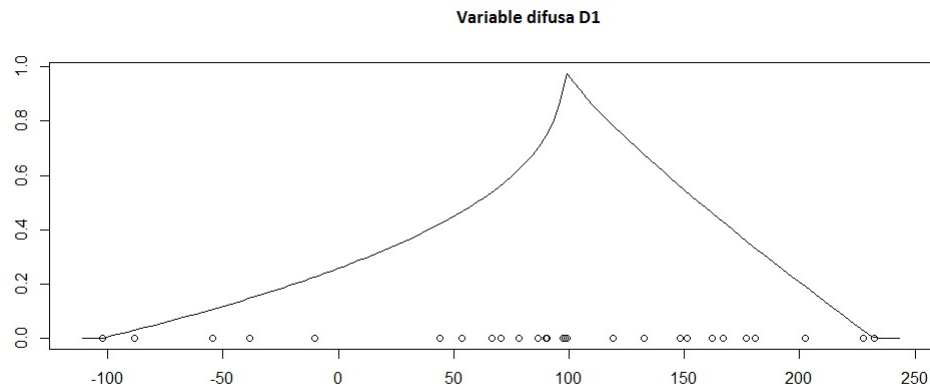


Figura 5.1: Gráfica de la variable difusa D^1

Obsérvese que en la figura 5.1 los círculos corresponden a las primeras diferencias de los datos observados (pertenecientes a la columna d^1 del cuadro 5.1).

Análogamente se obtienen las variables difusas D^2 y D^3 a partir de las diferencias de orden dos (d^2) y tres (d^3) respectivamente, teniendo la siguiente representación gráfica:

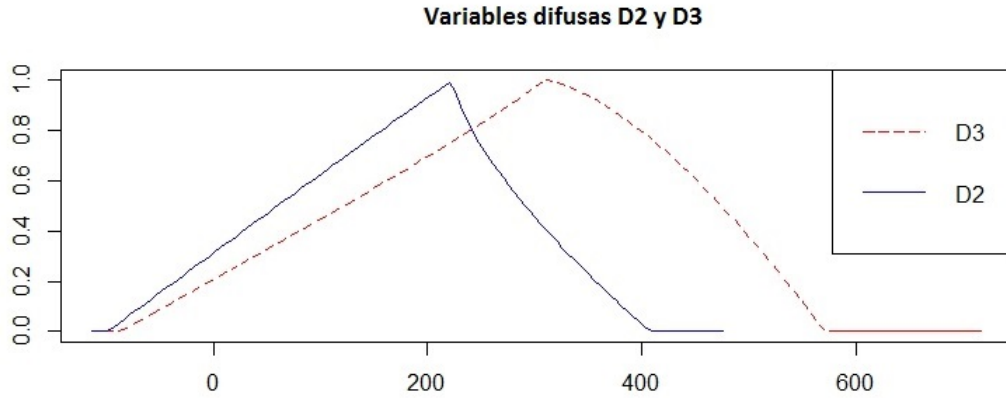


Figura 5.2: Variables difusas de los conjuntos 2-diferencias y 3-diferencias

Véase a continuación cómo funciona el modelo difuso $PD1$:

La última observación del conjunto de entrenamiento es $N = 28$, a partir de (5.13) se tiene que:

1. En la etapa 1,

$$F(29) = y_{28} + D^1$$

donde $F(29)$ es una variable difusa trasladada, y_{28} la última observación y D^1 la variable difusa de las diferencias de orden 1.

2. En la etapa 2

$$F(30) = E(F(29)) + D^1 = y_{28} + E(D^1) + D^1$$

donde $F(30)$ es una variable difusa trasladada y $E(D^1)$ el valor esperado credibilístico de la variable difusa D^1 .

3. En la etapa 3

$$F(31) = E(F(30)) + D^1 = y_{28} + 2E(D^1) + D^1$$

donde $F(31)$ es una variable difusa trasladada.

Nótese que como $E(a + \xi) = a + E(\xi)$ se tiene que:

$$E(F(29)) = y_{28} + E(D^1)$$

$$E(F(30)) = y_{28} + 2E(D^1)$$

$$E(F(31)) = y_{28} + 3E(D^1)$$

Así, en la **etapa 1** se tiene que:

$$F(29) = 7651.40 + D^1 = (7748.80, 199.70, 134.90, 0.41, 0.84)$$

donde la predicción puntual es el valor esperado calculado a partir de (5.8): $E(F(29)) = 7750.33$ y la predicción por intervalos se calcula a partir de la distribución de credibilidad (ecuaciones 5.9 y 5.10):

intervalos de predicción credibilista al 80 % $\rightarrow [7632.24, 7860.65]$

intervalos de predicción credibilista al 95 % $\rightarrow [7572.35, 7878.01]$

En la **etapa 2** se tiene que:

$$F(30) = 7750.33 + D^1 = (7847.73, 199.70, 134.90, 0.41, 0.84)$$

donde la predicción puntual (valor esperado) es: $E(F(30)) = 7849.26$ y la predicción por intervalos es calculada a partir de la distribución de credibilidad:

intervalos de predicción credibilista al 80 % $\rightarrow [7731.17, 7959.58]$

intervalos de predicción credibilista al 95 % $\rightarrow [7671.28, 7976.94]$

Finalmente en la **etapa 3** se tiene que:

$$F(31) = 7849.26 + D^1 = (7946.66, 199.70, 134.90, 0.41, 0.84)$$

donde la predicción puntual (valor esperado) es: $E(F(31)) = 7948.20$ y la predicción por intervalos es calculada a partir de la distribución de credibilidad:

intervalos de predicción credibilista al 80 % $\rightarrow [7830.10, 8058.51]$

intervalos de predicción credibilista al 95 % $\rightarrow [7770.21, 8075.88]$

Nótese que como $q_\gamma(a + \xi) = a + q_\gamma(\xi)$ se tiene que:

$$IC(F(29)) = y_{28} + IC(D^1)$$

$$IC(F(29)) = y_{28} + E(D^1) + IC(D^1)$$

$$IC(F(29)) = y_{28} + 2E(D^1) + IC(D^1)$$

donde IC es el intervalo de predicción credibilista.

Ahora véase como funciona el modelo difuso por pasos $PDxp$
La etapa 1 es análoga al modelo $PD1$:

$$F(29) = y_{28} + D^1$$

En la etapa 2 ya se observa el primer cambio respecto del modelo $PD1$, utilizando las diferencias de orden 2:

$$F(30) = y_{28} + D^2$$

Y finalmente en la etapa 3, se utilizan las diferencias de orden 3:

$$F(31) = y_{28} + D^3$$

Como ya se comentó la primera etapa es común a los dos modelos.
Se tiene que la variable difusa es $D^1 = (97.40, 199.70, 134.90, 0.41, 0.84)$

