

UNIVERSIDAD LOYOLA ANDALUCÍA



TESIS DOCTORAL

“ANÁLISIS DE LA DEUDA SOBERANA
EN LA UE-15”

Doctorando/a: Carlos Galnares Jiménez-Placer

Directores: Mariano Carbonero Ruz,
Alfonso Carlos Martínez Estudillo y
Pilar Campoy Muñoz

Sevilla, 2023

UNIVERSIDAD LOYOLA ANDALUCIA



TESIS DOCTORAL

“ANÁLISIS DE LA DEUDA SOBERANA
EN LA UE-15”

DOCTORADO EN CIENCIA DE LOS DATOS

Doctorando: Carlos Galnares Jiménez-Placer

Directores: Mariano Carbonero Ruz,
Alfonso Carlos Martínez Estudillo y
Pilar Campoy Muñoz

Sevilla 2023



AGRADECIMIENTOS

A mi familia, y en especial a mi padre que no ha podido disfrutar del final de este proyecto, que tanto me ha apoyado a lo largo de estos años.

A todos los profesionales de la Universidad Loyola Andalucía que no solo me han apoyado con sus conocimientos y experiencia sino que también han conseguido animarme en los momentos de mayor flaqueza.

RESUMEN

La calificación del riesgo país mediante la emisión de ratings es de gran trascendencia tanto para los inversores como para los gestores públicos. El coste del servicio de la deuda puede verse afectado por una mala clasificación e incluso algunos inversores rechazan invertir en productos que tengan un rating inferior a determinado nivel.

Estas clasificaciones son realizadas por entidades privadas, conocidas como “agencias de rating”, que basan su negocio en cobrar sus servicios a los emisores. Adicionalmente también califican deuda para la que no han sido contratadas basándose en la información pública. Esta vinculación comercial sólo con algunas de las entidades clasificadas hace que se planteen dudas sobre su neutralidad e independencia.

La tardía reacción en la crisis de 2008, y en particular el caso de la clasificación de Grecia, hizo que los gestores públicos se preocuparan por la falta de transparencia de estas entidades y así se han emitido distintas disposiciones regulatorias tanto en Europa como en Estados Unidos para evitar esta opacidad. A pesar de esto al intervenir comités que asignan los ratings en función de información de diversa índole sus metodologías siguen siendo poco transparentes.

Es por ello que a lo largo de la historia se han realizado diversos intentos de estimar los ratings de deuda soberana con distinto nivel éxito. La revisión de la literatura pone de manifiesto que para estimar los ratings se han utilizado una gran cantidad de variables y de metodologías.

En este trabajo se propone seleccionar un conjunto limitado de variables partiendo de un modelo macroeconómico de dinámica de la deuda. Estas variables serán verificadas con distintos tipos de clasificadores, tanto de la estadística clásica como los propuestos por modernas técnicas de aprendizaje automático.

Para el estudio se han escogido los quince países que formaban la Unión Europea en el año 2000, durante un periodo que va desde el año 2002 al año 2017. De esta forma se consideran los años previos a la crisis financiera de 2008 como la posterior recuperación.

Además, partiendo de los análisis que demuestran una asimetría en los movimientos de rating se ha aplicado una transformación a la serie clasificada, denominada “filtrado”, que aproxima mejor las estimaciones de rating a las clasificaciones otorgadas por las agencias, capturando mejor su comportamiento.

Palabras clave: Ratings sobre deuda soberana, agencias de rating, UE-15, aprendizaje automático.

ABSTRACT

The sovereign ratings are a mayor issue to policy makers and investors. The service of the debt can be increased by a bad rating and some investor only buy assets with a minimum rating.

The ratings are determined by private entities known as credit rating agencies. These agencies are based on an issuer pay model. They also provide unsolicited ratings with the public data available. This commercial relation with only some of the issuers raise some doubts about their neutrality and independence.

The late reaction in the 2008 financial crisis, and in particular with the rating of Greece, made the policy makers to worry and different regulation has been issued to reduce the opacity. Despite this regulation the intervention in the rating assignation of ratings committees that asses different type of information the methodologies are still little transparent.

That's why many authors have try to estimate the sovereign rating with different success rates. The literature review shows that a great number of determinant and methodologies has been used.

In this work the variable selection is made through a macroeconomic framework of the debt's dynamics. These variables are tested with a variety of classification methodologies, some of which belongs to the classical statistics and other are proposed by state of the art machine learning methodologies.

The countries of the European Union in the year 2000 have been selected for this study. The period analyzed, from year 2002 to year 2017, includes the pre-crisis years and the recovery period.

Furthermore, based on the rating movement asymmetries, a transformation of the classified data has been made, known as “filtering”, that makes the performance of our classification better, like the real rating agencies behavior.

Key words: sovereign credit ratings, credit ratings agencies, UE-15, machine learning.

ÍNDICE GENERAL

CAPÍTULO 1 INTRODUCCIÓN	1
1. LA VALORACIÓN DEL RIESGO PAÍS MEDIANTE LAS CALIFICACIONES DE DEUDA SOBERANA.....	2
2. OBJETIVO DE INVESTIGACIÓN.....	8
3. CONTRIBUCION DE LA TESIS	10
4. ESTRUCTURA DE LA TESIS	11
5. DIFUSIÓN DE LA TESIS	13
CAPÍTULO 2 MODELO DE ESTIMACIÓN DE RATINGS DE DEUDA SOBERANA.....	15
1. LA INDUSTRIA DE LAS AGENCIAS DE CALIFICACIÓN DE CRÉDITO	16
2. LA CALIFICACIÓN DEL RIESGO SOBERANO.....	21
3. UNA PROPUESTA DE CALIFICACIÓN DE RIESGO SOBERANO.....	30
3.1 . Revisión de la literatura	30
3.2 La solvencia y la dinámica de la deuda publica	36
3.3. Modelo teórico para la estimación de los SR.....	44
3.4. El filtrado de resultados	47
CAPÍTULO 3 LA SELECCIÓN DE LA MUESTRA	53
1. INTRODUCCIÓN	54
2. SELECCIÓN DE LA MUESTRA Y PERIODO TEMPORAL	54
3. FUENTES DE INFORMACIÓN E INDICADORES	56

4. PROCESAMIENTO DE LOS INDICADORES	59
5. ASPECTO FINAL DE LA MUESTRA	64
6. ENTRENAMIENTO Y GENERALIZACIÓN	66
CAPÍTULO 4 LOS CLASIFICADORES	71
1 INTRODUCCIÓN	72
2. CLASIFICADORES NOMINALES	74
2.1. Regresión Logística	74
2.2. Redes Neuronales	75
2.3. Máquinas de Vectores de Soporte	76
2.4. Random Forest	77
3. CLASIFICADORES ORDINALES	82
3.1 Métodos considerados NAÏVE	83
3.1.1 Support Vector Classifier with OneVsOne SVC1V1	83
3.1.2 Support Vector Classifier with OneVsAll SVC1VA	83
3.1.3 Support Vector Machines for regression SVR	83
3.1.4 Cost-Sensitive Support Vector Classifier CSSVC	84
3.2. Métodos de descomposición	84
3.2.1 Neural Network with Ordered Partition NNOP	84
3.2.2 Extreme Learning Machine with Ordered Partition ELMOP	84
3.2.3 Hierarchical Partial Order Label Decomposition HPOLD	84
3.3 Modelos de límites	85
3.3.1 Kernel Discriminant Learner for Ordinal Regression KDLOR	85

3.3.2 Ordinal Projection Based Ensemble OPBE	85
3.3.3 Reduction applied to Support Vector Machines REDSVM	85
3.3.4 Boosting ensemble for Ordinal Regression ORBOOST	85
3.3.5 Neural Network based on Proportional Odd Model NNPOM	85
3.3.6 Support Vector for Ordinal Regression with Explicit constraints SVOREX	86
3.3.7 Support Vector for Ordinal Regression with Implicit constraints SVORIM	86
3.3.8 Linear version of the SVORIM method SVORIMLIN	86
4 MÉTRICAS DE ERROR	86
4.1. Métrica nominal	87
4.1.1 Datos correctamente clasificados CCR	87
4.2 Métricas ordinales	87
4.2.1 Datos aproximadamente clasificados ACR	88
4.2.2 Error absoluto medio MAE	88
4.2.3. Error absoluto medio por clases AMAE	88
5. SOFTWARE UTILIZADO	89
CAPITULO 5 EXPERIMENTACIÓN.....	91
1 INTRODUCCIÓN	92
2. HOLD-OUT	93
2.1 Consistencia de los resultados	95
2.2 Comparativa con otros trabajos.....	96

2.3 Análisis de sensibilidad	98
3. 10 FOLD.....	100
3.1 Random Forest.....	101
3.2 Regresión Logística	104
3.3 Red Neuronal	105
3.4 Máquina de Vector de Soporte	108
3.5 Clasificadores Ordinales.....	110
3.6 Análisis conjunto.....	111
4. FILTRADO DE DATOS	115
5. ANÁLISIS POR PAÍSES	120
5.1 Países referentes	120
5.2 Países estables	121
5.2.1 Bélgica.....	121
5.2.2 Austria	123
5.2.3 Finlandia.....	126
5.2.4 Francia	128
5.2.5 Reino Unido.....	130
5.2.6 Italia.....	133
5.3 Países inestables.....	135
5.3.1 España	136
5.3.2 Irlanda	139
5.3.3 Portugal	143

5.3.4 Grecia.....	145
6. RESULTADOS POR PAÍSES Y POR AÑOS	149
7. TIPIFICACIÓN DE LOS ERRORES.....	152
CAPITULO 6 CONCLUSIONES.....	155
1. CONCLUSIONES DE LA INVESTIGACIÓN	156
2 CONTRIBUCIÓN DE LA TESIS.....	159
3. LINEAS FUTURAS DE INVESTIGACIÓN	159
BIBLIOGRAFIA	161
ANEXOS	181

ÍNDICE DE GRÁFICOS

Gráfico 1: Distribución del mercado europeo de ratings por categorías 2015..	20
Gráfico 2: Evolución del PIB en distintos países europeos tras la crisis de 2008	42
Gráfico 3: Modelo de resiliencia ESPON.....	43
Gráfico 4: Representación gráfica del modelo propuesto.....	45
Gráfico 5: Proceso de clasificación y filtrado	49
Gráfico 6: Red neuronal:	75
Gráfico 7: Ejemplo de árbol de decisión	78
Gráfico 8: Random Forest	81
Gráfico 9: Resultados del ensamble de clasificadores para Austria.....	124
Gráfico 10: Resultados de Random Forest para Austria	125
Gráfico 11: Resultado del ensamble de clasificadores para Finlandia	127
Gráfico 12: Resultados de Random Forest para Finlandia.....	128
Gráfico 13: Resultados del ensamble de clasificadores para Francia.....	129
Gráfico 14: Resultados de Random Forest para Francia	130
Gráfico 15: Resultado del ensamble de clasificadores para Reino Unido	132
Gráfico 16: Resultados de Random Forest para Reino Unido.....	132
Gráfico 17: Resultados del ensamble de clasificadores para Italia	134
Gráfico 18: Resultados de Random Forest para Italia.....	135
Gráfico 19: Resultado del ensamble de clasificadores para España.....	138
Gráfico 20: Predicción de Random Forest para España	139
Gráfico 21: Resultado del ensamble de clasificadores para Irlanda.....	142
Gráfico 22: Resultados de Random Forest para Irlanda	142
Gráfico 23: Resultados del ensamble de clasificadores para Portugal.....	144

Gráfico 24: Resultados de Random Forest para Portugal.....	145
Gráfico 25: Resultados del ensamble de clasificadores para Grecia	147
Gráfico 26: Resultados de Random Forest para Grecia.....	148

ÍNDICE DE TABLAS

Tabla 1: Número de ratings emitidos por categoría por las CRA	18
Tabla 2: Cuota de mercado de las agencias de rating en Europa 2016.....	19
Tabla 3: Sistemas de ratings de Moody's S&P y Fitch	24
Tabla 4: Indicadores usados por las agencias de rating	26
Tabla 5: Trabajos sobre determinantes e SR. 1996-2018.....	31
Tabla 6: Variables detectadas en la revisión de la literatura y su frecuencia ...	35
Tabla 7: Solvencia de la persona física respecto a la solvencia de un estado soberano	41
Tabla 8: Ratings asignados por la agencia Moody's	59
Tabla 9: Frecuencia de ratings en la muestra	61
Tabla 10: Ratings y clases asignadas	62
Tabla 11: Variables y su tratamiento	63
Tabla 12: Distribución de clases por año.	65
Tabla 13: Frecuencia de patrones en el conjunto de generalización.....	93
Tabla 14: Resultados de 30 hold-out 70/30.....	94
Tabla 15: Media de los resultados en diferentes particiones.....	96
Tabla 16: Mínimos y máximos de datos correctamente clasificados en diferentes trabajos	97
Tabla 17: Medias de las 30 particiones con conjuntos de variables propuestos por otros autores:.....	98
Tabla 18: Resultados eliminando una variable.....	98
Tabla 19: Resultados tras añadir una variable al modelo propuesto.....	100
Tabla 20: Media y desviación típica de las 30 repeticiones de Random Forest.....	101

Tabla 21: Resultados de Random Forest.....	102
Tabla 22: Casos en los que las repeticiones de Random Forest dan resultados diferentes:	102
Tabla 23: Matriz de confusión de Random Forest.....	103
Tabla 24: Frecuencia de errores absolutos por países en Random Forest....	103
Tabla 25: Resultados Regresión Logística.....	104
Tabla 26: Matriz de confusión de la Regresión Logística	104
Tabla 27: Errores absolutos por países en la Regresión Logística	105
Tabla 28: Media y desviación típica de las 30 repeticiones de la Red Neuronal	105
Tabla 29: Resultados de la Red Neuronal.....	106
Tabla 30: Casos en los que la frecuencia de la moda de la Red Neuronal es menor al 70%.....	106
Tabla 31: Matriz de confusión de la Red Neuronal.....	107
Tabla 32: Errores absolutos por países en la Red Neuronal.....	108
Tabla 33: Resultados de la Máquina de Vector Soporte	108
Tabla 34: Matriz de confusión de la Máquina de Vector Soporte	108
Tabla 35: Errores absolutos por países en la Máquina de Vector Soporte	109
Tabla 36: Resultados con clasificadores ordinales.....	111
Tabla 37: Resultados del ensamble de todos los clasificadores	112
Tabla 38: Matriz de confusión del ensamble de clasificadores.....	112
Tabla 39: Promedio de error absoluto por clase.....	113
Tabla 40: Clasificación de los patrones en los distintos clasificadores.....	114
Tabla 41: Datos correctamente clasificados tras el filtrado	116
Tabla 42: Datos clasificados en el entorno de una clase tras el filtrado	117

Tabla 43: Error Absoluto Medio tras el Filtrado.	118
Tabla 44: Error Absoluto Medio por Ponderado por Clase tras el Filtrado	119
Tabla 45: Errores de clasificación por países de la serie original.....	149
Tabla 46: Errores de clasificación por Países sobre la serie filtrada	150
Tabla 47: Frecuencia de errores por año	151
Tabla 48: Errores absolutos por año de la serie transformada.....	152
Tabla 49: Frecuencia de errores por tipo del ensamble de clasificadores.....	153
Tabla 50: Frecuencia de errores por tipo en Random Forest.....	154

CAPÍTULO 1

INTRODUCCIÓN

1. LA VALORACIÓN DEL RIESGO PAÍS MEDIANTE LAS CALIFICACIONES DE DEUDA SOBERANA.

Las consecuencias económicas de la pandemia de la COVID-19 han supuesto un importante aumento de los déficits fiscales y de las ratios de deuda pública en la Unión Europea (Singh et al., 2021). En este contexto, la valoración del riesgo soberano tiene una importancia primordial, tanto por su peso relativo, ya que la deuda soberana supone más del 60% de la deuda emitida (House of Lords, European Union Committee, 2011); como por la limitación que suponen para las empresas que operan en sus territorios, ya que sus valoraciones de riesgo de impago suelen tener esta valoración como límite superior (Afonso et al., 2007; Al-Sakka & ap Gwilym, 2009; Bank For International Settlements, 2011) y su papel como instrumento de política monetaria (Rodríguez de Codes Elorriaga et al., 2020).

La valoración del riesgo soberano se lleva a cabo mediante de varios indicadores alternativos. Por un lado, los diferenciales de las permutas de incumplimiento crediticio (CDS spreads, en inglés) y los rendimientos de la deuda pública (sovereign bond yields, en inglés). Los primeros son una medida directa del riesgo de impago por parte del deudor, mientras que los segundos muestran el coste de su endeudamiento. Ambas medidas están basadas en el mercado, sin embargo los hechos han mostrados que son propensas a las interferencias políticas (Gade et al., 2021). Así, durante la crisis de deuda soberana europea, los bancos centrales adoptaron medidas que modificaron los precios de los bonos soberanos, tanto en los mercados primarios como secundarios (*European Central Bank Asset Purchase Programmes*, 2023), y las autoridades prohibieron

CAPÍTULO 1: Introducción

realizar compras al descubierto de CDS sobre países de la UE (International Monetary Fund, 2013.) De esta forma, ambas medidas dejaron de expresar exclusivamente las expectativas de riesgo de impago de los inversores y comenzaron a incorporar previsiones sobre las medidas de rescate.

Por otro lado, la calificación crediticia de la deuda soberana (sovereign rating, SR, en inglés), en forma de indicador alfabético sobre el riesgo de impago de deuda por un país (Cantor y Packer, 1995). Este indicador es emitido por empresas financieras especializadas, llamadas agencias de calificación. Si bien su uso se ha extendido gracias al incremento en el número de países calificados actualmente¹, su capacidad informativa se ha puesto en tela de juicio debido a su papel en las crisis económico-financieras más recientes, bien por su incapacidad para anticiparlas (Méjico en 1994-95 o Asia en 1997-99), bien porque sus actuaciones agravaron dichas crisis con bajadas excesivas de rating (Mora, 2006, Ferri y Stiglitz, 1999) o no justificadas, como en el caso europeo (Boumparis, 2015; Lennkh & Moshammer, 2018).

Lo anterior ha suscitado un debate sobre la falta de transparencia en los modelos de determinación de los SR tanto a nivel político-legal como científico. Los SR son asignados por un comité de expertos a partir de información económica-financiera sobre el país, pero también considerando otros aspectos más cualitativos, con un peso relativo indeterminado en este proceso de decisión (De Moor et al., 2018; Ozturk, 2014). Se trata, por tanto, de un proceso de caja negra (Alexe, 2003), donde existen dudas sobre cuáles son los determinantes de la capacidad de pago de un país (Brewer & Rivoli, 1990, Haque et al 1997) y en los

¹ Moody's Investor Services, una de las principales agencias de calificación, registra calificaciones para 172 estados soberanos. (Moody's, 2022).

CAPÍTULO 1: Introducción

que se puede dar cierto sesgo regional por parte de alguna de las agencias calificadoras (Haque et al., 1997).

Esta falta de transparencia ha motivado la actuación de los organismos financieros reguladores, especialmente tras la crisis financiera mundial y la crisis de deuda soberana europea (Binici et al., 2018). Así, en 2008 la Organización Internacional de Comisiones de Valores (en inglés, International Organization of Securities Commissions, IOSCO) revisó los “Fundamentos del Código de Conducta de las Agencias de Calificación Crediticia” para abordar cuestiones de conflicto de intereses, transparencia, competencia e independencia. Posteriormente, en Estados Unidos, como resultado de la Ley Dodd-Frank, se creó la Oficina de Calificaciones Crediticias (en inglés, Office of Credit Ratings, OCR) para supervisar y regular las agencias de calificación crediticia. Los informes de esta entidad, especialmente el del año 2015², ponen de manifiesto la existencia de problemas en el funcionamiento de las CRA y el incumplimiento de las normas del regulador. De forma análoga la Unión Europea emitió nuevas normas y reglamentos sobre las CRA, que se concretan, entre otras cosas, en la creación de la Autoridad Europea de Valores y Mercados (en inglés, European Securities and Markets Authority, ESMA) para la regulación de las CRA, y en la obligación de registros en el Comité de Responsables Europeos de Reglamentación de Valores por parte de las CRA que operan en Europa. A pesar

² El informe de 2015 documenta continuos incumplimientos legales de las CRA. En la Eurozona Grecia, Irlanda y Portugal tuvieron su deuda clasificada como bono basura por alguna de las agencias en la primavera de 2010. Desde las administraciones públicas se informó que estas bajadas de rating aceleraron la crisis financiera en la Eurozona. Estas críticas tuvieron cierta incidencia en la nueva regulación de las CRA. En un memorándum de la Comisión Europea se afirma que las agencias de rating tienen un gran impacto en los mercados financieros en los que los cambios de rating son seguidos por todos los actores. Los SR tienen un efecto crucial en los países pues una bajada de rating tiene un efecto inmediato en el coste de la deuda. La nueva legislación obliga a las CRA que operan en Europa a registrarse en el Committee of European Securities and Markets y su regulación depende de la European Securities and Markets Authority

CAPÍTULO 1: Introducción

de estos avances, las mejoras metodológicas implementadas por las agencias desde entonces, esto es, un mayor protagonismo de las variables económico-financieras de carácter cuantitativo, no han podido ser confirmadas mediante el análisis empírico (De Moor et al., 2018; Boumparis et al., 2015).

Junto a la falta de transparencia, otras de las cuestiones ampliamente debatidas es la prociclicidad de las rebajas de calificación crediticia. Desde el ámbito académico se ha prestado especial atención a la práctica generalizada de las agencias de las CRA de revisar a la baja las calificaciones crediticias durante periodos de crisis (Rodríguez de Codes Elorriaga et al., 2020) debido a sus efectos negativos sobre la economía financiera y la economía real. En primer lugar, la rebaja de la calificación crediticia afecta a los canales de transmisión de la política monetaria a través de la reducción del número de activos elegibles, tanto para su adquisición por los bancos centrales en el marco de los programas de compras, como para su uso como garantía por los intermediarios financieros en las operaciones de financiación³. Además, los emisores afectados por estas rebajas pueden padecer un endurecimiento en sus condiciones de financiación, tanto en el mercado de financiación bancaria como en el mercado de deuda, siendo especialmente preocupante si reciben una calificación por debajo del grado de inversión. Todo ello redundará en una contracción de la inversión agregada que se lleva a cabo en la economía afectada por dichas rebajas (Acharya et al., 2012; Almeida et al., 2017; Holthausen & Leftwich, 1986). Debido a lo anterior, desde el ámbito regulatorio, la cuestión de la prociclicidad ha sido

³ En Europa el BCE tuvo que tomar medidas de flexibilización temporal de los criterios de admisión de activos de garantía. Estas medidas se tomaron en el año 2020 y se irán eliminando gradualmente hasta 2024. También en los programas extraordinarios de compra de deuda se plantearon excepciones para la compra de deuda griega.

CAPÍTULO 1: Introducción

analizada por el Consejo de Estabilidad Financiera (en inglés, Financial Stability Board, FSB), estableciendo un conjunto de principios para que, entre otras cuestiones, los bancos centrales sustituyeran el uso de las calificaciones otorgadas por las CRA por evaluaciones internas (Financial Stability Board, 2014). En Estados Unidos, ya con anterioridad a la aparición de estos principios, la Sección 939A de la Ley Dodd-Frank trataba sobre el exceso de confianza vinculado a la regulación basada en las calificaciones crediticias. Siguiendo dicha sección, todas las agencias federales estadounidenses debían eliminar las calificaciones crediticias de sus legislaciones y sustituirlas por otras medidas alternativas de solvencia (De Pascalis, 2016). Por su parte, en la Unión Europea, el Reglamento 462/2013⁴, en su artículo 5, letras a, b y c, regula en esta línea y establece el propósito de suprimir tales referencias antes de 2020, una vez que se encuentren alternativas adecuadas (Reglamento (UE) n° 462/2013 por el que se modifica el Reglamento (CE) n° 1060/2009 Sobre Las Agencias de Calificación Crediticia, 2013). Desde entonces se han implementado varias iniciativas, entre las que destacan el análisis de las distintas fuentes de evaluación crediticia empleadas y la autorización de nuevos modelos internos de calificación crediticia (Directorate-General for Internal Market, 2014). Sin

⁴ La legislación europea para reforzar el marco regulador y de supervisión de las agencias de calificación crediticia se ha introducido en tres fases consecutivas, conocidas como CRA I, II y III (Jones et al., 2022). El primer conjunto de normas, que entró en vigor a finales de 2009, estableció un marco regulador para las CRA e introdujo un régimen de supervisión reglamentaria, en virtud del cual las CRA debían estar registradas y eran supervisadas por las autoridades nacionales competentes. Además, las agencias de calificación crediticia debían evitar los conflictos de intereses y disponer de metodologías de calificación sólidas y actividades de calificación transparentes. En 2011, estas normas se modificaron para tener en cuenta la creación de la Autoridad Europea de Valores y Mercados, que supervisaba las agencias de calificación crediticia registradas en la UE. En 2013 se introdujo otra modificación para reforzar las normas y subsanar las deficiencias relacionadas con las calificaciones crediticias de la deuda soberana.

CAPÍTULO 1: Introducción

embargo, la dependencia de la información suministradas por las CRA es todavía elevada (Rodríguez de Codes et al., 2020).

Ligada a la prociclicidad de las calificaciones, aparece la cuestión de la asimetría en los ciclos de calificación (Broto & Molina, 2016). Esta cuestión se focaliza en la respuesta de las agencias de rating ante los cambios las variables determinantes de las calificaciones, también conocidas como fundamentos de los países, y las condiciones del mercado financiero, dando lugar a modificaciones en las calificaciones crediticias. En este sentido, se entiende por un ciclo de calificación el periodo que transcurre entre una fase de descenso de la calificación crediticia y una fase de recuperación posterior, sin que ésta ulterior mejora en la calificación suponga recuperar la calificación inicial (Koopman et al., 2009; Sy, 2002);. Los ciclos de calificación se caracterizan por sus fuertes asimetrías, ya que su duración y amplitud difieren en las fases de bajada y subida (Broto & Molina, 2016). Así, el análisis de los ciclos de calificación muestra que los descensos en las calificaciones son más bruscos, mientras que los ascensos suelen ser mucho más lentos (Koopman et al., 2009), esto es, una vez un país ve rebajada su calificación crediticia necesita de un periodo de tiempo más prolongado para conseguir recuperar dicha calificación.

La literatura empírica apunta dos vías para explicar el comportamiento de las agencias en los ciclos de calificación. Por un lado, el uso de una estrategia “point-in-time”, de forma que los cambios de calificación reflejan las condiciones del país de forma actualizada (Hu et al., 2002). Bajo esta perspectiva, la menos extendida, las asimetrías en los ciclos de calificación pueden explicarse por asimetrías en la propia información empleada por las agencias para actualizar las calificaciones durante los periodos de recesión y recuperación. Por otro, el

CAPÍTULO 1: Introducción

uso de la metodología “through-the-cycle”, seguida por las agencias para evitar causar perturbaciones en los mercados (Löffler, 2004). Bajo este enfoque, las agencias aplican un principio de prudencia en la actualización de las calificaciones, primando el componente de riesgo de créditos permanente sobre las fluctuaciones a corto plazo en los fundamentos del país (Altman & Rijken, 2005). En este caso, las asimetrías en los ciclos de calificación serían el resultado del dilema entre estabilidad y precisión que deben afrontar las CRA (R. Cantor & Mann, 2007). Cabe señalar que la mayoría de las investigaciones sobre el ajuste de las calificaciones crediticias a los fundamentales se focaliza en periodos de crisis financiera (Ferri et al., 1999; Mora, 2006; Mulder & Monfort, 2000), mientras que los que analizan las calificaciones de los países y su relación con el ciclo económico son más bien escasos. Entre estos últimos, cabe señalar el trabajo de Kiff et al. (2013) donde se muestra cierto desfase en los periodos de rebaja de la calificación y una tendencia a suavizar los periodos de recuperación de las calificaciones. De forma similar, Broto y Molina (2016) apuntan que las agencias de calificación ralentizan las rebajas de calificación ante unos fundamentales favorables, pero estos no parecen acelerar las subidas de calificación.

2. OBJETIVO DE INVESTIGACIÓN

Dada la actual relevancia de las calificaciones de riesgo soberano y los problemas que éstas muestran, la presente investigación pretende corroborar si es posible la estimación de SR a partir de datos económicos de carácter cuantitativo mediante la aplicación de técnicas de aprendizaje automático o

CAPÍTULO 1: Introducción

machine learning (ML) y la mejora de los resultados obtenidos mediante el tratamiento de los datos clasificados para asimilar la clasificación al comportamiento asimétrico de las agencias de rating. Esta aportación sería de gran utilidad para que los gobiernos anticiparan posibles cambios de calificación, mientras que los inversores y las instituciones financieras comprobaran si las calificaciones se desvían de lo que indican los fundamentos de un país (Overes & van der Wel, 2022). Además, el uso de ML es de gran utilidad para explorar cuestiones financieras, pues la solidez de los algoritmos subyacentes permite generar modelos optimizados para explicar los datos, sin necesidad de acudir a supuestos exógenos y limitando posibles sesgos (Li et al., 2020), por lo que son ampliamente usados también para pronósticos y predicciones (Lee et al., 2014; Wang et al., 2020); Esto redundará en una mejora de las evaluaciones y de la toma de decisiones,

Para ello se formulan los siguientes objetivos generales y específicos:

Objetivo general 1: Proponer un modelo de estimación de SR a partir de variables macroeconómicas

- Objetivo específico 1.1. Obtener un conjunto de variables que representen la solvencia de un país, basándose en fundamentos macroeconómicos.
- Objetivo específico 1.2. Aplicar distintas técnicas de inteligencia artificial al conjunto variables seleccionadas para comprobar la representatividad de las variables con distintos clasificadores.

Objetivo general 2: A partir de los resultados obtenidos de las diferentes técnicas de clasificación, analizar posibles efectos de la asimetría en las estimaciones

CAPÍTULO 1: Introducción

realizadas y proponer mecanismos que mejoren las clasificaciones obtenidas teniendo en cuenta estas asimetrías en los ciclos de rating.

- Objetivo específico 2.1. Utilizando la técnica de filtrado (R. Cantor & Mann, 2007) tratar las clasificaciones obtenidas con las distintas técnicas de inteligencia artificial para comprobar la posible mejora.

3. CONTRIBUCION DE LA TESIS

La consecución de los objetivos anteriores supone la obtención de un modelo teórico para la determinación del SR a partir de información económica de carácter público. De esta forma se ofrece una medida objetiva, transparente y accesible del riesgo país, ofreciendo una alternativa frente a las limitaciones actuales de los modelos que han sido apuntadas por la literatura científica. En este sentido, el modelo propuesto se vertebra en torno al concepto de solvencia y las variables que los componen recogen factores directamente relacionadas con la determinación de la capacidad de pago de un país. Estas variables se alimentan con datos públicos lo que asegura su transparencia y la replicabilidad del modelo para periodos y, en su caso, para contextos geográficos, distintos del recogido en la aplicación empírica. A partir de los resultados obtenidos de las diferentes técnicas de clasificación, analizar posibles efectos de la asimetría en las estimaciones realizadas y proponer mecanismos que mejoren las clasificaciones obtenidas teniendo en cuenta estas asimetrías en los ciclos de rating. Además, Las técnicas de aprendizaje automático utilizadas para evaluar la robustez del modelo propuesto han empleado las implementaciones

CAPÍTULO 1: Introducción

proporcionadas por software estándar, con el detalle mostrado, con el fin de favorecer la reproducibilidad de los resultados. .

Por último, el análisis de la asimetría en los ciclos de rating y la introducción del filtrado, como técnica para mejorar la clasificación, permiten una mejor evaluación de la capacidad de estimación de las distintas técnicas de ML al introducir un mecanismo que permite asimilar los resultados obtenidos al comportamiento real de las agencias de rating. Dicha mejora se basa en la automatización de los patrones de comportamiento detectados en las variaciones de los SR ante cambios en la solvencia de los distintos países y como se trasladan estos cambios por las agencias de rating a lo largo del periodo estudiado.

4. ESTRUCTURA DE LA TESIS

La estructura de la tesis se detalla a continuación. En el capítulo 2 se describen en primer lugar las agencias de rating y su funcionamiento. Posteriormente se hace una revisión de la literatura en la que se describen los principales trabajos de estimación de ratings de deuda soberana que se han realizado indicando sus metodologías y las variables que se han utilizado. Posteriormente se procede a realizar un análisis teórico de solvencia para lo que se describen algunos conceptos macroeconómicos entre los que destaca el concepto de dinámica de la deuda. Sobre este particular concepto: dinámica de la deuda se hace una revisión de la literatura con las principales teorías que se han publicado. Con toda esta información se propone un modelo teórico para estimar los ratings de la deuda soberana en el que se definen las variables que deben incluirse en el

CAPÍTULO 1: Introducción

modelo. Por último, con la información obtenida tras la clasificación, con el análisis de las metodologías publicadas por las agencias de rating y a partir de un modelo de tratamiento de datos “filtrado” que explica el comportamiento real de las agencias de rating se propone aplicar esta metodología a nuestro modelo para mejorar las clasificaciones.

El capítulo 3 define como se ha procedido para seleccionar la muestra y que tratamiento se ha dado a los distintos indicadores que intervienen en el modelo planteado. Posteriormente se explica cómo se ha tratado la muestra y como se han solventado los problemas específicos de la muestra seleccionada.

El capítulo 4 hace referencia a los clasificadores utilizados en la experimentación. Se describen los clasificadores nominales y ordinales que se han utilizado así como las métricas utilizadas para su evaluación. Por último, se informa sobre el software utilizado en la experimentación.

El capítulo 5 describe la experimentación. En primer lugar, se analizan los resultados obtenidos con técnicas de hold-out con los clasificadores nominales. Sobre estos resultados se hace un análisis de sensibilidad añadiendo y quitando variables. También se comparan los resultados con otros autores y se incluyen las variables seleccionadas por otros autores en nuestro modelo para comprobar cómo se comportan estas variables comparadas con las propuestas. Seguidamente se procede a hacer un análisis con 10-fold. Este análisis permite un estudio de los resultados de la muestra completa. Se trabaja con clasificadores nominales y ordinales. Posteriormente se hace un análisis conjunto de todos los clasificadores comprobando las ocasiones en los que los errores de clasificación son más frecuentes. También se aplica la técnica del filtrado y se comprueban las mejoras obtenidas en las clasificaciones.

CAPÍTULO 1: Introducción

El capítulo 5 se hace un análisis por país de los resultados obtenidos por el ensamble de todos los clasificadores utilizados y de los resultados obtenidos con la aplicación de “Random Forest”.

El capítulo 6 presenta las conclusiones del trabajo realizado como se contribuye a los objetivos planteados y que aplicaciones tendría el modelo propuesto. También se informa de las limitaciones que tiene el análisis realizado y los futuros trabajos a realizar en este campo.

5. DIFUSIÓN DE LA TESIS

La difusión de la presente tesis se ha llevado a cabo mediante la participado en dos congresos internacionales y la publicación un artículo en una revista indexada en el JCR, con el detalle que se muestra a continuación.

- Galnares Jiménez-Placer Carlos, Martínez Estudillo Alfonso Carlos y Carbonero Ruz Mariano (2018) “Estimating de Rating in the UE-15 2002-2017” Comunicación presentada en Ecomod 2018 International Conference on Economic Modelling Universidad Ca’ Foscari Venecia.
- Galnares Jiménez-Placer Carlos, Martinez Estudillo Alfonso Carlos y Carbonero Ruz Mariano (2019). “Estimación de ratings de deuda soberana en la ue15 de 2002 a 2017 mediante clasificadores ordinales”. Comunicación presentada en el XXXIII Congreso Internacional de Economía Aplicada Asepelt celebrado en Vigo donde se presentó el trabajo titulado que incluía el modelo definitivo analizándose desde la perspectiva de los clasificadores ordinales.

CAPÍTULO 1: Introducción

Una parte de la investigación ha sido publicada en una revista indexada en el Journal Citation Report

- Galnares Jiménez-Placer Carlos, Martínez Estudillo Alfonso Carlos, Carbonero Ruz Mariano y Campoy Muñoz Pilar (2022). “Revisiting the determinants of sovereign debt ratings in Europe through artificial intelligence techniques” Applied Economic Letters <https://doi.org/10.1080/13504851.2022.2097171>).