Así en la **etapa 1** se obtiene:

$$F(29) = 7651.40 + D^1 = (7748.80, 199.70, 134.90, 0.41, 0.84)$$

donde la predicción puntual (valor esperado) es: $E(F(29)) = 7750.33$ y la predicción por intervalos se calcula a partir de la distribución de credibilidad:

intervalos de predicción credibilista al 80 % $\rightarrow [7632.24, 7860.65]$

intervalos de predicción credibilista al 95 % $\rightarrow [7572.35, 7878.01]$

Ahora en la **etapa 2**, se calcula la variable difusa D^2 a partir del conjunto de 2-diferencias d^2 (columna del cuadro 5.1):

$$D^2 = (223.95, 322.75, 184.05, 1.02, 0.69)$$

Y la predicción difusa quedará como sigue

$$F(30) = 7651.40 + D^2 = (7875.35, 322.75, 184.05, 1.02, 0.69)$$

Lo que se está haciendo es utilizar las 2-diferencias para estimar una variable difusa con la que se hacen predicciones a dos etapas. Así la predicción puntual (valor esperado) es: $E(F(30)) = 7831.30$ y la predicción por intervalos se calcula a partir de la distribución de credibilidad:

intervalos de predicción credibilista al 80 % $\rightarrow [7615.83, 8033.14]$

intervalos de predicción credibilista al 95 % \rightarrow [7568.38, 8053.00]

La **etapa 3** se procederá de forma análoga a la etapa anterior, pero partiendo del conjunto de las 3-diferencias d^3 :

$$D^3 = (308.60, 396.10, 262.3, 0.92, 1.51)$$

$$F(31) = 7651.40 + D^3 = (7960, 199.70, 134.90, 0.41, 0.84)$$

Al igual que en la etapa anterior lo que se está haciendo es utilizar las 3-diferencias para estimar una variable difusa con la que se hacen predicciones a tres etapas. Así la predicción puntual (valor esperado) es: $E(F(31)) = 7944.01$ y la predicción por intervalos se calcula a partir de la distribución de credibilidad:

intervalos de predicción credibilista al 80 % \rightarrow [7649.19, 8147.22]

intervalos de predicción credibilista al 95 % \rightarrow [7585.38, 8202.74]

En este apartado se ha visto un ejemplo de aplicación de los principales modelos de predicción difusa propuestos. Para analizar la calidad de las predicciones véase a continuación un experimento numérico en el cual se comparan los resultados obtenidos con los proporcionados al aplicar modelos estadísticos clásicos de predicción (SE y *ARIMA*).

5.3. Experimentos numéricos

Se utiliza la base de datos de series temporales de la Competición **M4** del año 2018 para el contraste y comparación de los modelos propuestos con los modelos clásicos de predicción suavizado exponencial (SE) y *ARIMA*. Estas competiciones son eventos muy importantes en la comunidad de investigadores especializados en series temporales desde 1982. Es un desafío para el cual los investigadores desarrollan y aplican modelos de predicción a grandes conjuntos de series temporales de datos reales.

En la Competición **M4** hay cien mil series. Se seleccionan para este experimento series con datos anuales (ya que tienen estacionalidad 1), en total 23000 series. Debido a la sensibilidad de los modelos ante el número de observaciones, se aplica un filtro a las series anuales para obtener series de tamaño $N > 30$; con esta acción se asegura que las series tienen un tamaño mínimo ya que esta condición es necesaria para el cálculo de las variables difusas.

Después de aplicar dicho filtro se obtiene una base de datos que consta de 9060 series. Tanto los datos de entrenamientos como los datos de contraste son los proporcionados por dicha competición. La base de datos de entrenamiento está compuesta por 9060 series de distintos tamaños (de entre treinta observaciones hasta casi setecientas observaciones), y los datos de contraste son de tamaño homogéneo para todas las series y constan de seis observaciones a predecir. Resumiendo, cada una de las series es de la siguiente forma: sin pérdida de generalidad sea N el tamaño del conjunto de datos de una de las series $Y(t)$, entonces se subdividen los datos de entrenamiento $\{y_1, \dots, y_N\}$ y los datos de contraste $\{y_{N+k}\}$ para $k = 1, \dots, 6$. La Competición **M4** proporciona solamente las series de entrenamiento pero, una vez concluida, también hace públicas las series de contraste. Todas ellas están en el paquete de R: *M4comp2018*.

Objetivos en este experimento: modelar la incertidumbre de los datos desde un punto de vista credibilista mediante variables difusas; obtener predicciones puntuales desde una a seis etapas, esto se debe a que en la Competición **M4** se requieren hacer predicciones de hasta 6 etapas en las series anuales; calcular los intervalos de predicción usando la distribución de credibilidad de las variables difusas calculadas; medir la precisión de la predicción de los métodos PD propuestos con medidas del error de predicción habituales y estimar el coste computacional de los modelos estudiados.

Principalmente se van a comparar cinco modelos: dos clásicos (suavizado exponencial -SE- y *ARIMA*) y tres modelos difusos. Estos son: el modelo *PD1*, el modelo multiplicativo de *PD1* denotado como *LPD1*; y finalmente también se comparará con el modelo por pasos *PDxp*.

El experimento va a consistir en hacer predicciones puntuales con origen N a seis etapas $1 \leq k \leq 6$, donde la predicción puntual \hat{y} corresponderá al valor esperado credibilístico de la variable difusa $F(N+k)$ condicionado por toda la información disponible hasta el momento N del conjunto de datos $\{y_1, \dots, y_N\}$. También se comparará la predicción por intervalos así como los tiempos de computación.

Para cada modelo $M_j = \{SE, ARIMA, PD1, LPD1, PDxp\}$ y cada serie $Y_i(t)$ con $1 \leq i \leq 9060$, se calculan los errores como sigue:

$$e_{i,M_j}(t) = \hat{y}_{M_j(t,i)} - y_{(t,i)} \quad t = 1, \dots, 6$$

siendo $\hat{y}_{M_j(t,i)}$ la predicción en la etapa t de la serie obtenida mediante el modelo M_j y donde $y_{(t,i)}$ es el valor observado de los datos de contraste en

el instante t . Para comparar la calidad de la predicción, se consideran los siguientes errores medios de predicción:

El porcentaje de error medio del valor absoluto (MAPE):

$$MAPE_{M_j}(t) = 100 * \frac{1}{9060} \sum_{i=1}^{9060} \left| \frac{e_{i,M_j}(t)}{y_{(t,i)}} \right| \quad t = 1, \dots, 6$$

El porcentaje de error medio del valor absoluto simétrico (SMAPE):

$$SMAPE_{M_j}(t) = 100 * \frac{2}{9060} \sum_{i=1}^{9060} \frac{|e_{i,M_j}(t)|}{|\hat{y}_{M_j(t,i)} + y_{(t,i)}|} \quad t = 1, \dots, 6$$

A lo largo de esta memoria para los criterios de contraste y evaluación puntual de los modelos se habían utilizado los errores RMSE, MAD y BIAS. Pero, en este experimento debido a la gran heterogeneidad de las series se hace necesario trabajar con errores relativos, que no tengan en cuenta las unidades de los datos de cada serie temporal, es por ello que se utilizan los errores MAPE y SMAPE.

Para la predicción por intervalos se comparará el cubrimiento empírico, i.e., el porcentaje de veces que el valor a observar está comprendido dentro del intervalo de predicción. Finalmente también se calcularán tiempos de computación.

Resumiendo, se realiza un experimento con series de datos anuales de la Competición **M4** con filtro de tamaño $N_i > 30$, obteniéndose un total de 9060 series, realizándose predicciones de una a seis etapas y donde los errores de predicción utilizados son el MAPE y el SMAPE. En los diferentes cuadros donde se muestran los resultados obtenidos, se denota la etapa i -ésima como $E.i$ con $i = 1, \dots, 6$.

5.3.1. Predicción Puntual

En el cuadro 5.3 vienen recogidos, para los distintos modelos, los errores de predicción del MAPE y SMAPE de las distintas etapas y en el cuadro 5.4 sus medias (de las seis etapas). El cuadro 5.3 se organiza cronológicamente de la primera etapa a la sexta etapa, ordenados de izquierda a derecha y de arriba a abajo.