CAPÍTULO 2

MODELO DE ESTIMACIÓN DE RATINGS DE DEUDA SOBERANA

1. LA INDUSTRIA DE LAS AGENCIAS DE CALIFICACIÓN DE CRÉDITO

Las CRA existen desde hace más de 150 años. Durante este tiempo, han pasado de ser meras recopiladoras y proveedoras de información a árbitros cuasi-oficiales del riesgo crediticio en el sistema financiero mundial (Dimitriadou et al., 2021). En la década de 1850, Henry Poor, editor del American Railroad Journal, decidió publicar información estadística sobre la propiedad de los ferrocarriles, que resultó de gran utilidad para decisiones de inversión, ya que hasta el momento los inversores individuales obtenían este tipo de información a través de relaciones comerciales o bancarias. Posteriormente, hacia 1909, John Moody comenzó a proporcionar información crediticia, en forma de calificación, sobre la solvencia de varios bonos, basándose en información del sector y en análisis estadísticos. En 1913, John Knowles Fitch crea Fitch Publishing Company, con la finalidad de publicar estadísticas financieras sobre acciones y bonos, y desarrolla un sistema de calificación de deuda basado en la capacidad de una empresa para atender sus obligaciones.

Estas iniciativas fueron el germen de las actuales tres grandes CRA que, en sus inicios, instauraron el modelo de pago, en forma de cuotas de suscripción, por los servicios de información y calificación recibidos por los inversores (“investor-pays system”). Tras el desplome del mercado bursátil estadounidense en 1929, la Ley Glass-Steagall (1933) limitó la inversión de los intermediarios financieros a valores calificados con “grado de inversión” por las CRA, otorgando un rol formal a éstas en los mercados financieros. Posteriormente, la recesión de la década de 1970 llevó a la Comisión del Mercado de Valores de los Estados Unidos a designar a estas tres CRA como Organizaciones de Calificación

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

Estadística Nacionalmente Reconocidas (NRSPO, en inglés). Desde entonces, la confianza de los reguladores en las NRSPO fue en aumento, lo que se tradujo en una creciente demanda de valoración de activos financieros por parte de inversores y emisores ante el incremento del riesgo y del coste de financiación ocasionado por la “selección adversa”. De forma paralela, las CRA viraron hacia un modelo de emisor-pagador (“issuer-pays system”) que persigue calificar nuevas emisiones de deuda, especialmente en el mercado primario⁵ (Sylla, 2002).

La crisis financiera mundial supuso un cambio drástico en la percepción de la actividad de las CRA, recibiendo severas críticas desde los ámbitos económico-financiero, político-legal y académico (HOUSE OF LORDS European Union Committee, 2011). Las principales críticas versaban sobre los conflictos de intereses inherentes a su modelo de negocio, los cambios *ex post* en las calificaciones crediticias, la falta de control normativo y de supervisión, debido especialmente a la excesiva dependencia de instituciones financieras e inversores en las calificaciones otorgadas por las agencias, la responsabilidad civil limitada de las agencias, y la falta de competencia en la industria. De hecho, en el año 2000 solo existían tres agencias reconocidas como NRSPO (Dimitriadou et al., 2021). Desde entonces, como se ha señalado en el capítulo precedente, se ha legislado para fomentar la transparencia, la responsabilidad e incrementar el número de CRA en la industria (Vu et al., 2022). Aunque actualmente el número de NRSPO asciende a diez (U.S. Securities and

⁵ También publican ratings “no solicitados” que se generan sin que sean solicitados por el emisor o la entidad clasificada (HOUSE OF LORDS European Union Committee, 2011).

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

Exchange Commission, 2023), el sector de las agencias de rating sigue presentando una arquitectura oligopolista, tal y como reflejan las cuotas de mercado recogidas en la siguiente tabla:

Tabla 1: Número de ratings emitidos por categoría por las CRA

	Sector público		Resto		Total	
	2010	2017	2010	2017	2010	2017
A.M.Best	No registr.	No registr.	7.159	8.275	7.159	8.275
DBRS	13.533	18.865	29.051	30.783	42.584	49.648
EJR	19	No registr.	1.026	16.730	1.045	16.730
Fitch	363.897	205.674	141.127	90.491	505.024	296.165
HR Ratings	0	374	0	744	0	1.118
JCR	52	440	684	3.362	736	3.802
KBRA	59	72	17.565	8.980	17.624	9.052
Moody's	841.235	598.614	197.952	127.070	1.039.187	725.684
Mornigstar	No registr.	No registr.	8.322	2.871	8.322	2.871
R&I	1.031	0	3.387	0	4.418	0
S&P	965.900	920.306	224.600	158.560	1.190.500	1.078.866
Total	2.185.726	1.744.345	630.873	447.866	2.816.599	2.192.211

Fuente: U.S. Securities and Exchange Commission

En 2010, solo tres CRA (Fitch, Moody's y S&P) representaban aproximadamente el 97% de todas las calificaciones crediticias en circulación, si bien varían entre categorías, en el caso de las calificaciones crediticias de deuda pública, municipal y soberana este porcentaje se eleva hasta el 99%, siendo S&P y Moody's las que ostentaban mayor volumen de negocio. En 2017, este porcentaje se reduce ligeramente hasta el 95,8%, si bien siguen copando el mercado de las obligaciones emitidas por entes públicos.

En Europa la obligación de certificación y/o registro aparece con posterioridad⁶, y, aunque en 2017 las CRA certificadas/registradas eran veintiséis, la posición hegemónica de Fitch, Moody's y S&P se repite tanto a nivel global, con una cuota

⁶ Reglamento (CE) nº 1060/2009 del Parlamento Europeo y del Consejo, de 16 de septiembre de 2009, sobre las agencias de calificación crediticia (Reglamento sobre las agencias de calificación crediticia).

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

de mercado del 93,2 % (Tabla 2), como en el caso de la deuda pública (Gráfico 1).

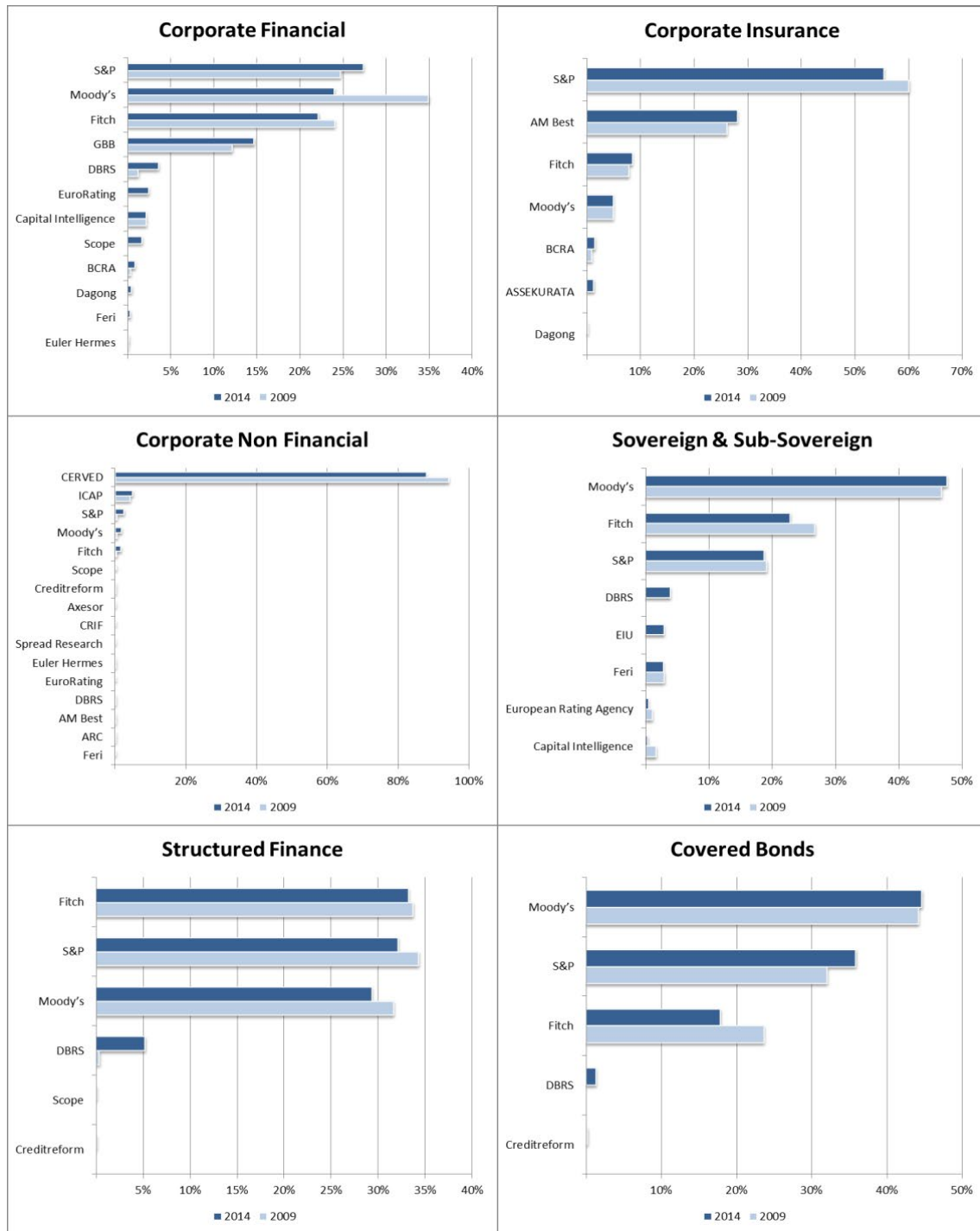
Tabla 2: Cuota de mercado de las agencias de rating en Europa 2016

Agencia	Cuota	Evolución
S&P Global Ratings	46,26	↑
Moody's Investors Service	31,27	↓
Fitch Ratings	15,65	↓
DBRS Ratings	1,87	↓
CERVED Rating Agency	0,97	↑
AM Best Europe Rating Services	0,90	↓
The Economist Intelligence Unit	0,69	↓
CreditReform Rating	0,53	↑
Scope Ratings	0,46	↑
GBB-Rating	0,35	↑
Assekurata	0,23	↑
Euler Hermes Rating	0,22	↑
Capital Intelligence Ratings	0,13	↓
ICAP	0,12	↑
ModeFinance	0,08	↑
Spread Research	0,07	↓
Dagong Europe Credit Rating	0,07	↑
ARC Ratings	0,05	↑
Axesor Rating	0,03	↓
CRIF Ratings	0,03	↓
BCRA Credit Rating Agency	0,02	↓
EuroRating	0,01	↑
INC Rating	<0,01	↑
European Rating Agency	<0,01	↑
Rating-Agentur Expert RA GmbH	<0,01	↑

Fuente: ESMA

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

Gráfico 1: Distribución del mercado europeo de ratings por categorías 2015



Fuente: ESMA

2. LA CALIFICACIÓN DEL RIESGO SOBERANO

Las calificaciones crediticias miden el riesgo relativo de incumplimiento de los compromisos financieros (pago de intereses y devolución del principal) por parte de una entidad en el tiempo establecido. En el caso de las calificaciones crediticias de riesgo soberano (SR), se considera que el país soberano incurre en impago cuando, o bien no atiende puntualmente las obligaciones de pago de la deuda pública emitida, o bien intenta conseguir un canje sobre la deuda original. El riesgo de impago es medido por Fitch⁷ mediante la probabilidad de impago, en el caso de los emisores, y la pérdida esperada, en el caso de los instrumentos, mientras que S&P solo considera la probabilidad de impago y Moody's la pérdida esperada, tal y como reflejan las definiciones de calificación crediticia de estas agencias:

*"Las calificaciones crediticias expresan el riesgo en orden relativo, es decir, son medidas ordinales del riesgo crediticio y no predicen una frecuencia específica de impago o pérdida. Las calificaciones crediticias de **Fitch Ratings** no abordan directamente ningún riesgo distinto del riesgo crediticio, las calificaciones no tratan el riesgo de pérdida de valor de mercado de un valor calificado debido a cambios en los tipos de interés, la liquidez y otras consideraciones de mercado." Fitch (2022).*

⁷Dada la importancia relativa de Fitch, S&P y Moody's en la industria de calificaciones crediticias y, especialmente, en el segmento de riesgo soberano, el presente apartado se focaliza en estas tres CRA. Además, la bibliografía empleada tendrá como referencia temporal el periodo comprendido entre los años 2002 y 2017, correspondiente al marco temporal del análisis empírico realizado en la presente tesis doctoral, con especial atención a los años siguientes a la crisis financiera global, debido a los cambios acontecidos.

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

*“Las calificaciones crediticias de **Standard & Poor's** están diseñadas principalmente para proporcionar clasificaciones relativas entre emisores y obligaciones de solvencia general; las calificaciones no son medidas de probabilidad absoluta de impago. La solvencia abarca la probabilidad de impago y también incluye la prioridad de pago, la recuperación y la estabilidad crediticia” (Standard & Poor's RatingsDirect, 2009).*

*“**Moody's** mantiene dos sistemas o escalas de calificación de bonos distintos. Una de ellas -Moody's Global- se aplica a las calificaciones asignadas a instituciones financieras y no financieras, emisores soberanos y sub-soberanos fuera de Estados Unidos y obligaciones de financiación estructurada. La Escala Global es una correspondencia entre las categorías de calificación y las tasas relativas de pérdidas esperadas en múltiples horizontes. La pérdida esperada comprende una evaluación de la probabilidad de impago, así como la expectativa de pérdida en caso de impago. La intención de Moody's es que la tasa de pérdida esperada asociada a una con un símbolo de calificación y un horizonte temporal determinados sea el mismo para todas las obligaciones y emisores calificados en la Escala Global. Las metodologías de calificación de Moody's, las prácticas de calificación y los sistemas de supervisión del rendimiento están diseñados para garantizar la coherencia de significado. [...] Un sistema de calificación independiente -Moody's US Municipal Scale- engloba las calificaciones asignadas a gobiernos estatales y locales, organizaciones sin ánimo de lucro y entidades afines que emiten*

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

deuda en el mercado estadounidense de bonos exentos de impuestos. Las tasas históricas de impago y pérdidas de las obligaciones calificadas en los Estados Unidos son significativamente inferiores a las de las obligaciones de empresas con calificaciones similares”(Moody’s Investor Service, 2009).

Esto implica que, en teoría, las calificaciones otorgadas por Fitch y S&P sobre el mismo emisor serían diferentes de las otorgadas por Moody’s, puesto que la pérdida esperada puede aproximarse como el producto de la probabilidad de impago y la gravedad de la pérdida esperada. Sin embargo, en la práctica, existe escasa divergencia en las calificaciones, especialmente cuando estas están pertenecen al subgrupo de grado de inversión (FMI, 2010).

La calificación se establece mediante una clasificación ordinal de letras, como puede observarse en la Tabla X. De mayor a menor solvencia, Fitch y S&P emplean AAA, AA, A y BBB para el riesgo de crédito a largo plazo con grado de inversión, y BB, B, CCC, CC, C y D para el riesgo de crédito a largo plazo "especulativo". La nomenclatura es prácticamente idéntica en el caso de Moody’s, que emplea una combinación de letras mayúsculas y minúsculas y asigna al impago la letra “C”. Además, se utilizan modificadores para distinguir y jerarquizar las calificaciones dentro de cada una de las clasificaciones más amplias: Fitch y S&P utilizan los símbolos más y menos, mientras que Moody's utiliza números. Al asignar estas calificaciones Fitch y S&P diferencian a veces entre obligaciones en moneda local y extranjera, y la diferencia suele ser a favor de las primeras, lo que refleja el mayor acceso del soberano a la moneda local, sin embargo Moody’s apenas hace esta distinción (Moody’s Investors Service, 2010). Por último, la intención de las CRA de modificar las calificaciones suele

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

ser indicada con antelación mediante notificaciones negativas de “revisión” u “observación”, lo que indica que es probable una rebaja en la calificación en un horizonte de 3 meses.

Tabla 3: Sistemas de ratings de Moody's S&P y Fitch

Caracterización de la deuda y el emisor (fuente: Moody's)		Agencia de rating		
		Moody's	S&P	Fitch
Máxima calidad	Grado de inversión	Aaa	AAA	AAA
Alta calidad		Aa1	AA+	AA+
		Aa2	AA	AA
		Aa3	AA-	AA-
Fuerte capacidad depago		A1	A+	A+
		A2	A	A
		A3	A-	A-
Capacidad de pago adecuada		Baa1	BBB+	BBB+
		Baa2	BBB	BBB
	Baa3	BBB-	BBB-	
Probable cumplimiento de las obligaciones depago, incertidumbre en curso	Ba1	BB+	BB+	
	Ba2	BB	BB	
	Ba3	BB-	BB-	
Alto riesgo de credito	B1	B+	B+	
	B2	B	B	
	B3	B-	B-	
Muy alto riesgo de credito	Caa1	CCC+	CCC+	
	Caa2	CCC	CCC	
	Caa3	CCC-	CCC-	
Cerca del default con posibilidad de recuperación	Ca	CC	CC C	
	C	SD D	DDD DD D	
Default				

Fuente: Afonso et al., (2007)

Ninguna de las CRA ha cuantificado explícitamente las escalas empleadas, si bien es posible establecer una distribución de las probabilidades de impago por calificación a partir de los resúmenes ex post de los impagos por grados de calificación (FMI, 2010). Además, las CRA no se ocupan especialmente de mantener constantes las tasas de impago, pues esto conllevaría cambiar las

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

calificaciones "en masa" en respuesta a cambios en las condiciones cíclicas (R. M. Cantor & Mann, 2003). Las CRA se han esforzado en aclarar esta idea de ordenación del riesgo, tanto por las críticas realizadas a sus modelos como por las incompatibilidades con algunos de los usos previstos por las autoridades monetarias⁸. Otro de los puntos ampliamente clarificados, y relacionado con el anterior, se refiere a la estabilidad en las calificaciones. Las CRA han pretendido que los grados de calificación más altos sean más estables que los grados más bajos, lo que minimizaría los costes relacionados con las transacciones que se realizarían ante cambios frecuentes en las calificaciones (R. Cantor & Mann, 2007). Para ello han utilizado un enfoque "a lo largo del ciclo", en el que se considera la capacidad de supervivencia del emisor ante depresiones cíclicas y las calificaciones otorgadas solo se modifican ante cambios en factores fundamentales. Esto implicaría que fenómenos como una recesión o una restricción de liquidez a nivel mundial, por sí solos, no ocasionarían una rebaja en la calificación (FMI, 2010).

Para asignar una calificación crediticia, las CRA consideran factores cuantitativos y cualitativos con los que calibran la capacidad y la voluntad de un país de reembolsar su deuda⁹ (Fitch Ratings, 2012; Moody's, 2013; Standard & Poor's, 2013). Aunque cada CRA tiene una metodología a la hora de determinar los SR, pues difieren en la forma de clasificar y agrupar los factores

⁸ En Basilea II, el método estándar se basaba en calificaciones AAA/AA que implican una probabilidad del 0,10% de impago durante un período de tres años, las calificaciones A una probabilidad del 0,25%, el 1,00% para los instrumentos BBB, y así sucesivamente (Basel Committee on Banking Supervision, 2006). En el caso del Eurosistema, los activos de garantía aportados contra operaciones de política monetaria debían contar con una calificación BBB, lo que implicaba una probabilidad de impago del 0,40% en un año (Banco Central Europeo, 2008).

⁹ La disposición a pagar evalúa el riesgo potencial de que, aunque el soberano tuviera la capacidad de pagar, estuviera dispuesto a hacerlo si considera que los costes sociales o políticos son demasiado elevados.

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

individuales¹⁰, el análisis conjunto de las tres CRA permite resaltar el uso común de variables macroeconómicas, así como variables relativas a la solidez financiera del sector público, la solidez de la financiación exterior y la calidad política e institucional (Choy et al., 2021), provenientes tanto de fuentes públicas, como de las autoridades del país, tal y como se muestra en la Tabla 4.

Tabla 4: Indicadores usados por las agencias de rating

	Moody's	Fitch	S&P
Macro/ Crecimiento	PIB per cápita Volatilidad a largo plazo del output nominal Escala de la economía Integración en zonas económicas y de comercio	PNB y PIB per cápita Consistencia de la política monetaria y Fiscal y credibilidad del marco político a largo plazo Competitividad de la Economía Profundidad de la demanda de moneda local Capacidad de implementar políticas contra cíclicas Composición de la Cuenta	Ratio y modelo de crecimiento económica Rango y eficiencia de las herramientas de política monetaria Tamaño y composición de los ahorros y las inversiones Dinero y expansión del crédito Comportamiento de los precios en los ciclos económicos
Finanzas públicas	Capacidad del gobierno para elevar impuestos, recortar gastos, vender activos u obtener moneda extranjera (por ej. De las reservas oficiales)	Activos financieros del gobierno Posición financiera neta de activos soberanos Volatilidad de los Ingresos Públicos Ratio Ingresos/PIB Dinámica de la deuda a largo plazo Credibilidad del marco de la política fiscal y las instituciones empresas no financieras del sector público Flexibilidad financiera	Tendencias en los ingresos generales del gobierno, gastos y déficit Compatibilidad de la posición fiscal con la monetaria y los factores externos Eficiencia y flexibilidad de los incrementos de ingresos Efectividad del gasto y presiones Tamaño y salud de las empresas no financieras del sector público
Deuda	Nivel de deuda Pagos e ingresos por intereses Estructura de la deuda pública Carga del reembolso de la deuda Dinámica de la deuda Pasivos condicionales Profundidad financiera	Tamaño y tasa de crecimiento de la deuda pública Composición de la deuda pública (madurez, tipo de interés y moneda) Madurez y estructura en divisas de los activos y pasivos financieros por sector Historial de pagos	Deuda pública bruta y neta; deuda externa bruta y neta Porcentaje de los ingresos dedicados a intereses Carga del servicio de la deuda Madurez, perfil y composición monetaria Acceso a fondos concesionales Deuda y pulso del mercado local de capitales
Sector financiero	Fortaleza del sector público Pasivos contingentes del sector bancario	Indicadores Macro-Prudentes de riesgo Calidad del sector bancario y su supervisión	Robustez del sector financiero Efectividad del sector financiero

¹⁰ Fitch agrupa las variables en cuatro pilares (Fitch Ratings, 2010, 2017); S&P ha reducido de nueva a cinco el número de agrupaciones (Beers et al., 2008; Standard & Poor's, 2017); mientras que Moody's emplea cuatro agrupaciones (Moody's Investor Service, 2008, 2016b).

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

		Pasivos contingentes del sector bancario	
		Propiedad extranjera del sector bancario	
Finanzas públicas	Dinámica de la balanza de pagos Reservas en moneda extranjera Acceso al cambio de moneda extranjera Indicador de vulnerabilidad externa	Flujos de capital Demanda de los no residentes de aumentar el crédito y comprar activos domésticos Porcentaje del output corriente destinado al servicio de la deuda externa Adecuación de las reservas	Impacto de la política fiscal y monetaria en las cuentas externas Composición de los flujos de capital Adecuación de las reservas
Tipo de cambio	Regímenes de tipos de cambio Indexación y dolarización	Regímenes de tipos de cambio Indexación y dolarización	Compatibilidad del régimen de tipo de cambio y las metas financieras Indexación y dolarización
Políticas	Guerra Grado de consenso político Caos político Eficiencia y predictibilidad De las acciones de gobierno Nivel de transparencia política	Riesgo de guerra Legitimación del régimen Político Relaciones con la comunidad internacional y las instituciones	Estabilidad y legitimización de las instituciones políticas Participación popular en el proceso político Orden en la sucesión de líderes Transparencia en la política económica decisión y objetivos Seguridad pública Riesgo geopolítico
Estructural/ Institucional	Transparencia Nivel de innovación Inversión en capital humano Respeto a los derechos de propiedad	Efectividad del gobierno Apertura flujos de capital internacional y al comercio Fortaleza del ambiente de negocios, capital humano y gobernanza Imperio de la ley, respeto a los derechos de propiedad Control de la corrupción	Eficiencia del sector público Factores institucionales como independencia del banco central Plazos, cobertura y transparencia en la información Competitividad y beneficios del sector privado
Otros	Terremotos Huracanes Crisis especulativa	Ratios de ahorro Apertura de la economía al comercio Dependencia de commoditys	Prosperidad, diversidad y grado de la orientación del mercado Discrepancias de ingresos Proteccionismo y otras influencias fuera del mercado Flexibilidad laboral

Fuente: IMF (2010) Global financial stability report

Toda esta información es considerada por un comité de calificación que determina la importancia de los factores y considera otras circunstancias cambiantes, lo que se traduce en una variación de la importancia relativa de cada factor a lo largo del tiempo (Kiff et al. 2010). Además, en esta valoración impera el enfoque “through the cycle” (TTC), que implica que un SR no se cambie si se estima probable una reversión de este cambio en el corto plazo (Moody’s, 2009).

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

Esto explica las posibles diferencias en SR si se empleara el enfoque "point in time" (PIT) donde la estimación depende de los datos relacionados con el período de estimación (Carey y Mark, 2001; Altman y Rijken, 2006; Topp y Perl, 2010; Kiff et al ; 2013). Varias investigaciones han intentado modelizar este comportamiento. Luffler (2004) analizó las metodologías anticíclicas de las agencias de rating en el caso de los ratings corporativos y cómo estos presentan una mayor estabilidad y menor capacidad de predicción. Altman y Rijken (2004) definieron los límites en los que se realizan cambios en las calificaciones crediticias. Más tarde, Luffler (2005) definió un modelo estadístico de estimaciones de ratings basado en cómo las agencias de rating evitan rebotes en las clasificaciones. En respuesta, la agencia Moody's argumentó la importancia de la estabilidad de los ratings y detalló cómo esta estabilidad afectaba a la precisión, proponiendo, además, usar las publicaciones de vigilancia de rating "ratings outlooks" y "ratings watches" para mejorar las precisiones de los SR (Cantor & Mann, 2007). A pesar de ello, las CRA siguen mostrando un comportamiento asimétrico en la determinación de los SR¹¹. Broto y Molina (2016) analizaron y definieron el "ciclo de rating" en el caso de la deuda soberana, concluyendo que las degradaciones tienden a ser más profundas y rápidas que las subidas, y midieron la respuesta asimétrica a las variables fundamentales en las subidas y las bajadas. Por su parte Ayed et al. (2018)

¹¹En estadística, este tipo de comportamiento se conoce como función de pérdida asimétrica. Hay estudios que evidencian que distintas predicciones tienen una función de pérdida asimétrica. Un análisis realizado sobre el comportamiento de los analistas en sus predicciones sobre beneficios futuros concluye que hay asimetría en sus predicciones (Clatworthy et al., 2012), otro análisis similar se ha realizado recientemente con las predicciones de la Reserva Federal de EE. UU. y similares conclusiones (Caunedo J et al., 2020). Dado que el rating es una predicción sobre la solvencia futura, es posible inferir que la relación de los ratings con la información económico-financiera que los sustenta no es necesariamente una función lineal y, por tanto, pueden aparecer comportamientos asimétricos en los movimientos a lo largo de las escalas de calificación crediticia.

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

analizaron las duraciones de las variaciones de ratings a partir de modelos de duración condicional autoregresiva y encontraron que en los periodos de crisis los episodios de rating tendían a ser más cortos y eran más probables las bajadas y que las probabilidades de bajada no solo dependían del historial de rating sino que también dependían del contexto económico.

Por último, aunque las CRA publican los componentes de los SR, el comité de calificación suele introducir modificaciones arbitrarias adicionales (Reusens & Croux, 2017). Por ello, la literatura empírica sobre determinantes del SR apunta que la identificación de la parte de información dura o variables cuantitativas sobre las que se emite una calificación crediticia es relativamente sencilla, mientras que la parte cualitativa es mucho más compleja, empleándose distintas aproximaciones (Afonso, 2003; Afonso et al., 2009; Butler & Fauver, 2006; R. Cantor & Packer, 1996; Ozturk, 2014; Slapnik, 2019). Esto no hace sino reforzar la crítica hacia la transparencia de los modelos de las CRA, lo que unido a las críticas ya apuntadas con, promueve la estimación de SR por agencias independientes (Dimitriadou et al., 2021). Esto ha sido posible gracias a la ocurrencia paralela de otros fenómenos, tales como i) la abundancia y accesibilidad a datos pertinentes; ii) un procesamiento informático más barato y rápido; iii) el progreso significativo en las técnicas de previsión tradicionales; y iv) el amplio uso de nuevas técnicas de previsión, esto es, sistemas de aprendizaje automático e inteligencia artificial.

3. UNA PROPUESTA DE CALIFICACIÓN DE RIESGO SOBERANO

Considerando el contexto descrito en las secciones precedentes, el presente apartado ofrece una propuesta para la determinación del riesgo soberano. Para ello, en primer lugar, se revisa la literatura empírica sobre determinantes del riesgo soberano, prestando especial atención a los factores que explican los SR, así como las técnicas empleadas para su análisis. A continuación, se expondrá una propuesta de modelo para la determinación de los SR, resaltando especialmente sus fundamentos teóricos, que sustentarán la aplicación empírica en los siguientes capítulos. Finalmente, se explicita la mejora implementada en el modelo con el objetivo de mejorar las estimaciones bajo el supuesto de comportamiento asimétrico de las CRA.

3.1 . Revisión de la literatura

La literatura empírica sobre los determinantes de los SR tiene su origen en el trabajo seminal de Cantor y Packer (1996), donde se analizan datos de Moody's y S&P asignados a 23 países industrializados y 26 países en desarrollo el 29 de septiembre de 1995. El modelo de estimación, mediante mínimos cuadrados ordinarios, alcanza un R-cuadrado ajustado de 0,92 para la media de los ratings

Tabla 5: Trabajos sobre determinantes de SR. 1996-2018

Autores	Año	Título	Revista	Alcance	Periodo	Métodos utilizados	Agecias de Rating				Resultados	
							Clases	Moody's	S&P	Fitch	Variables	R cuadrado
WHOLE SAMPLE (TRAIN=TEST)												
Cantor & Parker	1996	Determinants and impact of sovereign credit ratings	Economic Policy Review, Federal Reserve Bank of New York	Mundial	1995	OLS	16	1	1		8	0,92
Monfort Mulder	2000	Using credit ratings for capital requirements on lending to emerging market economies	IMF Working Papers	Emergentes	1993-1999	OLS	21	1	1		18	0,82
Eliasson	2002	Sovereign credit ratings	Working Papers Deutsche Bank	Emergentes	1990-1999	OLS	25		1		9	0,93
Hu et all	2002	The estimation of transition matrices for sovereign credit ratings.	Journal of Banking & Finance,	Mundial	1981-1988	Ordered probit estimation Transition Matrix	14		1		12	0,82
Afonso	2003	Understanding the determinants of sovereign debt ratings: evidence for the two leading agencies	Journal of Economics and Finance	Mundial	2001	Logistic & Esponential transf.	16	1	1		7	0,86
Alexe	2003	A non-recursive regression model for country risk rating	Rutgers Center for Operational Research	Mundial	1999	OLS non recursive multiregression	21		1		12	89,3
Borio & Packer	2004	Assessing new perspectives on country risk	BIS Quarterly Review	Mundial	1996-2003	OLS	17	1	1		18	0,95
Canuto et all	2004	Macroeconomics and Sovereign Risk Ratings	mimeo.	Mundial	1998-2002	OLS	24	1	1	1	9	0,88
Bsissondoyal	2005	An analysis of the determinants of sovereign ratings	Global Finance Journal	Mundial	1995-1999	ordered probit estimation	9	1	1		11	0,54
Bsissondoyal	2006	Determinants of sovereign ratings: A comparison of case-based reasoning and ordered probit approaches	Global Finance Journal	Mundial	2002	ordered probit estimation case-based reasoning	9	1	1	1	8	0,66
Mora	2006	Sovereign credit ratings: Guilty beyond reasonable doubt?	Journal of Banking & Finance,	Emergentes	1986-2001	fix & random efects x Linear & ordered probit estimation	21	1	1		10	0,78
Afonso Gomes Rother	2007	Sovereign credit ratings	European Central Bank Working Paper	Mundial	1970-2005	ordered probit y random efects oredered probit estimation	17	1	1	1	11	0,7
Afonso et al	2011	Short- and Long-run Determinants of Sovereign Debt Credit Ratings	International Journal of Finance and Economics	Mundial	1995-2005	ordered probit y random efects oredered probit estimation	16	1	1	1	11	0,95
Afonso et al	2011	Do Fiscal Imbalances Deteriorate Sovereign Debt Ratings?	Revue économique	Mundial	2008-2011	random efects, ordered probit & random efects ordered probit	17	1	1	1	6	0,7

CAPÍTULO 2: Modelo de estimación de ratings de deuda soberana

M Martin-G et al	2013	Evolution of Sovereign Rating Models in the Current Crisis	Globalización, Competitividad y Gobernabilidad de Georgetown/Universia	Mundial	2004-2011	OLS	17	1	1	1	12	0,82
Basu et al	2013	Sovereign Ratings in the Post-Crisis World: An Analysis of Actual, Shadow and Relative Risk Ratings	Policy Research working paper. The World Bank	Mundial	2008-2012	OLS, pooled, Ordered probit	60	1	1	1	12	0,9
Broto et al	2014	SOVEREIGN RATINGS AND THEIR ASYMMETRIC RESPONSE TO FUNDAMENTALS	Journal of Economic Behavior and Organization	Mundial	1994-2013	ordered logit model, pooled OLS	21		1			0,92
Boumparis et al	2015	Has the crisis affected the behavior of the rating agencies? Panel evidence from the Eurozone	Economic Letters	Eurozona	2002-2013	pooled OLS, fixed effects y random effects	21	1	1	1	10	0,9
Miricescu et al	2016	THE DETERMINANTS OF THE SOVEREIGN DEBT RATING: EVIDENCE FOR THE EUROPEAN UNION COUNTRIES	Economic Computation and Economic Cybernetics Studies and Research	EU25	2005-2012	logistic transformation random effect Afonso (2011)	14		1		10	0,9
De Moor et al	2018	Subjectivity in sovereign credit ratings	Journal of Banking and Finance	Mundial	1995-2014	logit random effects, random forest	21	1	1	1	13	0.51-0.58 0.92-0.93

OUT OF SAMPLE (TRAIN#TEST)

Benell et al	2006	Modelling sovereign credit ratings: Neural networks versus ordered probit	Expert System with Applications	Mundial	1989-1999	neural network, ordered probit	16	1	1	1	8	0,384-0,453
Van Gestel et al	2006	A process model to develop an internal rating system: Sovereign credit ratings	Decision Support System	Mundial	1997-2002	ordinal linear regression, SVM	17	1			13	0,397-0,448
Fernandez Navarro et al	2013	Addressing the EU Sovereign Ratings Using an Ordinal Regression Approach	IEEE Transactions on Cybernetix	EU27	2007-2010	ordinal regression, MLR, RBFN, MLP, SVM, POM, SVMRank, SVOR, Ensemble Aproaches	5		1		11	0,63-0,85
S Moned. et al	2014	A guided data projection technique for classification of sovereign ratings: The case of European Union 27	Applied Soft Computing	EU27	2007-2010	Pairwise Class Distances projection	5	1	1	1	11	0,37-0,85
Huseyin Ozturk et al	2016	Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample	Economic Modelling	Mundial	1999-2010	SvM, MLP, Classif & regr trees, Bayesian networks, Naive Bayes	16	1			9	0,34-0,79

Fuente : Elaboración propia

de las dos agencias, considerando un total de 8 variables explicativas¹². El enfoque propuesto no resultó tan eficiente para marcos temporales distintos, lo que motivó análisis posteriores en los que se ha continuado indagando, tanto sobre las variables explicativas, como sobre los métodos de estimación. La Tabla 5 resume las principales características de los trabajos más relevantes para el periodo 1996-2018.

La revisión de la literatura no ha permitido encontrar patrón alguno en la selección de los países soberanos objeto de estudio o los periodos temporales analizados. En algunos trabajos se opta por el ámbito mundial y periodos de varios años (Bennell et al., 2006; Broto & Molina, 2016; Canuto et al., 2012; Hu et al., 2002), mientras que otros se focalizan en grupos de países más homogéneos, como pueden ser países emergentes (Eliasson, 2002; Mora, 2006; Mulder & Monfort, 2000), pertenecientes a la Unión Europea (Fernández-Navarro et al., 2013; E.-C. Miricescu et al., 2016; Sánchez-Monedero et al., 2014) o incluso a la Eurozona (Boumparis et al., 2015), y periodos de un año (Afonso, 2003; R. Cantor & Packer, 1996).

En cuanto a la metodología utilizada es común el hacer varias aproximaciones en el mismo trabajo. El modelo inicial de mínimos cuadrados ordinarios ha ido evolucionando (Borio & Packer, 2004; R. Cantor & Packer, 1996; Canuto et al., 2012; Eliasson, 2002; Martín-García et al., 2014; Mulder & Monfort, 2000). Por una parte, complicándose e introduciendo variantes más complejas como transformaciones logísticas y exponenciales (Afonso, 2003), regresiones no

¹² PIB per cápita, crecimiento real del PIB, inflación, déficit (o superávit), balanza de pagos, deuda externa, desarrollo económico, e historial de “default”

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

recursivas (Alexe et al., 2003) o mínimos cuadrados agrupados (Basu et al, 2013; Boumpais et al., 2015; Broto y Molina, 2016;). Otra evolución importante la introducción de modelos de elección discreta, en su mayor parte probit ordenado, tanto con efectos fijos como con efectos aleatorios (Afonso et al., 2011; Afonso & Gomes, 2011; Bissoondoyal-Bheenick, 2005; Bissoondoyal-Bheenick et al., 2006; Hu et al., 2002; E.-C. Miricescu et al., 2016; Mora, 2006; Van Gestel et al., 2006). Sólo muy recientemente y, en pocos artículos, se ha empezado a utilizar metodologías de inteligencia artificial (Fernández-Navarro et al., 2013; Ozturk et al., 2016; Sánchez-Monedero et al., 2014), no distinguiendo en algún caso entre los grupos de entrenamiento y generalización (De Moor et al., 2018).