	<i>SE</i>	<i>ARIMA</i>	<i>PD1</i>	<i>LPD1</i>	<i>PDxp</i>		<i>SE</i>	<i>ARIMA</i>	<i>PD1</i>	<i>LPD1</i>	<i>PDxp</i>
E.1						E.2					
MAPE	7.70	8.14	7.55	8.86	7.55	10.80	11.38	10.26	13.06	10.08	
SMAPE	7.36	7.70	7.31	8.34	7.31	9.89	10.25	9.63	11.58	9.37	
E.3						E.4					
MAPE	13.44	13.77	12.64	16.55	12.14	16.48	16.70	15.67	21.30	14.86	
SMAPE	12.10	12.10	11.69	14.08	11.17	14.30	14.12	13.95	17.20	13.16	
E.5						E.6					
MAPE	18.80	18.60	17.63	25.39	16.53	22.05	22.02	20.93	31.99	19.67	
SMAPE	16.38	16.02	15.82	20.27	14.89	18.08	17.52	17.62	23.63	16.63	

Cuadro 5.3: Errores de predicción varias etapas

MEDIAS	<i>SE</i>	<i>ARIMA</i>	<i>PD1</i>	<i>LPD1</i>	<i>PDxp</i>
MAPE	14.88	15.10	14.11	19.53	13.47
SMAPE	13.02	12.95	12.67	15.85	12.09

Cuadro 5.4: Media de errores de predicción para las 6 etapas

Nótese que para el experimento planteado, los errores de predicción (MAPE y SMAPE) para todas las etapas que se presentan en los cuadros 5.3 y 5.4 el modelo *PDxp* es el que mejor funciona, concretamente mejor que los modelos clásicos de predicción del suavizado exponencial (*SE*) y *ARIMA*. El modelo multiplicativo propuesto es el que obtiene peores resultados.

Se lleva a término un contraste estadístico que estudia si existen diferencias significativas entre los errores MAPE de predicción de los diferentes modelos, los análisis comparativos de los errores del SMAPE arrojan resultados análogos. Se utiliza un test no paramétrico para muestras emparejadas, el test de Wilcoxon, que permite ver si existen, o no, diferencias significativas entre las medianas de dos muestras. También se ajustan los p-valores para comparaciones múltiples utilizando el método Holm. Un análisis paramétrico con el test-t produce resultados similares.

En el cuadro 5.5 vienen recogidas los p-valores de las seis etapas para la comparación del error MAPE. En la primera matriz, la submatriz inferior se compone por los p-valores de la primera etapa, mientras que en la submatriz superior se compone por los p-valores de la segunda etapa. En la segunda matriz, la submatriz inferior se compone por los p-valores de la etapa tres, mientras que la submatriz superior se compone de los p-valores de la etapa cuatro. Y finalmente en la tercera matriz, la submatriz inferior se compone por los p-valores de la etapa 5, mientras que la submatriz superior se compone de los p-valores de la etapa 6.

E.1/E.2	<i>SE</i>	<i>ARIMA</i>	<i>PD1</i>	<i>LPD1</i>	<i>PDxp</i>
<i>SE</i>		0.32	< 2e-16	< 2e-16	< 2e-16
<i>ARIMA</i>	0.005		<2e-16	<2e-16	<2e-16
<i>PD1</i>	7.2e-16	< 2e-16		< 2e-16	7e-07
<i>LPD1</i>	< 2e-16	< 2e-16	< 2e-16		< 2e-16
<i>PDxp</i>	2.2e-15	< 2e-16	1.00	< 2e-16	
E.3/E.4	<i>SE</i>	<i>ARIMA</i>	<i>PD1</i>	<i>LPD1</i>	<i>PDxp</i>
<i>SE</i>		9.7e-06	2.3e-16	< 2e-16	< 2e-16
<i>ARIMA</i>	7.3e-07		1.0e-09	< 2e-16	< 2e-16
<i>PD1</i>	7.2e-16	< 2e-16		< 2e-16	0.00038
<i>LPD1</i>	< 2e-16	< 2e-16	< 2e-16		< 2e-16
<i>PDxp</i>	< 2e-16	< 2e-16	4e-04	< 2e-16	
E.5/E.6	<i>SE</i>	<i>ARIMA</i>	<i>PD1</i>	<i>LPD1</i>	<i>PDxp</i>
<i>SE</i>		2.2e-06	4.3e-16	< 2e-16	< 2e-16
<i>ARIMA</i>	9.7e-06		8.0e-08	< 2e-16	4.0e-15
<i>PD1</i>	< 2e-16	1.0e-09		< 2e-16	0.012
<i>LPD1</i>	< 2e-16	< 2e-16	< 2e-16		< 2e-16
<i>PDxp</i>	< 2e-16	0.00085	0.00081	< 2e-16	

Cuadro 5.5: P-valores del test Wilcoxon dos a dos ajustados con el método de Holm del MAPE, para las etapas 1, 2, 3, 4, 5 y 6

Nótese que la primera etapa es común para los modelos *PD1* y *PDxp* (son exactamente iguales), por lo que es trivial el valor del p-valor. En media de SMAPE y MAPE, el modelo *PDxp* es el que mejor calidad de predicción obtiene, concretamente es menor que el suavizado exponencial (*SE*) y que el *ARIMA*, que a su vez tienen un error inferior que el modelo *LPD1*. En las comparaciones múltiples del error MAPE a todos los horizontes de predicción (etapas del uno al seis) se encuentran diferencias significativas entre el modelo *PDxp* y el resto de modelos (en particular los modelos estadísticos *SE* y *ARIMA*), cabe resaltar el hecho que los errores de predicción en media del modelo *PDxp* son inferiores en todas las etapas con respecto a los modelos estadísticos de predicción.

Entre los modelos estadísticos de predicción, el modelo *ARIMA* pese a tener un error superior en la predicción en las primeras etapas con respecto del *SE*, no se encuentran diferencias significativas en la segunda etapa y en las últimas etapas cuando ya existen diferencias significativas (entre *SE* y *ARIMA*), siendo la calidad de la predicción sensiblemente mejor con el modelo *ARIMA*.

El modelo que peor calidad de predicción obtiene es el *LPD1*, siendo las diferencias con el resto de modelos estadísticamente significativas.

5.3.2. Predicción Intervalar (cubrimiento)

En el cuadro 5.6 vienen recogidos los porcentajes de los cubrimientos que corresponden a los datos de contraste que pertenecen a los intervalos provistos por las predicciones por intervalos, siendo la primera matriz el cubrimiento al 80 % y la segunda matriz el cubrimiento al 95 %. La última columna de ambas matrices corresponde a la media aritmética de todas la etapas de los porcentajes de cubrimiento.

80 %	E.1	E.2	E.3	E.4	E.5	E.6	MEDIA
<i>SE</i>	0.65	0.72	0.76	0.76	0.77	0.78	0.74
<i>ARIMA</i>	0.51	0.53	0.56	0.56	0.56	0.57	0.55
<i>PD1</i>	0.66	0.79	0.89	0.92	0.93	0.94	0.86
<i>LPD1</i>	0.71	0.81	0.87	0.90	0.89	0.91	0.85
<i>PDxp</i>	0.66	0.80	0.86	0.83	0.78	0.73	0.78
95 %							
<i>SE</i>	0.81	0.87	0.89	0.89	0.88	0.89	0.87
<i>ARIMA</i>	0.49	0.56	0.65	0.68	0.69	0.72	0.63
<i>PD1</i>	0.74	0.86	0.93	0.95	0.96	0.96	0.90
<i>LPD1</i>	0.80	0.89	0.95	0.96	0.96	0.97	0.92
<i>PDxp</i>	0.74	0.87	0.93	0.91	0.86	0.83	0.86

Cuadro 5.6: Cubrimiento varias etapas en tantos por uno.