Tampoco existe uniformidad a la hora de definir las clases, con independencia de usar transformaciones lineales, logísticas u ordenadas. Algunos autores optan por limitarse a las clases principales sin tener en cuenta los modificadores que tienen algunas clases, de forma que trabajan con variables explicadas con 9 etiquetas (Bissoondoyal-Bheenick, 2005; Bissoondoyal-Bheenick et al., 2006). Otros, la mayoría, incluyen los modificadores que existen para algunas calificaciones, pero agrupan las situadas en escalas inferiores, obteniendo un número que oscila entre las 14 y las 25 etiquetas (Eliasson, 2002; Hu et al., 2002). Por último, hay quién se decanta por incluir la perspectiva como un modificador de cada subclase llegando a las 60 clases (Basu et al., 2013).

No existe, tampoco, homogeneidad a la hora de considerar las distintas CRA. Hay autores que sólo utilizan los ratings de una (Fernández-Navarro et al., 2013; C. Miricescu, 2014), los de dos (Afonso, 2003; Mora, 2006) o los de las tres agencias principales (Bennell et al., 2006; De Moor et al., 2018; Sánchez-Monedero et al., 2014). Cuando se consideran los SR de más de una CRA el

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

tratamiento tampoco es uniforme: en algunos trabajos se realiza una estimación del SR promedio de las CRA consideradas (Borio & Packer, 2004); en otros, se estima un modelo para cada agencia (Boumparis et al., 2015; Sánchez-Monedero et al., 2014); o bien se consideran ambas opciones (Canuto et al., 2012; Martín-García et al., 2014).

Pero si en algo existen realmente discrepancias es a la hora de seleccionar las variables explicativas de los SR (véase antepenúltima columna de la Tabla 5). En la Tabla 6 se recoge la totalidad de las variables empleadas¹³ en los trabajos revisados, siendo éstas 41, y la frecuencia con la que son consideradas como relevantes en dichas investigaciones, lo que oscila entre 1 y 24 ocasiones.

Tabla 6: Variables detectadas en la revisión de la literatura y su frecuencia

Variable	Frec.	Variable	Frec.	Variable	Frec.
Inflación	24	Importaciones	6	Previsión PIB	1
PIB per cápita	23	Región	6	Invers. Extranjera	1
Crecimiento PIB	21	Deuda a CP	5	Crisis	1
Déficit	19	Índice Corrupción	5	Indep. B. Central	1
Deuda	16	Tipo interés	5	Calid. Instituciones	1
Deuda externa	16	Tipo de cambio	4	Termino Comercio.	1
Reservas Int.	15	Servicio deuda	3	Riesgo financiero	1
Exportaciones	14	Crédito	3	Relación con EEUU	1
Balance CC	14	Prima de riesgo	3	Remesas	1
Hist. Default	10	Costes lab. unit.	3	Gasto en salud	1
Desarrollo	9	Riesgo político	3	Escuela secundaria	1
Desempleo	9	Imperio Ley	3	Precio petróleo	1
Efect. Gobierno	8	Ahorros e Inv.	3	Número móviles	1
Balanza de pagos	7	PIB	2		

Fuente: elaboración propia

¹³ En el recuento se ha ignorado las transformaciones logarítmicas de las variables; y, en caso de uso de ratios, se han considerado cada operador de la ratio.

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

Como puede observarse, ninguna variable es usada en la totalidad de las investigaciones revisadas. Aun así, la variable inflación es la más usada, siendo considerada en 24 de los 25 trabajos; la deuda en relación al PIB, como indicador del endeudamiento de un país, es utilizada en 16 de las 25 investigaciones. Las previsiones de crecimiento del PIB, de gran importancia para conocer la evolución futura de la economía, y por tanto la capacidad de pago de un país, sólo se utilizan en un artículo. Lo anterior pone de manifiesto que no existe una motivación clara en la selección de variables explicativas, sino que estas han sido elegidas en función del grado de ajuste y significación estadística.

3.2 La solvencia y la dinámica de la deuda pública

La solvencia de un estado soberano puede entenderse como la capacidad de responder a sus obligaciones financieras en tiempo y forma (Afonso et al., 2007), dependiendo para ello de la evolución futura de su actividad económica. Por tanto, un país puede ser solvente aunque su deuda nominal aumente indefinidamente, siempre y cuando la relación entre esta deuda y su PIB se sitúe en un porcentaje razonable (Agenor y Montiel, 2008). Esta afirmación tiene su reflejo práctico, por ejemplo, en el artículo 126 del Tratado de Funcionamiento de la Unión Europea que, para los estados miembros, establece la obligación de mantener la deuda por debajo del 60% respecto al PIB (Versión Consolidada Del Tratado de Funcionamiento de La Unión Europea, 2012).

El estudio de los niveles óptimos de deuda pública y su evolución en el tiempo por reconocidos economistas ha puesto de manifiesto dos aspectos relevantes (Domar, 1994; Tobin, 1984, Blanchard, 1990; Lane 1992; Dornbusch, 1997;

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

Canzoneri et al., 2001; Cherif y Hasanov 2012). Por un lado, los mecanismos de corrección de niveles excesivos quedan condicionados por la independencia de las autoridades monetarias (bancos centrales). Si esta es elevada, la política monetaria será el instrumento empleado para reconducir los niveles de deuda, normalmente ocasionando subidas generalizadas de precios; en caso contrario, la política fiscal será el instrumento más idóneo. De cara a un inversor la independencia del banco central es, por tanto, clave para determinar la seguridad de su inversión. La posibilidad de cobrar al vencimiento en una moneda devaluada supone un riesgo significativo (en caso de deuda en moneda nacional) o de no cobrar por las dificultades de obtener divisas por la devaluación de la moneda (en caso de deuda en moneda extranjera).

Por otro, la dependencia de la ratio deuda/PIB de la evolución de cinco magnitudes: i) el nivel de deuda respecto al PIB, que proporciona información sobre el nivel de endeudamiento de un país y determina el servicio de la deuda; ii) el tipo de interés de la deuda que, además de determinar el servicio de la deuda, indica, en términos relativos y comparado con el exigible a otras economías, la percepción del mercado del riesgo de insolvencia de un país; iii) el excedente primario que, junto al servicio de la deuda, informa sobre la necesidad o exceso de financiación de un país; iv) la tasa de crecimiento de la economía, que informa sobre la capacidad de un país de reducir el peso de su deuda relativa mediante el crecimiento; y, finalmente v) la tasa de inflación, que informa de los efectos de la corrección de precios en el valor relativo de la deuda. Si bien todos esos factores son considerados en las investigaciones sobre determinantes de los SR, el caso de la inflación la mayoría de los economistas consideran que el déficit presupuestario anunciado por los gobiernos está

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

sobrestimado en una cuantía equivalente al valor actualizado del stock de deuda pública, empleando para ello la tasa de inflación correspondiente. (Mankiw, 2014). Eso previene el uso de la variable inflación pues ya es considerada en a través del déficit presupuestario.

En las investigaciones previamente analizadas está implícita la idea de sostenibilidad de la deuda, que es desarrollada desde el punto de vista teórico por Blanchard et al. (1990), partiendo del concepto de restricción presupuestaria del sector público:

$$D = \Delta B \vee \Delta H \cong 0 \quad (1)$$

Donde D representa el déficit público en términos nominales, B la deuda pública en términos nominales e ΔH el crecimiento de la masa monetaria. Según esta ecuación el déficit público nominal sólo se podría financiar con incrementos de deuda por parte del estado.

El déficit público se puede desglosar en la parte correspondiente al pago de intereses, que equivale a la deuda en circulación por el tipo de interés aplicable Bi , y la parte resultante de la diferencia entre la recaudación por impuestos y otros ingresos del estado y los gastos, inversiones públicas y transferencias del estado excluido el pago de intereses, esto es, el excedente primario (X). La ecuación anterior se puede reformular de la siguiente manera

$$D = \Delta B = Bi - X \quad (2)$$

La variación de la deuda pública, esto es, el déficit, es el resultado de restar a los pagos por intereses el excedente primario.

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

Por otra parte, la ecuación (3) muestra la relación entre la deuda pública B y el PIB nominal PY (siendo P el nivel de precios e Y la renta), denotada mediante b , mientras que la ecuación (4) permite conocer su evolución

$$b = \frac{B}{PY} \quad (3)$$

$$\Delta b = \Delta \frac{B}{PY} = \frac{\Delta B * PY - B * \Delta PY}{PY^2} \quad (4)$$

Realizando las oportunas transformaciones, la variación de la deuda pública podría expresarse como:

$$\Delta b = \frac{\Delta B}{PY} - \frac{B}{PY} \left(\frac{\Delta P}{P} + \frac{\Delta Y}{Y} \right) \quad (5)$$

Sustituyendo en la ecuación (5), la variación de los precios de una economía por el deflactor del PNB $\left(\pi = \frac{\Delta P}{P} \right)$ y denotando el crecimiento real de la economía como g , es posible obtener una expresión más compacta de la variación de la deuda pública (6):

$$\Delta b = \frac{\Delta B}{PY} - b(\pi + g) \quad (6)$$

Teniendo en consideración el déficit público, definido en la ecuación 2, y su relación con el PIB nominal, la variación de la deuda pública quedaría definida por la siguiente ecuación:

$$\Delta b = ib - x - b(\pi + g) \quad (7)$$

Así, en (7) es posible distinguir tres elementos: los pagos nominales de intereses, la contribución a la estabilización de la deuda por parte del excedente primario; y la contribución por parte del crecimiento de la renta nominal. Los dos

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

primeros están relacionados con la variación de la deuda real, esto es B , y el tercero con la variación del PIB (PY). En caso de utilizar el tipo de interés real ($r = i - \pi$) la variación de la deuda vendría dada por:

$$\Delta b = -x + b(i - \pi - g) = -x + b(r - g) \quad (8)$$

Esta ecuación indica que cuando el interés real es superior a la tasa de crecimiento es necesario generar un excedente primario para poder mantener el nivel de deuda (Blanchard, 1990; Dornbusch et al., 1991; Spaventa, 1987).

En el caso de economías estables con bancos centrales independientes, monedas estables y tasas de inflación moderadas se puede hacer el mismo análisis, pero simplificando y sustituyendo la evolución del crecimiento del PIB real y la inflación por la evolución del PIB nominal. Bajo este supuesto, la variación de la deuda quedaría reflejada en la ecuación (9), donde g_N refleja la variación del PIB nominal de un ejercicio al siguiente y el PIB nominal del ejercicio en cuestión:

$$\Delta b = bi - x - bg_N \quad (9)$$

Como los propios autores reconocen la sostenibilidad de la deuda está vinculada a la gestión de las cuentas públicas de manera análoga a la gestión de un presupuesto familiar, aunque con una salvedad importante (Blanchard et al., 1990): la duración del periodo de amortización de la deuda. Así, mientras que las personas tienen una esperanza de vida finita, los estados no tienen previsión de caducidad; lo que se traduce en la exigencia de cancelación de la deuda a su vencimiento en el primer caso, frente a la exigencia de mantener un nivel razonable de endeudamiento. Por otra parte, esta comparación entre la solvencia de las personas físicas y de los estados resulta de utilidad para la

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

determinación de los factores determinantes de la solvencia. La Tabla 7 resumen la comparativa entre las variables consideradas en la determinación de la capacidad de pago de una persona física y de un Estado:

Tabla 7: Solvencia de la persona física respecto a la solvencia de un estado soberano

Persona física	Estado Soberano
% Renta dedicado a pago de préstamos	% Renta respecto al PIB
Renta neta mensual	PIB
Situación personal y familiar	PIB per cápita
Previsible evolución de los ingresos	Evolución del % deuda respecto al PIB
Seguridad de los ingresos	Resiliencia
Garantías patrimoniales	N/A*

* Las garantías patrimoniales no son aplicables a la deuda soberana pues los bienes de un estado no son reclamables por el acreedor en caso de impago.

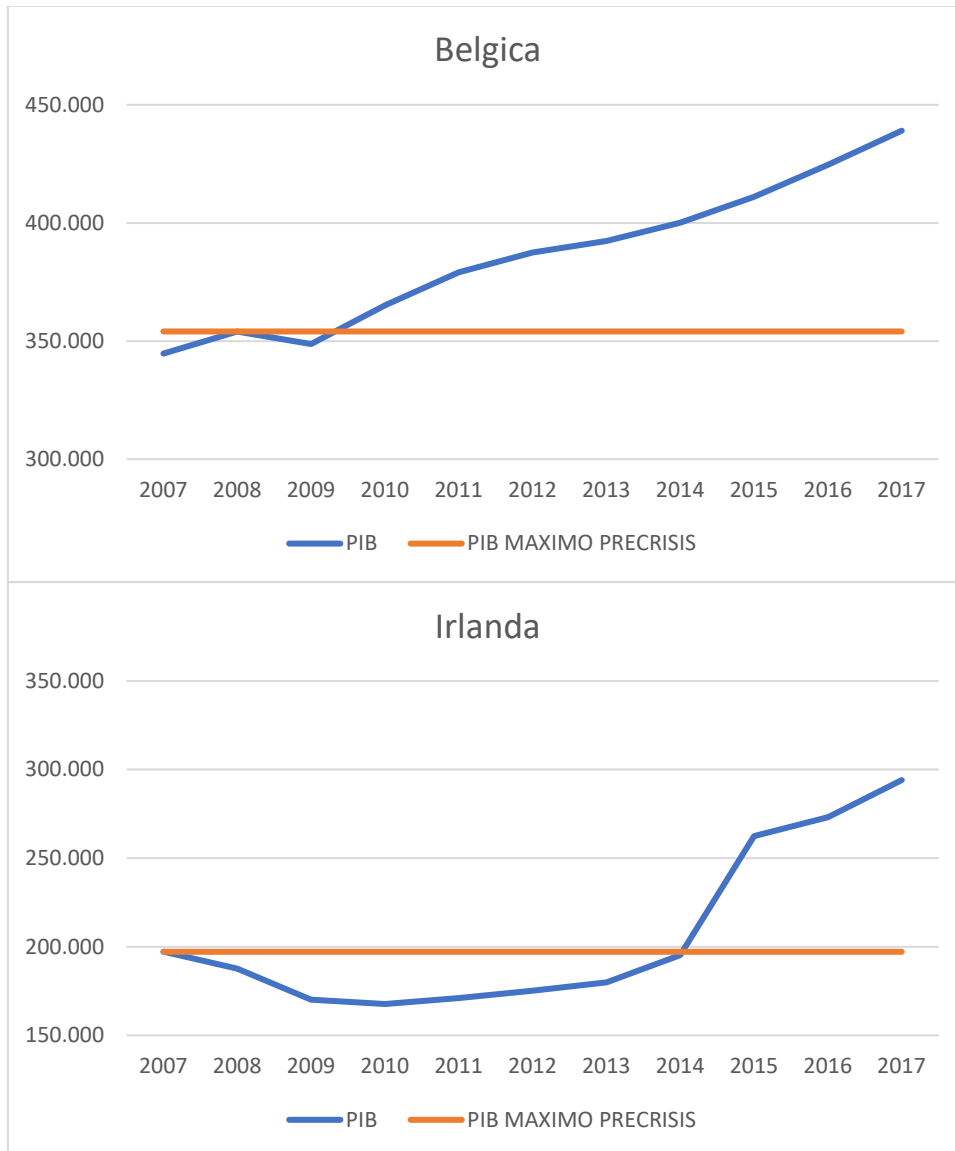
Fuente: elaboración propia

Esta perspectiva incorpora otros dos factores explicativos para la estimación de los SR: el PIB per cápita y la idea de resiliencia. A pesar de que el PIB es una medida denostada por sus limitaciones para medir la riqueza de un país, sigue siendo el único estándar comparable internacionalmente. Sin embargo es necesaria su consideración en términos per cápita para realizar comparaciones de nivel de riqueza entre distintos países (Afonso et al., 2007; Bennell et al., 2006; Erdem & Varli, 2014; Gültekin-Karakaş et al., 2011; Mora, 2006). Por su parte, la resiliencia se concibe como la magnitud y duración de la desviación del nivel de rendimiento de un sistema (Vugrin et al., 2010). Para evaluar la resiliencia económica, primero es necesario comprender el nivel de rendimiento de la economía. Dada la falta de un entendimiento común sobre el nivel de rendimiento de la economía, la resiliencia económica se suele equiparar al nivel de rendimiento del PIB (Briguglio et al., 2009; Oprea et al., 2020; Pontarollo & Serpieri, 2020). La crisis financiera mundial resultó ser un experimento natural para evaluar la resiliencia económica de los países, dada la disparidad en la

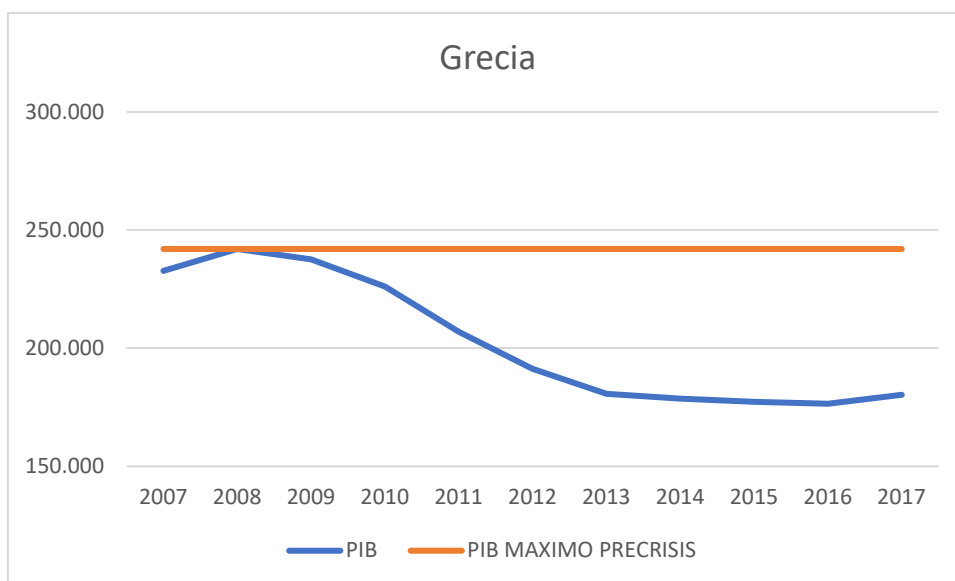
CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

evolución de su PIB, tal y como se muestra en los ejemplos recogidos en el Gráfico 2.