A la vista de los resultados, los modelos *SE* y *PDxp*, tiene en media un comportamiento similar. Obsérvese que en el cuadro 5.6 los modelos propuestos *PD1* y derivados, para el caso del cubrimiento al 95 % funcionan sensiblemente mejor en media que el suavizado exponencial y el modelo difuso *PDxp*, siendo el modelo semisuma el que mejor funciona para dicho cubrimiento ya que es el modelo que más se ajusta al porcentaje de cubrimiento dado, el modelo con menor porcentaje de cubrimiento en media es el *ARIMA*. En el caso del cubrimiento al 80 % no se observan mejoras significativas con respecto al suavizado exponencial y el modelo difuso *PDxp*, de nuevo el modelo *ARIMA* es el que menor porcentaje de cubrimiento obtiene.

Nótese que el suavizado exponencial tiene, para todos los cubrimientos (80 % y 95 %) y en todas las etapas, porcentajes de cubrimientos menores al

porcentaje de cubrimiento dado, mientras que los modelos propuestos a un paso ($PD1, LPD1$) en este capítulo las primeras etapas los porcentajes son menores que el cubrimiento dado y en las últimas etapas son mayores que el cubrimiento dado, notorio son los porcentajes tan bajos obtenidos con el modelo $ARIMA$ que en todo momento son muy inferiores al resto de modelos, en este sentido el modelo que peor funciona es el $ARIMA$. El modelo $PDxp$ en el cubrimiento al 80%, empieza con un porcentaje bajo, luego en la etapa 2, da exactamente como porcentaje del cubrimiento empírico (80%) el cubrimiento teórico esperado, y en las etapas sucesivas oscila alrededor del cubrimiento teórico. Finalmente para el caso del cubrimiento teórico al 95% el modelo $PDxp$ siembre da porcentajes inferiores, siendo el valor más próximo (obtenido en la etapa 3) del 93%.

5.3.3. Coste Computacional

En cuanto al coste computacional se calcula en segundos y por ejecución. Una ejecución se entiende como el tiempo transcurrido desde que se le pasan los datos de entrenamiento al algoritmo hasta que devuelven las seis predicciones. En el cuadro 5.7 vienen recogidos los estadísticos descriptivos que permitirán comparar la eficiencia computacional de los distintos modelos.

milisegundos	SE	$ARIMA$	$PD1$	$LPD1$	$PDxp$
MEDIA	10.80	80.91	0.75	0.78	2.61
DES. TÍPICA	27.81	53.60	17.51	2.51	2.07
TOTAL	97876.75	733054.9	6788.27	7104.61	23605.61

Cuadro 5.7: Estadísticos descriptivos de los tiempos de computación para las 9060 series

En el coste computacional de los modelos descritos en el cuadro 5.7 existe una notable diferencia con respecto al suavizado exponencial y al modelo $ARIMA$, teniendo los modelos clásicos estadísticos de predicción tiempos mucho mayores, siendo el modelo $ARIMA$ el que peor funciona en términos de coste computacional. Entre los modelos $PD1$ y $LPD1$ no hay grandes diferencias en cuanto al coste computacional (como era de esperar). En términos computacionales el modelo $PDxp$ funciona más lento que los modelos $PD1$ y $LPD1$, sin embargo sigue funcionando mucho mejor que los métodos clásicos de predicción estadística.

5.3.4. Resultados de los otros modelos propuestos

A continuación se presentan los resultados obtenidos con los otros modelos difusos de predicción propuestos:

En el cuadro 5.8 se presentan los errores de predicción de la primera etapa hasta la sexta etapa y en el cuadro 5.9 se presentan las medias de las seis etapas, para los modelos de los modelos $SPD1$, $LPDxp$ y $SPDxp$.

$SPD1$	E1	E2	E3	E4	E5	E6
MAPE	8.03	11.27	13.75	17.21	19.91	24.46
SMAPE	7.65	10.21	12.08	14.36	16.58	28.88
<hr/>						
<i>LPDxp</i>						
MAPE	8.86	12.90	16.01	20.46	24.34	30.34
SMAPE	8.34	11.44	13.76	16.71	19.53	22.75
<hr/>						
<i>SPDxp</i>						
MAPE	8.03	11.10	13.27	16.45	18.89	23.10
SMAPE	7.65	10.04	11.71	13.82	15.82	18.04

Cuadro 5.8: Errores de predicción varias etapas

MEDIAS	$SPD1$	$LPDxp$	$SPDxp$
MAPE	15.77	38.81	15.14
SMAPE	13.29	41.91	12.85

Cuadro 5.9: Media de errores de predicción para las 6 etapas

En el cuadro 5.8 se observa que en la primera etapa los modelos $SPD1$ y $SPDxp$ obtienen los mismos resultados, esto es debido a que en los modelos $PD1$ y $PDxp$ la primera etapa es común por lo que los modelos que derivan de ellos también tienen la primera etapa es común (tienen los mismos resultados). Tanto en media como en el resto de etapas el modelo que mejor funciona es el $SPDxp$.

En el cuadro 5.10 vienen recogidos los porcentajes de los cubrimientos de los modelos $SPD1$, $LPDxp$ y $SPDxp$. En el cuadro 5.10 se observan comportamientos similares a los descritos en el apartado anterior 5.3.2, así el modelo $SPD1$ va aumentando el porcentaje etapa a etapa, mientras que los modelos $LPDxp$ y $SPDsp$ en las tres primeras aumenta y en las tres últimas disminuye.

80 %	E.1	E.2	E.3	E.4	E.5	E.6	MEDIA
<i>SPD1</i>	0.71	0.82	0.91	0.94	0.94	0.95	0.88
<i>LPDxp</i>	0.71	0.75	0.71	0.66	0.63	0.59	0.68
<i>SPDxp</i>	0.71	0.78	0.77	0.70	0.66	0.61	0.71
95 %							
<i>SPD1</i>	0.79	0.90	0.96	0.97	0.98	0.98	0.93
<i>LPDxp</i>	0.80	0.84	0.83	0.77	0.72	0.67	0.77
<i>SPDxp</i>	0.79	0.86	0.87	0.81	0.75	0.70	0.80

Cuadro 5.10: Cubrimiento varias etapas en tantos por uno.

En el cuadro 5.11 se presentan el coste computacional de los modelos *SPD1*, *LPDxp* y *SPDxp*. Obsérvese que el modelo más rápido de los tres, es el *SPD1*.

milisegundos	<i>SPD1</i>	<i>LPDxp</i>	<i>SPDxp</i>
MEDIA	1.53	2.59	5.19
DES. TÍPICA	19.69	2.51	2.99
TOTAL	13892.87	23439.95	47045.55

Cuadro 5.11: Estadísticos descriptivos de los tiempos de computación para las 9060 series

5.4. Conclusiones

En este capítulo se ha presentado un nuevo esquema de predicción, basado en la lógica difusa y las distribuciones de credibilidad de eventos difusos. Partiendo de la definición de variable difusa y función de distribución credibilista, se han definido dos modelos aditivos *PD1* y *PDxp*, y de estos se han derivado cuatro modelos más: dos modelos multiplicativos y otros dos modelos por combinación lineal (semisuma). Finalmente de la aplicación de los modelos sobre el conjunto de datos anuales extraídos de la M4-competición, se han proporcionado predicciones puntuales y por intervalos, calculándose adicionalmente el coste computacional.