Gráfico 2: Evolución del PIB en distintos países europeos tras la crisis de 2008



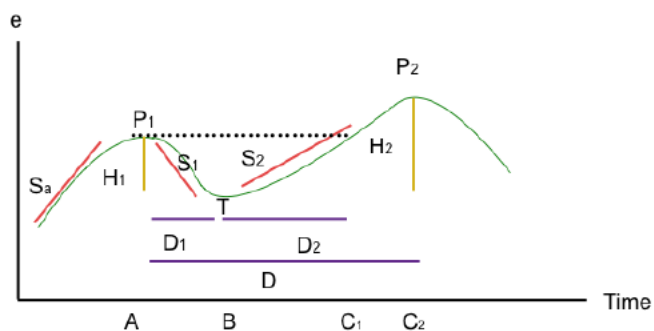
CAPÍTULO 2: Modelo de estimación de rating de deuda soberana



Fuente: elaboración propia a partir de datos de Eurostat

Para hacer operativa el concepto de resiliencia económica basado en la evolución del PIB, se acude al modelo de análisis de resiliencia temporal (Gráfico X) recogido en el programa “European Observation Network for Territorial Development and Cohesión” (ESPON):

Gráfico 3: Modelo de resiliencia ESPON



Fuente ESPON (2014)

En este modelo la recuperación tiene lugar en el punto C_1 cuando el PIB recupera el nivel previo a la crisis P_1 . El tiempo de recuperación viene dado por D_1+D_2 mientras que la totalidad del ciclo se representa por la distancia D (ESPON, 2014). Cuando menor sea la distancia D mayor capacidad de recuperación tendrá una economía y por tanto la exposición a una crisis económica será

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

menor.

Siguiendo a Manca et al. (2017), es posible calcular un indicador acumulativo de resiliencia que aproxima el riesgo de incrementar el déficit en el supuesto de un impacto externo. Para ello, en primer lugar, se define P_1 para cada país i como el PIB máximo previo a la crisis (MGP_i):

$$MGP_i = \max (PIB_{i2002} - PIB_{i2008}) \quad (10)$$

Posteriormente, a medida que pasen los años y mientras no se llegue a la recuperación en C_1 , el indicador va incrementándose. Para que los datos sean comparables se establece una relación con el PIB nominal del año analizado.

$$CR_{it} = (\sum_n^t MGP_i - PIB_{it})/PIB_{it} \quad (11)$$
$$\forall PIB < MGP; \forall t > n$$

3.3. Modelo teórico para la estimación de los SR

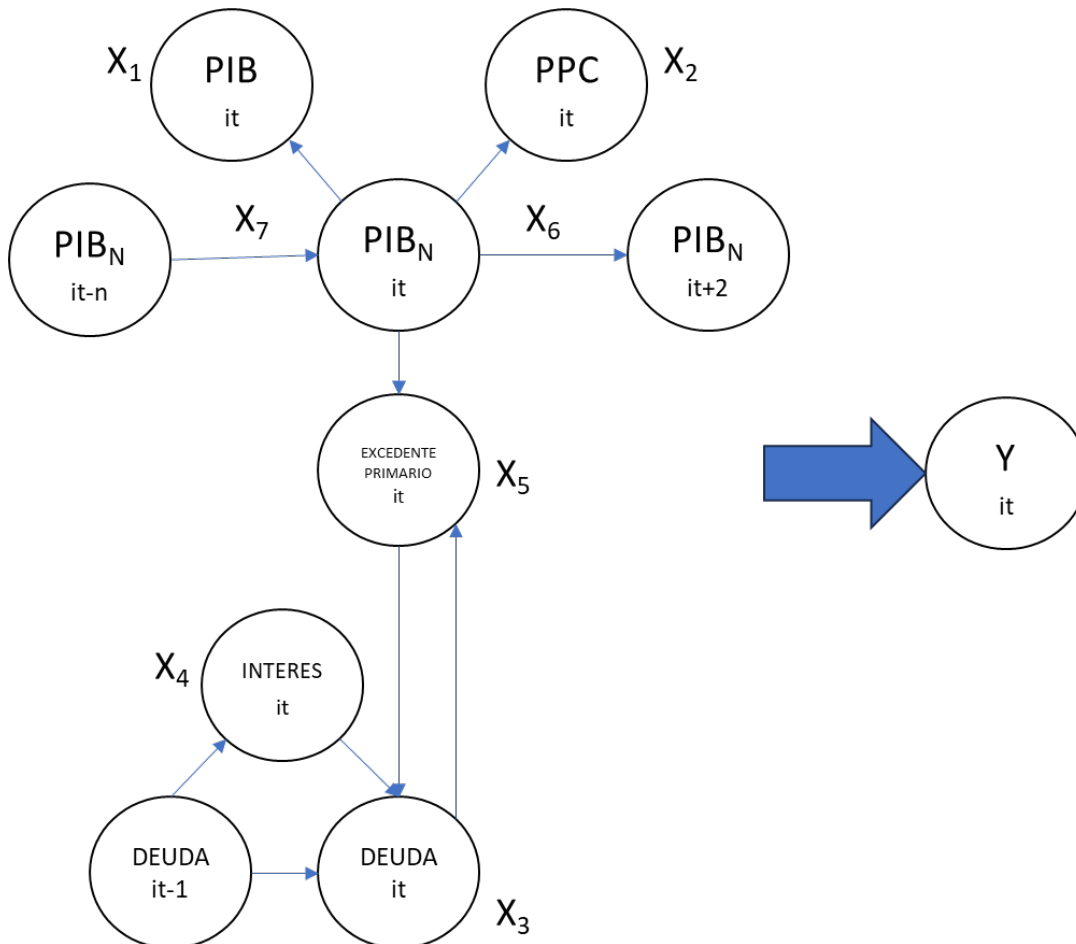
Los fundamentos teóricos explicitados en el apartado precedente sirven de guía para la elaboración del modelo de estimación de las SR, intentando superar así las limitaciones apuntadas en la revisión de la literatura empírica. En este modelo la variable explicada es la calificación de riesgo soberano (SR), que se determina a partir de un número limitado de variables explicativas, en total siete:

$$Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7) \quad (12)$$

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

EL Gráfico 4 muestra la representación del modelo, recogiendo tanto las variables, como las relaciones existentes entre las mismas:

Gráfico 4: Representación gráfica del modelo propuesto



Fuente: elaboración propia

Donde:

Y_{it} : Calificación crediticia asignada por la agencia Moody's a la deuda emitida en moneda extranjera para el país i en el ejercicio t .

$X_1 = PIB_{it}$. Definido como el Producto Interior Bruto el país i en el ejercicio t expresado en euros constantes.

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

$X_2 = PPC_{it}$. Es el Producto Interior Bruto Per Cápita en el del país i ejercicio t expresado en euros constantes.

$X_3 = b_{it} = \frac{B_{it}}{PIB_{Nit}}$. En la que b representa la relación entre la deuda calculada según el protocolo de déficit excesivo B y el PIB nominal PIB_N en el ejercicio t del país i .

$X_4 = i_{maxit}$. Máximo de la serie diaria del tipo de interés de los bonos a 10 años del país i en el periodo t .

$X_5 = x_{it}$. Porcentaje del excedente primario con respecto al PIB nominal del ejercicio t para el país i .

$X_6 = g_{N(t+2)it}$. Previsión¹⁴ de crecimiento del PIB nominal acumulado para los dos ejercicios siguientes al ejercicio t para el país i .

$X_7 = CR_t$. Indicador acumulativo de resiliencia calculado según Manca et al., (2017) para el ejercicio t del país i .

Para determinar el rating se utilizan las variables que se han identificado para asimilar el modelo de determinación de solvencia para un país al modelo usado para una persona física. El PIB y el PIB per cápita se equipararían a la renta neta mensual y la situación personal y familiar de una persona física. La relación entre la deuda y el PIB sería asimilable al porcentaje de la renta mensual que se dedica al pago de préstamos. El excedente primario, el tipo de interés y la previsión futura de crecimiento nominal del PIB sirven para calcular una previsible

¹⁴ La previsión de crecimiento nominal del PIB es el efecto combinado de la previsión del crecimiento real de la economía y del crecimiento de los precios.

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

evolución del endeudamiento. Por último, la resiliencia económica sería equivalente a la seguridad esperada de los ingresos futuros.

Para que los datos tengan una coherencia temporal se ha intentado utilizar un marco temporal en el que la información analizada esté disponible. Para el rating se ha considerado que toda la información presupuestaria está disponible el 30 de abril del ejercicio siguiente y se ha tomado esa fecha para su obtención. Los datos de deuda, PIB, PIB per cápita, excedente primario y los datos para el cálculo de la resiliencia se toman a 31 de diciembre del año anterior. Los datos de previsiones se han obtenido del informe de primavera de la Comisión Europea disponibles en la fecha elegida para toma de datos de rating. Por último, el dato referido al tipo de interés al ser un dato con una mayor disponibilidad se ha calculado partiendo de una antelación de 15 días a la fecha de toma de datos de rating; al ser este dato muy variable siendo lo más preocupante para determinar un rating que este valor sea alto se ha optado por utilizar el máximo del periodo temporal analizado.

3.4. El filtrado de resultados

Como se ha mencionado, el uso del enfoque TTC por parte de las agencias ha dificultado la replicación de sus resultados en cuanto asignación de SR, especialmente en el caso de aquellos trabajos de investigación que emplean técnicas de clasificación¹⁵. Para incluir esta asimetría en los cambios de rating la solución más evidente es considerar como variable del modelo la calificación del ejercicio anterior, pero es muy difícil que en una serie de carácter

¹⁵ La mayoría de las técnicas de aprendizaje automático son modelos con enfoque PIT.

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

principalmente estable el modelo prediga los cambios de categoría. Adicionalmente, si el procedimiento es incluir la clasificación del ejercicio anterior como variable independiente, implicaría considerar en el modelo una clasificación realizada a partir de los datos del ejercicio anterior que normalmente no son muy distintos a los del ejercicio a analizar. Salvo conjuntos de países en los que existan muchos cambios de rating el modelo tendería a asignar como rating el rating del ejercicio anterior y proporcionaría muy buenos resultados, pero no tendría en cuenta el resto de variables que interesa analizar. Supondría demasiada información de las agencias de rating como variable independiente, el rating se ha empelado se en contadas ocasiones y para conjuntos de países emergentes con un resultado no muy superior al obtenido con otras variables (Mulder & Monfort, 2000).

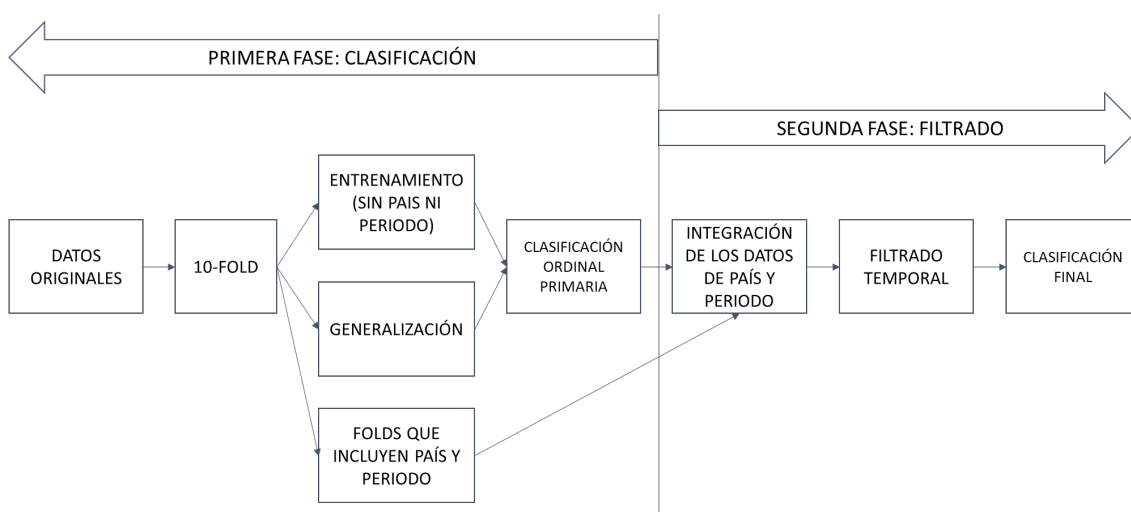
Como alternativa para evaluar el comportamiento de las agencias de rating aparece la técnica del filtrado de los resultados de una clasificación previa (Cantor y Mann, 2007; Kiff and Schumacher, 2013). Con esta técnica los datos clasificados se transforman en función de la clasificación anterior, permitiendo cambios sólo en el caso de que se haya alcanzado algún umbral o desfase temporal. De esta forma, resulta posible incluir una política de migración de clase en las estimaciones de SR.

En la práctica supone un proceso de clasificación de dos etapas:

- Primera etapa - clasificación ordinal: hacer una clasificación ordinal sin información sobre la serie temporal.
- Segunda etapa - filtrado: incluir la información del país y del año en los datos clasificados en la primera etapa y realizar una transformación para incluir los retrasos en la clasificación.

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

Gráfico 5: Proceso de clasificación y filtrado



Fuente: elaboración propia

Una vez que comparamos la clasificación final con la clasificación ordinal primaria podemos concluir si esta clasificación transformada supera a la anterior y, por tanto, la solución propuesta mejora la clasificación original.

Esta función de filtrado puede usarse tanto para valores numéricos como temporales. En estos casos los cambios de ratings no son automáticos ni inmediatos. En el primer caso, siempre que exista un indicador numeral que se transforme en rating, se exige superar un umbral por encima del límite inferior de la clase a la que se aspira para subir de clase y por tanto hay que estar a x de diferencia sobre el umbral mínimo de la clase superior, mientras que para bajar de clase hay que estar a y de diferencia con el límite superior de la clase inferior; si por ejemplo la máxima categoría exige un 9 en una escala de 0 a 10 el que provenga de una categoría inferior podría no ascender mientras no supere un umbral que podría ser 9,2, igualmente si está en la clase superior solo bajaría si bajara de 8,9. De esta forma, la asimetría en el comportamiento del rating se define como la diferencia entre x e y , en el ejemplo sería la diferencia entre el

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

margen de subida 0,2 y el de bajada 0,1. Este comportamiento permite una mayor estabilidad en el rating como pretenden las agencias; por una parte tiene cierta flexibilidad al permitir estar en el borde de la categoría actual sin cambiar de rating y un restricción para el cambio de rating al exigir un margen para que se produzca. En el segundo caso solo se tiene en cuenta el efecto temporal del retraso que existe entre el momento en el que se cumplen los requisitos de pertenecer a la clase superior o inferior y cuando se produce realmente la subida y la bajada de clase. Para subir de clase hay que estar un periodo adicional t_1 en valores que corresponden a la clase superior y para bajar de clase hay que estar un periodo adicional t_2 en valores que corresponden a la clase inferior (la asimetría la determina la diferencia entre t_1 y t_2).

Dada la naturaleza ordinal de los SR, se aplica un filtrado temporal. Para ello, se estima una función que determine el tiempo que transcurre en el cambio de SR asignado a un país, una vez que dicho país cumple con las condiciones necesarias para la asignación de la calificación correspondiente a la nueva o clase dentro de la escala de calificación de la correspondiente agencia:

$$TR_{i \rightarrow j} = \tau(C_i, C_j) \quad (13)$$

Donde TR sería el tiempo que tarda un rating de migrar de la clase i a la clase j una vez cumple los parámetros de la clase j . Se podría estimar una matriz de rango $q-1$ (siendo q el número de clases) que informara de los tiempos de cambio entre clases para las subidas y otra similar para las bajadas.

Una forma de simplificar el problema es el propuesto por Broto y Molina (2016) que consiste en distinguir solamente entre el comportamiento de las subidas (upgrades - UTR) y bajadas (downgrades - DTR):

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

$$UTR_{i \rightarrow j} = \tau(C_i, C_j); \forall i < j \quad (14)$$

$$DTR_{i \rightarrow j} = \tau'(C_i, C_j); \forall i > j \quad (15)$$

Que nos daría el tiempo de retraso entre el momento que se consigue los valores correspondientes a un rating superior y el momento en que el rating es asignado para todas las categorías y por otra parte el tiempo de retraso entre el momento que se tienen valores que corresponden a un rating inferior y se asigna el rating inferior para todas las categorías.

Aunque la función podría optimizarse, en una primera fase, se formula una propuesta sencilla para demostrar que el filtrado mejora la clasificación ordinal. A diferencia del trabajo de Broto et al (2016) en el que se estiman dos modelos distintos para uno para el caso de las subidas y otro para el caso de las bajadas en la propuesta presentada se transforma la serie clasificada en relación a las clasificaciones obtenidas en ejercicios anteriores. Esta transformación se puede formular de la siguiente manera:

$$UTR_{i \rightarrow j} = (1) \forall i < j; \quad (16)$$

$$DTR_{i \rightarrow j} = (0) \forall i > j \quad (17)$$

Esto implica que la categoría (rating) se degrada tan pronto como se conoce que la información que proporcionan las variables independientes pertenece a una calificación inferior a la estimada para el período anterior, pero se tarda un año en actualizar la categoría en el caso de un ascenso. Con esta transformación se evitan posibles rebotes de calificación en las mejoras, ya que una calificación no puede moverse hacia arriba un año y hacia abajo al siguiente por las

CAPÍTULO 2: Modelo de estimación de rating de deuda soberana

restricciones temporales impuestas. Esta suposición es fácilmente computable y se puede expresar de la siguiente manera:

$$y_{it}^* = \min (y'_{it}, y'_{it-1}) \forall t > 1 \quad (18)$$

Donde y_{it}^* representa la clasificación filtrada e y'_{it}, y'_{it-1} son las clasificaciones ordinales obtenidas a partir de las variables seleccionadas en el modelo propuesto de estimación de ratings para la deuda soberana.

CAPÍTULO 3

LA SELECCIÓN DE LA MUESTRA

1. INTRODUCCIÓN

Para analizar la idoneidad de los indicadores de solvencia que se han seleccionados se propone considerar una muestra que permita su validación.

Esta muestra debe reunir unas características adecuadas en la que se pretende evitar que factores adicionales a la solvencia inserten sesgos en la muestra que impidan una correcta interpretación de los resultados.