De los resultados obtenidos cabe destacar la mejora en predicción puntual del modelo *PDxp* con respecto a los errores del MAPE y SMAPE del suavizado exponencial y del *ARIMA*, a todos los horizontes de predicción, y que además como se han visto en el análisis posterior, muestra diferencias

estadísticamente significativas.

También cabe resaltar la mejora respecto al coste computacional de todos los modelos difusos con respecto a los modelos estadísticos, siendo los modelos de predicción difusa mucho más eficientes en términos computacionales. En cuanto a la predicción por intervalos, no se aprecian grandes diferencias de mejora con los porcentajes de cubrimiento entre los modelos propuestos y, el suavizado exponencial y el *ARIMA*.

Conclusiones y líneas de investigación futuras

Esta tesis ha cubierto amplia y satisfactoriamente los objetivos principales planteados en la introducción. Por un lado se han proporcionado distintos modelos de predicción difusa y, adicionalmente, se han extraído medidas predictivas difusas para utilizarlas en el problema de selección de carteras. También se ha trabajado dentro del marco de la teoría de conjuntos difusos, desde distintos puntos de vista, dando un enfoque posibilista en los primeros capítulos (del segundo al cuarto) y acabando la tesis con un enfoque credibilista (quinto capítulo). Y también se han podido obtener predicciones difusas a varias etapas y predicciones por intervalos.

En los modelos de predicción difusa (a una o varias etapas) presentados, en esta tesis, se ha comprobado general y extensamente su buen funcionamiento puntual, mejorando la calidad de las predicciones de otros modelos de predicción difusa y también con respecto de modelos de predicción estadísticos clásicos. La obtención de predicciones difusas como resultado de un proceso de predicción en las STD, ha permitido dar soporte a la toma de decisiones, dando información adicional que de otra manera no sería posible tener. También hacer mención del buen funcionamiento de las medidas predictoras difusas en el algoritmo heurístico aplicado al problema de selección de carteras, del capítulo 1.

En todo momento se ha querido poner en valor el hecho de que la predicción sea difusa, ya que la utilización los distintos momentos posibilistas o credibilistas nos permite extraer más información que la meramente puntual; poniendo en relieve que la mayoría de modelos de predicción en el campo de las STD, sólo se preocupan de la predicción puntual; de hecho en un quinto

capítulo se propone una predicción por intervalos. Se ha aportado a las matemáticas que envuelven las STD el concepto de tamaño de salto, para tratar de aproximar la tendencia de una serie temporal. Y se ha implementado una herramienta para un SAD que pretende automatizar y dar soporte a la toma de decisiones, y donde también se ha propuesto un predictor.

Algunas líneas de investigación futuras podrían ser:

Usar distintos momentos de la predicción difusa o la semi-desviación media absoluta en el problema de selección de carteras, ya que el capítulo 2 se limitó a aplicar el procedimiento de predicción difusa expuesto en dicho capítulo, sin hacerse comparaciones con otros algoritmos o aproximaciones. También desarrollar herramientas que trabajen con carteras a partir de las predicciones difusas.

Otra de las labores que quedó pendiente en el transcurso de esta tesis es definir los conceptos de tendencia o estacionalidad difusa y sus posibles implicaciones teóricas, así como trabajar con otro tipo de series temporales (con estacionalidad), ya que sólo se han trabajado con series con y sin tendencia. También tratar de proponer modelos de predicción difusa a varias etapas basados en procedimientos de series temporales difusas.

Otra investigación futura podría centrarse en la incorporación de reglas más sofisticadas para seleccionar a los diferentes operadores que trabajan en el SAD, que deben aplicarse para proporcionar predicciones más precisas para varios tipos de series temporales.

También cabe mejorar las predicciones por intervalos propuestas en el capítulo 5, ya que los resultados no han sido del todo satisfactorios y se tiene la convicción que su mejora es posible. Así como el estudio de las posibles implicaciones teóricas, que tendrían los modelos expuestos en dicho capítulo, dentro del marco de estudio de las series temporales difusas y la reinterpretación de la definición de serie temporal difusa.

También se podrían utilizar las predicciones difusas en otros problemas o modelos que requieran medidas predictoras alternativas o información adicional sobre el sistema a estudiar, y que se ajusten a las premisas que infiere la utilización de la lógica difusa.

Bibliografía

- Aladag, H., Basaran, A., Egrioglu, E., Yolcu, U., & Uslu, R. (2009). Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations. *Expert Systems with Applications*, 36(3):4228–4231.
- Arutchelvan, G., Srivatsa, S. K., & Jagannathan, R. (2010). Inaccuracy minimization by partitioning fuzzy data sets - validation of analytical methodology. *International Journal of Computer Science and Information Security*, 8(1):275–280.
- Askari, S. & Montazerin, N. (2015). A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering. *Expert Systems with Applications*, 42(4):2121 – 2135.
- Atashpaz-Gargari, E. & Lucas, C. (2007). Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition. In: *2007 IEEE Congress on Evolutionary Computation*, pages 4661–4667.
- Banzhaf, W. (2009). *Genetic programming: an introduction ; on the automatic evolution of computer programs and its applications*. Kaufmann.
- Bass, F. (1969). A new product growth model for consumer durables. *Management Science*, 15(5):215–227.
- Bellman, R. & Giertz, M. (1973). On the analytic formalism of the theory of fuzzy sets. *Information Sciences*, 5:149 – 156.
- Bellman, R. & Zadeh, L. (1970). Decision-making in a fuzzy environment. *Management science*, 17(4):B–141.
- Bermúdez, J., Segura, J., & Vercher, E. (2007a). Fuzzy portfolio optimization under downside risk measures. *Fuzzy Sets and Systems*, 158(7):769 – 782.

- Bermúdez, J., Segura, J., & Vercher, E. (2007b). A fuzzy ranking strategy for portfolio selection applied to the spanish stock market. *IEEE International Conference on Fuzzy Systems*, pages 3–6.
- Bermúdez, J., Segura, J., & Vercher, E. (2012). A multi-objective genetic algorithm for cardinality constrained fuzzy portfolio selection. *Fuzzy Sets and Systems*, 188(1):16 – 26. Theme: Decision and Optimisation.
- Bertsimas, D. & Shioda, R. (2009). Algorithm for cardinality-constrained quadratic optimization. *Comp. Opt. and Appl.*, 43:1–22.
- Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers.
- Box, G. & Jenkins, G. (1976). *Time series analysis: forecasting and control*. CA: Holdan-Day.
- Brockwell, P. & Davis, R. (1996). *Introduction to Time Series and Forecasting*. Springer.
- Brown, R. (1959). *Statistical forecasting for inventory control*. McGraw/Hill.
- Bulut, E. (2014). Modeling seasonality using the fuzzy integrated logical forecasting (filf) approach. *Expert Systems with Applications*, 41(4, Part 2):1806 – 1812.
- Campello, R. & Hruschka, E. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157(21):2858 – 2875.
- Carlsson, C., Fullér, R., & Majlender, P. (2002). A possibilistic approach to selecting portfolios with highest utility score. *Fuzzy Sets and Systems*, 131(1):13 – 21. Soft Decision Analysis.
- Chang, P.-T. (1997). Fuzzy seasonality forecasting. *Fuzzy Sets and Systems*, 90(1):1 – 10.
- Chang, T.-J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13):1271–1302.
- Chen, B.-T., Chen, M.-Y., Chiang, H.-S., & Chen, C.-C. (2011). *Forecasting Stock Price Based on Fuzzy Time-Series with Entropy-Based Discretization Partitioning*, pages 382–391. Springer Berlin Heidelberg.

- Chen, M.-Y. & Chen, B.-T. (2014). Online fuzzy time series analysis based on entropy discretization and a fast fourier transform. *Applied Soft Computing Journal*, 14:156–166.
- Chen, S.-M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems*, 81(3):311–319.
- Chen, S.-M. (2002). Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems*, 33(1):1–16.
- Chen, S.-M. & Chen, S. (2015). Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups and the probabilities of trends of fuzzy logical relationships. *IEEE Transactions on Cybernetics*, 45(3):391–403.
- Chen, S.-M. & Chung, N.-Y. (2006a). Forecasting enrollments of students by using fuzzy time series and genetic algorithms. *International journal of information and management sciences*, 17(3):1–17.
- Chen, S.-M. & Chung, N.-Y. (2006b). Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *International Journal of Intelligent Systems*, 21(5):485–501.
- Chen, S.-M. & Hsiao, H. (2006). A new method to estimate null values in relational database systems based on automatic clustering techniques. *International Journal of Intelligent Systems*, 21(5):485–5001.
- Chen, S.-M. & Hsu, C.-C. (2004). A new method to forecast enrollments using fuzzy time series. *International Journal of Applied Science and Engineering*, 3(2):234–244.
- Chen, S.-M. & Hwang, J.-R. (2000). Temperature prediction using fuzzy time series. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 30(2):263–275.
- Chen, T.-L., Cheng, C.-H., & Teoh, H. (2007). Fuzzy time-series based on fibonacci sequence for stock price forecasting. *Physica A: Statistical Mechanics and its Applications*, 380:377 – 390.
- Cheng, C., Chang, J.-R., & Yeh, C.-A. (2006). Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting {IT} project cost. *Technological Forecasting and Social Change*, 73(5):524 – 542.

- Cheng, C.-H. & Chen, C.-H. (2018). Fuzzy time series model based on weighted association rule for financial market forecasting. *Expert Systems*, 35(4):e12271. e12271 EXSY-Jul-17-163.R1.
- Cheng, C.-H., Chen, T.-L., Teoh, H.-J., & Chiang, C.-H. (2008a). Fuzzy time-series based on adaptive expectation model for taiex forecasting. *Expert Systems with Applications*, 34(2):1126–1132.
- Cheng, C.-H., Chen, Y.-S., & Wu, Y.-L. (2009). Forecasting innovation diffusion of products using trend-weighted fuzzy time-series model. *Expert Systems with Applications*, 36(2, Part 1):1826 – 1832.
- Cheng, C.-H., Cheng, G.-W., & Wang, J.-W. (2008b). Multi-attribute fuzzy time series method based on fuzzy clustering. *Expert Systems with Applications*, 34(2):1235 – 1242.
- Cheng, C.-H., Wang, J.-W., & Li, C.-H. (2008c). Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix. *Expert Systems with Applications*, 34(4):2568 – 2575.
- Chou, M.-T. (2017). An improved fuzzy time series theory with applications in the shanghai containerized freight index. *Journal of Marine Science and Technology (Taiwan)*, 25:393–398.
- Christensen, R. (1985). Entropy minimax multivariate statistical modeling–i: Theory. *International Journal of General Systems*, 11(3):231–277.
- Chu, H.-H., Chen, T.-L., Cheng, C.-H., & Huang, C.-C. (2009). Fuzzy dual-factor time-series for stock index forecasting. *Expert Systems with Applications*, 36(1):165–171.
- Daw, C., Finney, C., & Tracy, E. (2003). A review of symbolic analysis of experimental data. *Review of Scientific Instruments*, 74(2):915–930.
- Delgado, M., Vila, M., & Voxman, W. (1998). On a canonical representation of fuzzy numbers. *Fuzzy Sets and Systems*, 93(1):125 – 135.
- Dijkman, J., Haeringen, H., & de Lange, S. (1983). Fuzzy numbers. *Journal of Mathematical Analysis and Applications*, 92(2):301 – 341.
- Dubois, D. & Prade, H. (1980). Fuzzy sets and systems: theory and applications.
- Dubois, D. & Prade, H. (1987). The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24(3):279 – 300.

- Duru, O. & Bulut, E. (2014). A non-linear clustering method for fuzzy time series: Histogram damping partition under the optimized cluster paradox. *Applied Soft Computing*, 24:742 – 748.
- Efendi, R., Ismail, Z., & Deris, M. (2015). A new linguistic out-sample approach of fuzzy time series for daily forecasting of malaysian electricity load demand. *Applied Soft Computing*, 28:422 – 430.
- Egrioglu, E., Aladag, C., Yolcu, U., Basaran, M., & Uslu, V. (2009). A new hybrid approach based on sarima and partial high order bivariate fuzzy time series forecasting model. *Expert Systems with Applications*, 36(4):7424–7434.
- Elliot, R. (1994). *R.N. Elliott's Masterworks*. New Classics Library.
- Enayatifar, R., Sadaei, H., Abdullah, A., & Gani, A. (2013). Imperialist competitive algorithm combined with refined high-order weighted fuzzy time series (rhwfts-ica) for short term load forecasting. *Energy Conversion and Management*, 76:1104–1116.
- Fischer, R. & Fischer, J. (2001). *The New Fibonacci Trader: Tools and Strategies for Trading Success*. Wiley.
- Gardner, E. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1):1–28.
- Gath, I. & Geva, A. (1989). Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):773–780.
- Georgieva, P. & Popchev, I. (2013). Cardinality problem in portfolio selection. In: *Adaptive and Natural Computing Algorithms*, A. A. D. F. Tomassini, M. & P. Buesser, ed., pages 208–217.
- Gharavi, H. & Cheng, E. (2004). Information and communication technology in the stock broking industry An evolutionary approach to the diffusion of innovation. *Industrial Management & Data Systems*, 104:756–765.
- Goguen, J. (1967). L-fuzzy sets. *Journal of Mathematical Analysis and Applications*, 18(1):145 – 174.
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co and Inc.

- Gupta, P., Inuiguchi, M., & Mehlawat, M. K. (2011). A hybrid approach for constructing suitable and optimal portfolios. *Expert Systems with Applications*, 38(5):5620 – 5632.
- Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press. second edition, 1992.
- Holt, C. (1957). *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*, volume 52. ONR Memorandum.
- Huarng, K. (2001a). Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems*, 123(3):387–394.
- Huarng, K. (2001b). Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems*, 123:369–386.
- Huarng, K. & Yu, T.-K. (2006a). The application of neural networks to forecast fuzzy time series. *Physica A: Statistical Mechanics and its Applications*, 363(2):481 – 491.
- Huarng, K. & Yu, T.-K. (2006b). Ratio-based lengths of intervals to improve fuzzy time series forecasting. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(2):328–340.
- Huarng, K., Yu, T.-K., & Hsu, Y. (2007). A multivariate heuristic model for fuzzy time-series forecasting. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(4):836–846.
- Hwang, J.-R., Chen, S.-M., & Lee, C.-H. (1998). Handling forecasting problems using fuzzy time series. *Fuzzy Sets and Systems*, 100(1):217 – 228.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Cinelli, C., Khan, Y., Mayer, Z., Razbash, S., Schmidt, D., Shaub, D., Tang, Y., Wang, E., & Zhou, Z. (2015). *forecast: Forecasting functions for time series and linear models*. R package version 6.2.
- Hyndman, R. & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.
- Hyndman, R., Koehler, A., Snyder, R., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454.