2. SELECCIÓN DE LA MUESTRA Y PERIODO TEMPORAL

Para la selección de la muestra y el periodo temporal se han planteado las siguientes restricciones:

- Los países deben formar parte de un contexto económico único con el fin de evitar que diferencias distintas de la solvencia, como podrían ser el marco jurídico, la fortaleza de la moneda, etc., puedan tener influencia sobre el rating.
- El periodo debe incluir distintas fases del ciclo económico para que los resultados sean validados en distintas fases del contexto económico global.

Con estas restricciones se han seleccionado los países de la Unión Europea. Estos países, a pesar de ser considerados todos como economías desarrolladas tienen economías muy diferentes tanto en tamaño como por su dinámica, pero con un marco normativo armonizado (Unión Europea, 2017).

CAPÍTULO 3: La selección de la muestra

Se ha analizado el periodo comprendido entre el año 2002 y el 2017 para incluir años previos a la crisis financiera de 2008 y los posteriores a la recuperación.

Para el periodo seleccionado solo se han tomado como referencia los países que han pertenecido a la Unión Europea durante todo el periodo seleccionado. Estos países son los siguientes:

- Alemania
- Bélgica
- Francia
- Italia
- Luxemburgo
- Países Bajos
- Dinamarca
- Irlanda
- Reino Unido
- Grecia
- España
- Portugal
- Austria
- Finlandia
- Suecia

La mayoría de estos países están integradas en la moneda única: el euro, que fue lanzado el 1 de enero de 1999 como moneda virtual y el 1 de enero de 2002 se introdujeron los billetes y monedas. Sólo Reino Unido Dinamarca y Suecia mantienen su moneda (Unión Europea, 2017).

CAPÍTULO 3: La selección de la muestra

La Unión Europea entre el año 2002 y 2017 es, por tanto, un conjunto de países con cierta armonización normativa, con monedas estables y en un periodo que incluye los años previos a la crisis, la crisis y la posterior recuperación por lo que con este conjunto de países se cumplen sobradamente las restricciones planteadas inicialmente.

3. FUENTES DE INFORMACIÓN E INDICADORES

Para analizar los ratings de los países seleccionados se ha seleccionado los ratings de la agencia Moody's por ser una de las líderes del mercado (HOUSE OF LORDS European Union Committee, 2011), y el gran peso que tiene la deuda soberana en esta agencia al clasificar en 2022 a 172 países (Moody's, 2022).

Esta agencia no tiene una escala definida para los emisores. Utiliza, por tanto, la misma escala que se usa para las emisiones de obligaciones. La siguiente clasificación es aplicable a los emisores de deuda a largo plazo:

- Aaa: Obligaciones de la más alta calidad y con el riesgo de crédito más bajo.
- Aa: Obligaciones de alta calidad y con riesgo de crédito muy bajo.
- A: Obligaciones de calidad media-alta y riesgo de crédito bajo
- Baa: Obligaciones de calidad media, con riesgo de crédito moderado y que pueden tener ciertas características especulativas.
- Ba: Obligaciones especulativas y con riesgo de crédito significativo.
- B: Obligaciones especulativas con riesgo de crédito alto.
- Caa: Obligaciones especulativas de mala calidad con riesgo de crédito muy alto.

CAPÍTULO 3: La selección de la muestra

- Ca: Obligaciones altamente especulativas próximas al impago, pero con alguna perspectiva de recuperar el principal y el interés.
- C: Obligaciones peor clasificadas que se espera que sean impagadas y con bajas perspectivas de recuperación del principal y el interés.

Adicionalmente para las categorías comprendidas entre los valores Aa hasta Caa se añaden unos sufijos que van de 1 a 3 indicando el número 1 que la clasificación está en la parte más alta de su categoría, el número 2 la parte media y el número 3 la parte baja. Además, periódicamente se publican perspectivas sobre evolución del riesgo que puede ser positiva, negativa, estable o en desarrollo. Los ratings también pueden estar en revisión o retirados para esta agencia (Moody's Investor Service, 2016a).

Esta escala de 9 categorías en las que 6 tienen tres modificadores distintos hace que el total de posibilidades ascienda a 21. Esto genera una gran cantidad de posibilidades de rating por lo que procede una simplificación que será abordada en el siguiente apartado.

Para los datos económicos correspondientes a los países seleccionados se ha tomado como fuente la web de Eurostat¹⁶. Eurostat es la oficina estadística de la Unión Europea y su función es proporcionar a la Unión Europea estadísticas a nivel europeo que permitan comparaciones entre países y regiones (European Commission, 2018).

Su página web proporciona estadísticas basadas en SEC 2010 de todos los países europeos. También proporcionan estadísticas de otros datos de interés como el desempleo o la deuda pública calculada según el protocolo del déficit

¹⁶ Obtenidos en noviembre de 2018

CAPÍTULO 3: La selección de la muestra

excesivo. Solo el excedente primario y el indicador de resiliencia económica se ha calculado a partir de datos de Eurostat sin tener el resto de indicadores ningún tipo de tratamiento.

Para las previsiones económicas se han utilizado las publicaciones de primavera del “Directorate-General for Economic and Financial Affairs” que es un departamento dependiente de la Comisión Europea responsable de las políticas de la Unión Europea que promueven el crecimiento económico y la recuperación, alto nivel de empleo, finanzas públicas estables y estabilidad financiera. Este departamento realiza previsiones económicas y estudios de mercado que sirven como base a la política económica en la Unión Europea (Economic Andl Financial Affairs, 2023). El “Directorate-General for Economic and Financial Affairs” publica anualmente tres previsiones económicas sobre los países europeos. En nuestro caso se han obtenido de los informes de primavera de los años 2003 a 2018 las previsiones para el año en curso y para el ejercicio siguiente del:

- Producto Interior Bruto
- Deflactor del Producto Nacional Bruto

Con estos datos relativos al ejercicio en curso y el ejercicio siguiente se ha calculado la previsión de crecimiento nominal acumulada utilizada como variable. (European Commission Directorate-General for Economic and Financial Affairs, 2003-2018).

4. PROCESAMIENTO DE LOS INDICADORES

Como se ha comentado la escala de ratings de Moody's incluye 9 clases de las que 6 tienen modificadores que las subdividen en 3. Para evitar un exceso de clases y permitir un tratamiento adecuado de la muestra no se han tenido en cuenta los modificadores.

Para dar consistencia temporal a los datos se parte del supuesto de que los primeros avances de las cifras macroeconómicas de un país para un año concreto no se conocen hasta pasados unos meses desde la finalización del ejercicio que se trate. Por tanto, hemos dado un plazo de 4 meses para que la obtención del rating relativo al ejercicio anterior. De esta forma la información del rating es consistente temporalmente con los datos macroeconómicos y las previsiones publicadas en primavera.

La siguiente tabla nos muestra el rating asignado por la agencia Moody's, sin tener en cuenta los modificadores, a cada uno de los países objeto de análisis durante el periodo seleccionado el 30 de abril del ejercicio que se indica.

Tabla 8: Ratings asignados por la agencia Moody's

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Belgium	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa
Denmark	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa
Germany	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa
Ireland	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aa	Baa	Ba	Ba	Baa	Baa	Baa	A	A
Greece	A	A	A	A	A	A	A	A	B	C	C	Caa	Caa	Caa	Caa	B
Spain	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aa	A	Baa	Baa	Baa	Baa	Baa	Baa
France	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aa	Aa	Aa	Aa	Aa	Aa	Aa
Italy	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	A	Baa	Baa	Baa	Baa	Baa	Baa
Luxembourg	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa
Netherlands	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa
Austria	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aa	Aa
Portugal	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Aa	Baa	Ba	Ba	Ba	Ba	Ba	Ba	Ba
Finland	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aa	Aa
Sweden	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa
United Kingdom	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aa	Aa	Aa	Aa	Aa	Aa

Fuente: Moody's

En el caso de Luxemburgo el primer rating de la agencia Moody's es de septiembre de 2010 en el que su deuda fue clasificada como Aaa, la máxima

CAPÍTULO 3: La selección de la muestra

categoría, por lo que para realizar nuestro análisis hemos asignado ese mismo rating al periodo anterior ya que no existe ninguna razón económica que justifique un rating inferior previo al tratarse de un país muy poco endeudado con una importante renta per cápita.

Del análisis de la anterior tabla se obtienen dos importantes conclusiones:

- En primer lugar, hay que destacar que predominan los ratings Aaa al ser la clase más frecuente.
- En segundo lugar, los ratings de la mayoría de los países han empeorado a lo largo del periodo seleccionado.

La siguiente tabla muestra la frecuencia de los distintos ratings en la muestra seleccionada

CAPÍTULO 3: La selección de la muestra

Tabla 9: Frecuencia de ratings en la muestra

MOODY'S	FRECUENCIA
Aaa	143
Aa	1
	2
	3
A	1
	2
	3
Baa	1
	2
	3
Ba	1
	2
	3
B	1
	2
	3
Caa	1
	2
	3
Ca	0
C	2

Fuente: elaboración propia

Como se puede observar predominan las clases que implican una mayor solvencia. Además, existen dos categorías con solo dos casos. En estas situaciones si se utilizan distintos grupos para el entrenamiento y la generalización estas clases son muy difíciles de predecir. Por ello se propone unificar las clases menos frecuentes en grupos que tengan un mínimo número de casos.

Para ello los ratings Ba y B se unificarán en una clase con 11 casos distintos y las ratings Caa y C se unificarán en una clase con 6 casos.

CAPÍTULO 3: La selección de la muestra

Tabla 10: Ratings y clases asignadas

MOODY'S		CLASE	FRECUENCIA
Aaa		C6=Aaa	143
Aa	1	C5=Aa	51
	2		
	3		
A	1	C4=A	12
	2		
	3		
Baa	1	C3=Baa	17
	2		
	3		
Ba	1	C2=Ba-B	11
	2		
	3		
B	1		
	2		
	3		
Caa	1	C1≤Caa	6
	2		
	3		
Ca			
C			

Fuente: elaboración propia

Quedarían por tanto 6 clases a clasificar en línea con otros trabajos como el de Sánchez Monedero et al. (2014) o Fernandez et al. (2013) que utilizan 5 clases para un grupo de países similar al estudiado o Bissoondoyal (Bissoondoyal-Bheenick, 2005; Bissoondoyal-Bheenick et al., 2006) que utiliza 9 clases al usar también solo las clases principales sin modificadores.

Los datos de Eurostat no tienen ningún tipo de tratamiento a excepción del excedente primario que se ha calculado sumando a la partida del déficit el servicio de la deuda y la resiliencia que, como se ha indicado anteriormente, es un dato calculado a partir del PIB publicado por Eurostat.

CAPÍTULO 3: La selección de la muestra

Para el cálculo de la previsión de crecimiento nominal acumulada correspondiente a los próximos dos años se han utilizado los datos de la previsión económica y la previsión del deflactor del PNB para los próximos años del “Directorate-General for Economic and Financial Affairs”

$$PCN_{1+2} = (1 + Prev_1) * (1 + PDPNB_1) * (1 + Prev_2) * (1 + PDPNB_2) - 1 \quad (19)$$

Siendo PCN_{1+2} la previsión de crecimiento nominal para los dos próximos años, $Prev_i$ la previsión de crecimiento nominal para el año i y $PDPNB_i$ la previsión del deflactor del PNB para el año i .

La siguiente tabla recoge un resumen de las variables utilizadas, sus fuentes y si tienen o no tienen tratamiento.

Tabla 11: Variables y su tratamiento

Variable	Definición	Unidad	Fuente	Tratamiento
PIB _t	Producto Interior Bruto en el periodo t	Euros constantes	Eurostat	No
PPC _t	Producto Interior Bruto per cápita en el periodo t	Euros constantes	Eurostat	No
b _t	Deuda pública en el periodo t	PIB	Eurostat	No
i	Tipo de interés máximo del periodo	%	Eurostat	No
x	Excedente primario del periodo t	PIB	Eurostat	Calculado a partir del Déficit y el Servicio de la Deuda
g ^N	Previsión del crecimiento nominal para los 2 próximos años	%	Directorate General	Calculado a partir de previsiones de crecimiento y de deflactor del PNB
CR _t	Resiliencia económica para el año t	índice	Eurostat	Calculado a partir del PIB

Fuente: elaboración propia

CAPÍTULO 3: La selección de la muestra

La siguiente fórmula especifica como se han utilizado los indicadores y que criterios temporales se han usado para las variables:

$$R_{it+0,33} = f(PIB_{it}, PPC_{it}, b_{it}, i_{\max(it-0,7:it+0,3)}, x_{it}, g_N(it:it+2), CR_{it}) \quad (20)$$

Donde:

- $R_{it+0,33}$ es el rating del país i el 30 de abril del periodo $t+1$.
- PIB_{it} es el PIB del país i en el periodo t .
- PPC_{it} es el PIB per cápita del país i en el periodo t .
- b_{it} es el nivel de deuda respecto al PIB del país i en el periodo t .
- $i_{\max(it-0,7:it+0,3)}$ es el máximo del tipo de interés de los bonos a largo plazo del país i entre el 16 de marzo del periodo t y el 15 de marzo del periodo $t+1$.
- x_{it} es el excedente primario del país i en el periodo t .
- $g_N(it:it+2)$ es la previsión de crecimiento nominal acumulado del PIB nominal del país i entre el periodo t y el periodo $t+2$.
- CR_{it} es el valor del indicador de resiliencia del país i en el periodo t .

5. ASPECTO FINAL DE LA MUESTRA

Como se ha comentado, la distribución de clases tiene un importante sesgo temporal. La siguiente tabla presenta la distribución de clases por ejercicio analizado.

CAPÍTULO 3: La selección de la muestra

Tabla 12: Distribución de clases por año.

CLASE	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
C6 Aaa	11	11	11	11	11	11	11	10	9	9	7	7	7	7	5	5
C5 Aa	3	3	3	3	3	3	3	4	3	1	3	3	3	3	5	5
C4 A	1	1	1	1	1	1	1	1		2					1	1
C3 Baa									2		2	3	3	3	2	2
C2 Ba-B									1	2	2	1	1	1	1	2
C1≤Caa										1	1	1	1	1	1	

Fuente: elaboración propia

Esta distribución, en la que en la segunda mitad del periodo aparecen clases nuevas hace muy difícil un aprendizaje de carácter temporal, en el que el aprendizaje de los años previos estime los siguientes, pues en los años que aparecen clases nuevas, si el conjunto de entrenamiento está formado por los años previos, el sistema no va a poder entrenar las clases nuevas.

Si analizamos la muestra de ratings también observamos que estos tienden a ser estables en el tiempo, de las 225 posibilidades de cambio de clase sólo se han producido cambios en 18 ocasiones. Esto hace que exista una fuerte relación entre el rating asignado y el rating del ejercicio anterior.

Para abordar una serie de este tipo existen dos estrategias posibles, una primera sería centrarnos en los cambios y estudiar cuando se producen los cambios para analizar si el modelo es capaz de estimar las 18 ocasiones en que se producen. Este planteamiento presenta el problema derivado del escaso número de cambios que hace que cualquier diferencia pueda plantear sesgos importantes y diferencias significativas.

La alternativa que se propone es analizar los casos en los que no se producen cambios. Para ello se eliminan en los conjuntos de entrenamiento todas las referencias al país y al periodo y en el conjunto de generalización se vuelven a

CAPÍTULO 3: La selección de la muestra

incluir estos datos para ver si las series que se estiman tienen el mismo carácter estable que las series originales.

Tampoco se van a reconocer efectos diferenciados no recogidos en nuestras variables para cada país ni a lo largo del periodo analizado. El conjunto de variables utilizadas debe de tener el mismo efecto en todos los países y en todos los periodos.

Por tanto, el aprendizaje no se hará sobre el panel de datos tratado como tal, sino que, para entrenar el modelo, hemos convertido cada vector de datos de país y periodo en un vector sin referencia alguna al país y periodo que corresponden.

$$R_{it} = R_n = (X_{1n}, X_{2n}, \dots, X_{7n}) \quad (21)$$

Que nos dicen que el rating de un país i en un periodo t se va a tratar como un rating indeterminado y donde n representa el valor de la fila de la tabla de datos que ocupa el rating del país y el año en cuestión y tienen un valor entre 1 y 240 (15 países estudiados durante 16 periodos).

Con posterioridad para estudiar el conjunto de generalización si se utilizará el país y periodo para poder analizar las diferencias en las clasificaciones por países.

6. ENTRENAMIENTO Y GENERALIZACIÓN

La gran capacidad de aprendizaje de las modernas técnicas de Inteligencia Artificial hace que algunos clasificadores sean capaces de predecir casi toda la muestra si se entrena completa. Esta aproximación limita la capacidad de

CAPÍTULO 3: La selección de la muestra

análisis sobre la capacidad de predicción de los indicadores seleccionados por lo que siguiendo el ejemplo de los últimos trabajos en los que se utilizan estas técnicas (Bennell et al., 2006; Fernández-Navarro et al., 2013; Ozturk et al., 2016; Sánchez-Monedero et al., 2014; Van Gestel et al., 2006) se van a diferenciar los conjuntos de entrenamiento, que es conjunto de patrones que se utilizan para que el modelo se ajuste y defina, y el conjunto de generalización, que es el conjunto de patrones que se utilizan para evaluar la capacidad de predicción del modelo. Al utilizarse conjunto de patrones distintos para entrenar que para generalizar se consigue que los patrones sobre los que se hace un análisis de predicción no se hayan utilizado antes para entrenar.

Para ello se han utilizado dos técnicas: el hold-out (De Moor et al., 2018; Fernández-Navarro et al., 2013; Sánchez-Monedero et al., 2014) y el ten-fold (Ozturk et al., 2016).

El hold-out consiste en definir un conjunto aleatorio de patrones para entrenar y usar el resto de los patrones para generalizar. En nuestro caso se ha entrenado con el 70% y generalizado con el 30% seleccionados de forma aleatoria. Para esta selección aleatoria se suele incluir en la selección un estratificado por clases de forma que la representación de clases, tanto en los conjuntos de entrenamiento como en el de generalización sea proporcional al conjunto de clases de la muestra total.