- Indro, D.-C., Jiang, C.-X., Patuwo, B.-E., & Zhang, G.-P. (1999). Predicting mutual fund performance using artificial neural networks. *Omega*, 27(3):373 – 380.
- Inuiguchi, M. & Ramík, J. (2000). Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. *Fuzzy Sets and Systems*, 111(1):3 – 28.
- Jaynes, E. (1957). Information theory and statistical mechanics. *Physical Review Journal*, 106:620–630.
- Jilani, T. & Burney, S. (2008). Multivariate stochastic fuzzy forecasting models. *Expert Systems with Applications*, 35(3):691 – 700.
- Joan, C. (2009). *Operativa con candlestick para traders*. Noesis Analisis Financiero,S.L.
- Jorion, P. (1997). *Value at risk: the new benchmark for controlling derivatives risk*. McGraw-Hill.
- Kapur, J. (1989). *Maximum-entropy models in science and engineering*. New York: Wiley.
- Kemeny, J. & Snell, J. (1960). Finite markov chains. *Van Nostrand, Princeton, NJ*.
- Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. In: *Proceedings of the IEEE International Conference on Neural Networks*, pages 1942–1948.
- Kennedy, J. & Eberhart, R. (2001). *Swarm Intelligence*. Morgan Kaufmann Publishers Inc.
- Keogh, E. & Kasetty, S. (2003). On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4):349–371.
- Keogh, E., Lonardi, S., & Chiu, B.-C. (2002). Finding surprising patterns in a time series database in linear time and space. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 550–556. ACM.

- Kim, M., Min, S., & Han, I. (2006). An evolutionary approach to the combination of multiple classifiers to predict a stock price index. *Experts Systems with Applications*, 31:341–247.
- Kmenta, J. (1986). *Elements of Econometrics*. Macmillan.
- Konno, H. & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, 37(5):519–531.
- Kullback, S. & Leibler, R. (1951). On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86.
- Kuo, I.-H., Horng, S.-J., Kao, T.-W., Lin, T.-L., Lee, C.-L., & Pan, Y. (2009). An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. *Expert Systems with Applications*, 36(3 PART 2):6108–6117.
- Lee, L.-W., Wang, L.-H., Chen, S.-M., & Leu, Y.-H. (2006). Handling forecasting problems based on two-factors high-order fuzzy time series. *IEEE Transactions on Fuzzy Systems*, 14(3):468–477.
- Lee, M., Efendi, R., & Ismail, Z. (2009). Modified weighted for enrollment forecasting based on fuzzy time series. *Matematika*, 25(1):67–78.
- León, T., Liern, V., & Vercher, E. (2000). *Fuzzy Mathematical Programming for Portfolio Management*, pages 241–256. Physica-Verlag HD.
- Leu, Y., Lee, C.-P., & Jou, Y.-Z. (2009). A distance-based fuzzy time series model for exchange rates forecasting. *Expert Systems with Applications*, 36(4):8107 – 8114.
- Li, S. & Cheng, Y. (2007). Deterministic fuzzy time series model for forecasting enrollments. *Computers & Mathematics with Applications*, 53(12):1904 – 1920.
- Li, X., Qin, Z., & Kar, S. (2010). Mean-variance-skewness model for portfolio selection with fuzzy returns. *European Journal of Operational Research*, 202(1):239–247.
- Liu, B. (2006). A survey of credibility theory. *Fuzzy Optimization and Decision Making*, 5(4):387–408.

- Liu, B. & Liu, Y.-K. (2002). Expected value of fuzzy variable and fuzzy expected value models. *IEEE Transactions on Fuzzy Systems*, 10(4):445–450.
- Liu, H.-T. (2007). An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimization and Decision Making*, 6(1):63–80.
- Lu, W., Chen, X., Pedrycz, W., Liu, X., & Yang, J. (2015). Using interval information granules to improve forecasting in fuzzy time series. *International Journal of Approximate Reasoning*, 57:1–18.
- Mahajam, V., Muller, E., & Bass, F. (1990). New product diffusion models in marketing: A review and directions for research. *Journal Marketing*, 54(1):1–26.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2019). The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, page In press.
- Mandelbrot, B. & Hudson, R. (2004). *The (miss)Behavior of Markets*. Basic Books.
- Mansini, R. & Speranza, M. (1999). Heuristic algorithms for the portfolio selection problem with minimum transaction lots. *European Journal of Operational Research*, 114(2):219 – 233.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.
- Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). A proposal for the dartmouth summer research project on artificial intelligence. *AI Magazine*.
- McCulloch, W. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.
- Miller, G. (1956). *The magical number seven, plus or minus two: some limits on our capacity of processing information*, volume 63. The Psychological Review.
- Nadaraya, E. (1964). On estimating regression. *Theory of Probability and Its Applications*, 9(1):141–142.

- Nguyen, L. & Novák, V. (2019). Forecasting seasonal time series based on fuzzy techniques. *Fuzzy Sets and Systems*, 361:114 – 129. Theme: Clustering and Rule-based Models.
- Nison, S. (2001). *Japanese Candlestick Charting Techniques*. Prentice Hall Press.
- Panigrahi, S. & Behera, H. (2018). A computationally efficient method for high order fuzzy time series forecasting. *Journal of Theoretical and Applied Information Technology*, 96(21):7215–7226.
- Pathak, H. (2011). A new bandwidth interval based forecasting method for enrollments using fuzzy time series. *Applied Mathematics*, 02(04):504–507.
- Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*, 11(5):341–356.
- Pedrycz, W. & Vukovich, G. (2001). Abstraction and specialization of information granules. *Trans. Sys. Man Cyber. Part B*, 31(1):106–111.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Raynolds, J. & Mullin, L. (2005). Applications of conformal computing techniques to problems in computational physics: the fast fourier transform. *Computer Physics Communications*, 170(1):1 – 10.
- Ross, T. (1997). *Fuzzy Logic With Engineering Applications*. John Wiley & Sons Ltd.
- Rubio, A., Bermúdez, J., & Vercher, E. (2016). Forecasting portfolio returns using weighted fuzzy time series methods. *International Journal of Approximate Reasoning*, 75:1–12.
- Rubio, A., Bermúdez, J., & Vercher, E. (2017a). Improving stock index forecasts by using a new weighted fuzzy-trend time series method. *Expert Systems With Applications*, 76:12–20.
- Rubio, A., Bermúdez, J., & Vercher, E. (2017b). Introducing a fuzzy-pattern operator in fuzzy time series. In: *Advances in Computational Intelligence*, I. Rojas & A. C. Gonzalo, ed., pages 154–166. Springer International Publishing.
- Rutkowska, D. (2002). *Neuro-fuzzy Architectures and Hybrid Learning*. Physica-Verlag GmbH.

- Sadaei, H., Enayatifar, R., Lee, H., & Mahmud, M. (2016). A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting. *Applied Soft Computing Journal*, 40:132–149.
- Sadaei, H. & Lee, M. (2014). Multilayer stock forecasting model using fuzzy time series. *The Scientific World Journal*, 2014.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.
- Singh, P. & Dhiman, G. (2017). A fuzzy-lp approach in time series forecasting. In: *Pattern Recognition and Machine Intelligence*, B. U. Shankar, K. Ghosh, D. P. Mandal, S. S. Ray, D. Zhang, & S. K. Pal, ed., pages 243–253. Springer International Publishing.
- Singh, S. (2008). A computational method of forecasting based on fuzzy time series. *Math. Comput. Simul.*, 79(3):539–554.
- Smith, K. & Gupta, J. (2002). *Neural Networks in Business: Techniques and Applications*. PA: Idea Group Publishing. Hershey.
- Song, Q. (1999). Seasonal forecasting in fuzzy time series. *Fuzzy Sets and Systems*, 107(2):235 – 236.
- Song, Q. (2003). A note on fuzzy time series model selection with sample autocorrelation functions. *Cybernetics and Systemes*, 34(2):93–107.
- Song, Q. & Chissom, B. (1991). Forecasting enrollments with fuzzy time series part i. In: *The Annual Meeting Of the Mid-South Educational Research Association*.
- Song, Q. & Chissom, B. (1993a). Forecasting enrollments with fuzzy time series — part i. *Fuzzy Sets and Systems*, 54(1):1–9.
- Song, Q. & Chissom, B. (1993b). Fuzzy time series and its models. *Fuzzy Sets and Systems*, 54(3):269–277.
- Song, Q. & Chissom, B. (1994). Forecasting enrollments with fuzzy time series — part ii. *Fuzzy Sets and Systems*, 62(1):1–8.
- Song, Q. & Leland, R. (1996). Adaptive learning defuzzification techniques and applications. *Fuzzy Sets and Systems*, 81(3):321 – 329.

- Song, Q., R.P., L., & Chissom, B. (1995). A new fuzzy time-series model of fuzzy number observations. *Fuzzy Sets and Systems*, 73(3):341 – 348.
- Speranza, M. (1993). Linear programming model for portfolio optimization. *Finance*, 14:107–123.
- Stevenson, M. & Porter, J. (2009). Fuzzy time series forecasting using percentage change as the universe of discourse. *World Academy of Science, Engineering and Technology*, 55:154–157.
- Sullivan, J. & Woodall, W. (1994). A comparison of fuzzy forecasting and markov modeling. *Fuzzy Sets and Systems*, 64(3):279 – 293.
- Tanaka, H. & Guo, P. (1999). Portfolio selection based on upper and lower exponential possibility distributions. *European Journal of Operational Research*, 114(1):115 – 126.
- Teoh, H., Chen, T.-L., Cheng, C.-H., & Chu, H.-H. (2009). A hybrid multi-order fuzzy time series for forecasting stock markets. *Expert Systems with Applications*, 36(4):7888–7897.
- Ting, J. (2003). Casualties of the taiwan stock market. *Physica A*, 324:285–295.
- Tseng, F.-M., Tzeng, G.-H., & Yu, H.-C. (1999). Fuzzy seasonal time series for forecasting the production value of the mechanical industry in taiwan. *Technological Forecasting and Social Change*, 60(3):263 – 273.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.
- Vercher, E. & Bermúdez, J. (2013). A possibilistic mean-downside risk-skewness model for efficient portfolio selection. *IEEE Transactions on Fuzzy Systems*, 21(3):585–595.
- Vercher, E. & Bermúdez, J. (2015). Portfolio optimization using a credibility mean-absolute semi-deviation model. *Expert Systems with Applications*, 42(20):7121 – 7131.
- Vercher, E., Bermúdez, J., & Segura, J. (2007). Fuzzy portfolio optimization under downside risk measures. *Fuzzy Sets and Systems*, 158(7):769 – 782.
- Walczak, B. & Massart, D. (1999). Rough sets theory. *Chemometrics and Intelligent Laboratory Systems*, 47(1):1 – 16.

- Wand, M. (1995). *KernSmooth: Functions for kernel smoothing for Wand & Jones*.
- Wang, L., Liu, X., & Pedrycz, W. (2013). Effective intervals determined by information granules to improve forecasting in fuzzy time series. *Expert Systems with Applications*, 40(14):5673 – 5679.
- Wang, L., Liu, X., Pedrycz, W., & Shao, Y. (2014). Determination of temporal information granules to improve forecasting in fuzzy time series. *Expert Systems with Applications*, 41(6):3134–3142.
- Wang, L. & Mendel, J. (1992). Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427.
- Wang, N.-Y. & Chen, S.-M. (2009). Temperature prediction and {TAIFEX} forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series. *Expert Systems with Applications*, 36(2, Part 1):2143 – 2154.
- Watada, J. (1997). Fuzzy portfolio selection and its applications to decision making. *Tatra Mountains Mathematical Publication*, 13(4):219–248.
- Watson, G. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4):359–372.
- Weiss, S. & Kulikowski, C. (1991). *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. Morgan Kaufmann Publishers Inc.
- Widrow, B., Rumelhart, D., & Lehr, M. (1994). Neural networks: Applications in industry, business and science. *Commun. ACM*, 37(3):93–105.
- Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342.
- Wu, H., Long, H., & Jiang, J. (2019). Handling forecasting problems based on fuzzy time series model and model error learning. *Applied Soft Computing*, 78:109 – 118.
- Wu, W. (1986). Fuzzy reasoning and fuzzy relational equations. *Fuzzy Sets Syst.*, 20(1):67–78.

- Ye, F., Zhang, L., Zhang, D., Fujita, H., & Gong, Z. (2016). A novel forecasting method based on multi-order fuzzy time series and technical analysis. *Information Sciences*, 367:41 – 57.
- Yu, H.-K. (2005a). A refined fuzzy time-series model for forecasting. *Physica A: Statistical Mechanics and its Applications*, 346(3–4):657 – 681.
- Yu, H.-K. (2005b). Weighted fuzzy time series models for taiex forecasting. *Physica A: Statistical Mechanics and its Applications*, 349(3-4):609–624.
- Yu, T. & Huarng, K. (2008). A bivariate fuzzy time series model to forecast the taiex. *Expert Systems with Applications*, 34(4):2945–2952.
- Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3):338 – 353.
- Zadeh, L. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23(2):421 – 427.
- Zadeh, L. (1975a). The concept of a linguistic variable and its application to approximate reasoning-iii. *Information Sciences*, 9(1):43 – 80.
- Zadeh, L. (1975b). The concept of a linguistic variable and its application to approximate reasoning—i. *Information Sciences*, 8(3):199 – 249.
- Zadeh, L. (1975c). The concept of a linguistic variable and its application to approximate reasoning—ii. *Information Sciences*, 8(4):301 – 357.
- Zhang, G. (2004). Business forecasting with artificial neural networks. In: *Neural Networks in Business Forecasting*, pages 1–22. IGI Global.
- Zhang, L. & Li, Z.-F. (2012). Multi-period mean-variance portfolio selection with uncertain time horizon when returns are serially correlated. *Mathematical Problems in Engineering*, 2012.
- Zimmermann, H. (2001). *Fuzzy set theory and its application*. Boston: Kluwer Academic Publishers. fourth edition.