Este método tiene una principal limitación y es la enorme dependencia de los resultados de los conjuntos seleccionados en los casos en que la dificultad de clasificación de los patrones no sea homogénea. En nuestro caso nos encontramos con dos tipos de patrones:

CAPÍTULO 3: La selección de la muestra

- Casos con alta probabilidad de estimación. Son casos muy similares a otros de la serie y, por tanto, generan un buen aprendizaje en el conjunto de entrenamiento y que hace que los algoritmos tengan una buena capacidad de predicción y, por tanto, desempeñan bien en la generalización.
- Casos con baja probabilidad de estimación. Son casos en los que, en la mayoría de los casos, los algoritmos no son capaces de predecir y desempeñan mal en la generalización. Esto puede deberse a dos factores principales
 - o Singularidad, son casos muy particulares que podrían entrenarse. Al no coincidir nunca en entrenamiento y generalización si entrenan no generalizan y si generalizan no han podido entrenar. Por tanto, es muy improbable que sean estimados correctamente en la generalización.
 - o Diferencia, son casos, poco frecuente, similares a otros de la serie, pero con una clase distinta. Estos generan errores de clasificación en la generalización al asignar el algoritmo siempre la clase de los que tienen características similares.

Los casos singulares, en el caso de incluirse en el entrenamiento no generan problemas pues simplemente el modelo aprende un caso que no va a usar en generalización; sin embargo, si van a dar un error en generalización porque el modelo no va a tener ninguna indicación de cómo tratar ese patrón concreto y lo va a asimilar al que encuentre más parecido de los que ha entrenado que debe tener una clase diferente.

CAPÍTULO 3: La selección de la muestra

En los casos diferentes, al ser minoritarios, no deberían alterar mucho el algoritmo de clasificación, pero en el caso de generalizar nos van a dar un error al asignar el algoritmo la clase de los casos similares.

En ambos casos el que determinados patrones entren en un conjunto o en otro afectan a los resultados que proporcione el algoritmo por lo que los resultados van a depender de los patrones que entren en el conjunto de entrenamiento y el de generalización. Para evitar este sesgo se han repetido las clasificaciones 30 veces con distintos conjuntos de entrenamiento y generalización.

Otro problema que plantea el hold-out es que al entrenar un subconjunto de patrones y generalizar otro, sólo se puede hacer análisis sobre el subconjunto de generalización no siendo viable el análisis de todo el conjunto de patrones, ya que, aunque se generalicen todos los patrones en las 30 repeticiones estos se harían con distinta frecuencia lo que exigiría un tratamiento posterior de las 30 generalizaciones para determinar el conjunto completo de generalización.

La otra técnica utilizada es el k-fold. Para ello se divide la muestra en un número k de subconjuntos (en nuestro caso 10). Esta división se suele hacer de forma aleatoria y estratificada respecto a la variable dependiente para que todos los subconjuntos entrenen con todas las clases y en proporciones similares a la muestra. Una vez segregada la muestra el experimento se repite k veces utilizando uno de los k subconjunto generados para hacer la generalización y el resto, 9 en nuestro caso, para entrenar el algoritmo, cada experimento clasifica uno de los 10 subconjuntos generados y tras repetir el experimento 10 veces se obtiene una clasificación en la que todos los subconjuntos han sido clasificados sin que haya ninguna coincidencia entre entrenamiento y generalización.

CAPÍTULO 3: La selección de la muestra

Con esta técnica se consigue unos porcentajes de datos correctamente clasificados similares a la media de hacer las 30 repeticiones de hold-out. Esto permite, por una parte, proporcionar datos realistas sobre la capacidad de predicción de nuestro modelo consistentes con su capacidad real de predicción y, por otra, tener toda la muestra clasificada para poder realizar análisis pormenorizado de las clasificaciones teniendo toda la muestra clasificada.

CAPÍTULO 4

LOS CLASIFICADORES

1 INTRODUCCIÓN

En el presente capítulo se van a analizar varios clasificadores que se han utilizado para verificar la capacidad de predicción de las variables seleccionadas.

Nos encontramos ante un problema de clasificación multiclase ordinal. Es multiclase porque hay seis clases que tienen una denominación alfabética, Aaa, Aa, A, Baa, Ba-B y \leq Caa. Es ordinal porque las clases tienen un orden preestablecido: pertenecer a una clase supone tener mejor rating que las que le siguen y peor a las que le anteceden, pero no se tiene más información que el orden sin que se sepa si existe alguna relación o proporcionalidad entre las distintas clases.

Gutiérrez et al. (2016) justifican en su trabajo sobre problemas de clasificación ordinal indicando que el problema de predicción del rating puede ser considerado como un claro ejemplo de clasificación ordinal.

Se considera que un problema de clasificación es desbalanceado cuando predomina una clase sobre el resto. En este caso un clasificador que clasificara toda la muestra en la clase predominante proporcionaría algunas métricas de clasificación aceptables. En este caso es importante valorar cómo funciona el clasificador en las clases menos frecuentes para verificar que se cumplen los objetivos planteados. El caso que se estudia es desbalanceado porque predomina una clase "Aaa" sobre el resto (143 casos de 240) y predominan las clases que suponen una mejor clasificación que las que suponen una peor clasificación (las tres clases que presentan una mejor clasificación suponen 206

CAPÍTULO 4: Los clasificadores

Los resultados de la clasificación con los distintos clasificadores proporcionaran información sobre la robustez del modelo y su capacidad de predicción conforme al objetivo específico 1-2.

Para ello vamos a abordar el problema desde dos perspectivas, una estrictamente nominal donde los clasificadores trabajan asignando la clase sin tener en cuenta su posición en la escala de ratings y otra ordinal en el que los clasificadores sí tienen en cuenta el orden de las clases.

Se van a utilizar clasificadores que se corresponden con la estadística clásica y también técnicas más modernas que se corresponden con el área de conocimiento del aprendizaje automático.

Para que los resultados obtenidos con los distintos clasificadores sean comparables se han realizado dos particiones aleatorias estratificadas por clase: En primer lugar, se ha planteado un hold-out por ser el método que predomina en la literatura y que permite una mejor comparación de resultados con los trabajos de otros autores. En este caso se entrena con el 70% de la muestra y se generaliza con el 30% restante. Este método tiene el inconveniente de que en casos como el que se analiza donde existen casos más fáciles de predecir que otros es muy sensible a los casos concretos que pertenezcan al conjunto de generalización. Para evitar este inconveniente se han realizado 30 particiones distintas que se han utilizado en todos los casos. Una técnica para evitar este problema es el 10-fold en el que se realizan 10 particiones y repetidamente se entrena con 9 de ellas y se generaliza con la restante. Este proceso se repite 10 veces hasta completar la muestra con lo que todos los elementos de la muestra con independencias de que sea más fácil o difícil predecirlos son analizados.

CAPÍTULO 4: Los clasificadores

Estas dos particiones se han utilizados con todos los clasificadores utilizados en cada caso.

Para el análisis se ha recurrido a dos herramientas informáticas el software KNIME (Berthold et al., 2009) y el desarrollo sobre MatLab ORCA (Gutiérrez et al., 2016; Sánchez-Monedero et al., 2019).

2. CLASIFICADORES NOMINALES

2.1. Regresión Logística

La regresión logística es un método estadístico utilizado para adaptar una regresión a resultados nominales (Hosmer & Lemeshow, 2000).

Este método ha sido utilizado por diversos autores (Afonso, 2003; Broto & Molina, 2016; De Moor et al., 2018; E.-C. Miricescu et al., 2016).

Para ello se basa en determinar la probabilidad de un caso frente a otros con la siguiente función probabilística:

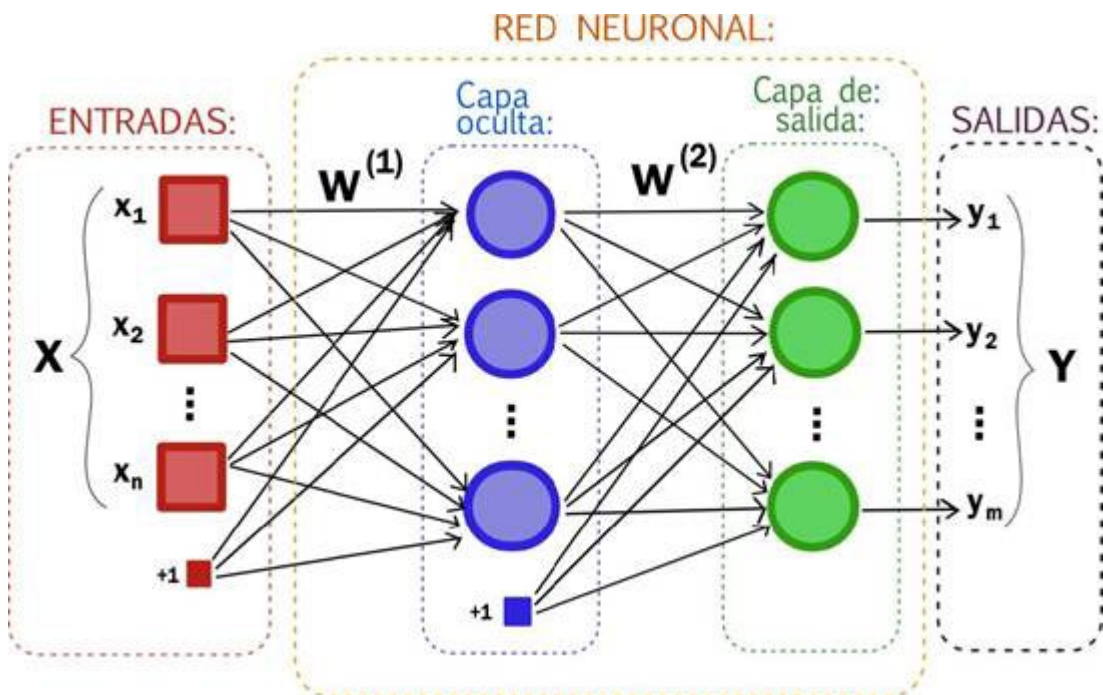
$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + i)}} \quad (22)$$

Donde p_i representa la probabilidad del caso i y las β son los parámetros de la regresión a estimar. Para la estimación se utiliza la implementación de KNIME (Berthold et al., 2009) con la opción de Stochastic Average Gradient. La clase se asigna a la que proporciona una mayor probabilidad.

2.2. Redes Neuronales

Una red neuronal es una relación funcional entre unas variables de entrada, en nuestro caso las distintas variables de entrada asociadas a cada país y periodo, y unas variables de salida, en nuestro caso el rating. Para ello se intenta emular el sistema nervioso biológico con una gran cantidad simple de elementos de procesamiento de información que se asimilan a las neuronas humanas.

Gráfico 6: Red neuronal:



Fuente: ArtFromCode

Para ajustar los valores de los pesos asignados a cada unión w_i se utiliza una función de error y un mecanismo de retropropagación que realiza pequeños cambios en los pesos de las conexiones con el fin de minimizar la función de error. Los pesos iniciales de las conexiones se establecen de forma aleatoria. Por tanto, los resultados pueden variar en distintas repeticiones según el acierto que tenga el método al asignar inicialmente los pesos de las conexiones.

CAPÍTULO 4: Los clasificadores

Podemos concluir que este método no es determinista pues distintas aplicaciones sobre el mismo conjunto pueden proporcionar distintos resultados.

Se ha utilizado la función RProp MLP Learner de KNIME (Berthold et al., 2009) basado en el trabajo de Martin Riedmiller y Heinrich Braun (1993). Para cada nodo de la capa oculta se utiliza una función que se conoce como función de activación. Pueden usarse distintas funciones de activación, aunque la más frecuente y la que se ha utilizado es la función sigmoide:

$$y = \frac{1}{1+e^{-x}} \quad (23)$$

En el caso de clasificación multiclase, como el que estamos utilizando, y con la asignación de pesos w_j que hace el algoritmo la red neuronal se describe con la siguiente función:

$$H_j(x_1, x_2 \dots x_k) = \frac{1}{1+e^{-(\sum_1^k w_{ji}x_i+w_{j0})}} \quad (24)$$

Este clasificador ha sido usado por diversos autores con muy buenos resultados (Bennell et al., 2006; Ozturk et al., 2016).

2.3. Máquinas de Vectores de Soporte

Los algoritmos de máquinas de vectores de soporte (SVM) se basan en encontrar los hiperplanos que maximicen la separación entre las diferentes clases según lo descrito por Chih-Wei Hsu et al. (2008). Para lograr este objetivo, se emplean criterios de optimización matemática que definen la separación máxima entre las clases.

CAPÍTULO 4: Los clasificadores

Se ha utilizado la función de base radial. Es importante destacar que las máquinas de vector soporte son clasificadores basados en problemas de optimización matemática y por tanto de carácter determinístico.

En este sentido, se ha utilizado la implementación SVM de KNIME (Berthold et al., 2009), utilizando la opción de funciones de base radial. Cabe destacar que las máquinas de vector soporte son algoritmos binarios que tienen la capacidad de separar dos clases. Sin embargo, para resolver problemas multiclase, se debe repetir el algoritmo comparando cada clase con el resto de las clases.

Se ha utilizado la función de base radial y se ha optimizado con particiones distintas de las usadas en el experimento final. Es importante destacar que las máquinas de vector soporte son clasificadores de carácter determinístico, es decir, aplicados a un mismo conjunto de datos siempre proporcionan el mismo resultado. Por lo tanto, no se ha repetido el experimento cuando no ha sido necesario.

Un ejemplo de utilización de máquinas de vector soporte es el estudio de Sánchez Monedero et al. (2014), que se basa en un modelo que emplea máquinas de vector soporte. Asimismo, Ozturk et al. (2016) y Gestel et al. (2006) también han utilizado este algoritmo en sus investigaciones.

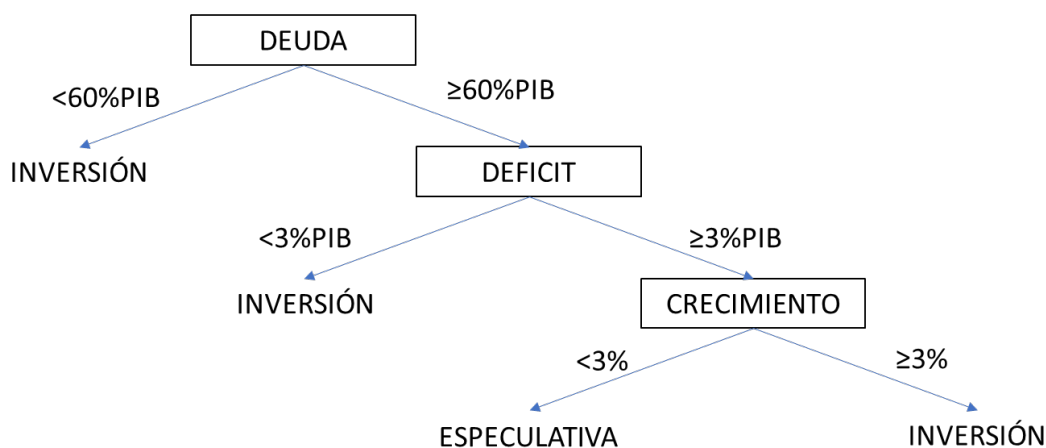
2.4. Random Forest

Para entender el método “Random Forest” (Breiman, 2001) hay en primer lugar que entender que es un árbol de decisión. Un árbol de decisión es “un conjunto de condiciones organizada en una estructura jerárquica, de manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se

CAPÍTULO 4: Los clasificadores

cumplan desde la raíz del árbol a cada una de sus hojas” (Hernández Orallo et al., 2005). A continuación, se representa gráficamente un ejemplo de árbol de decisión¹⁷:

Gráfico 7: Ejemplo de árbol de decisión



Hay que destacar que todos los nodos tienen una decisión alternativa y que las clases finales son excluyentes.

Para crear un árbol de decisión a partir de un conjunto de datos, se sigue un enfoque jerárquico que implica dividir los datos en particiones en cada nodo, con el objetivo de maximizar la separación entre las clases. Es crucial generar estas particiones de manera óptima, ya que una vez creadas, no se pueden modificar. Para lograrlo, es necesario determinar qué particiones deben considerarse y qué criterios se utilizan para seleccionarlas de manera adecuada. Existen diferentes

¹⁷ Árbol de decisión con 3 variables: deuda, déficit y crecimiento y 2 clases: inversión y especulativa

CAPÍTULO 4: Los clasificadores

algoritmos para aprender mediante árboles de decisión, uno de los más utilizados es el C4.5 desarrollado por Ross Quinlan en la década de los 80 y principio de los 90 (J. R. Quinlan, 1993). Este algoritmo genera todas las particiones posibles para cada uno de los atributos utilizando los valores nominales, en el caso de que el atributo sea nominal, o los valores intermedios de los intervalos en el caso de que el atributo sea numérico (si existen demasiados ejemplos se selecciona un subconjunto mediante técnicas de densidad u otras). Las particiones en el algoritmo C4.5 se van seleccionando en función de la capacidad de separación que tiene cada una de las particiones; para ello se mide el número de elementos ajenos que existe en cada una de las particiones resultantes utilizando la siguiente fórmula:

$$I(s) = \sum_{j=1..n} p_j \cdot f(p_j^1, p_j^2, \dots, p_j^c) \quad (25)$$

Donde n es el número de nodos de cada partición, p_j es la probabilidad de caer en el nodo j y p_j^1 es la proporción de elementos de la clase 1 en el nodo j , y así sucesivamente en las c clases.

La función que se suele usar en el algoritmo C4.5 es la ganancia de entropía (gain) que es la siguiente:

$$\sum p_j \cdot \log(p_j) \quad (26)$$

Este procedimiento puede dar lugar a árboles que lleguen a identificar individuo a individuo lo que luego no resultaría efectivo para clasificar a nuevos elementos por lo que el algoritmo C4.5 utiliza criterios de poda, unos previos, no abre nuevos nodos si la clasificación no aporta una determinada disminución de la

CAPÍTULO 4: Los clasificadores

impureza, otros a posteriori, analizando la aportación de impurezas del conjunto de nodos.

Una de las limitaciones inherentes a este algoritmo es su incapacidad para modificar la jerarquía de los nodos una vez que ha sido establecida. Esta jerarquía está fuertemente condicionada por el conjunto de entrenamiento utilizado. El modelo no explora la posibilidad de que un árbol que aporte una menor entropía en la primera división pueda tener un mejor resultado final en su conjunto. Esto hace que este algoritmo sea muy sensible al conjunto de entrenamiento y se le considere un aprendiz débil (*weak learner*) (Hernández Orallo et al., 2005).

Para solucionar este problema existen técnicas que intentan disminuir la sensibilidad del modelo al conjunto de entrenamiento. Una de ellas es el Bagging (Breiman, 1996; J. Quinlan, 1996) que deriva del mecanismo denominado “bootstrap aggregation”, que lo que hace es generar aleatoriamente y con reemplazamiento subconjuntos de entrenamiento del mismo tamaño que el original. Así con cada subconjunto de datos se puede generar un árbol de decisión distinto.

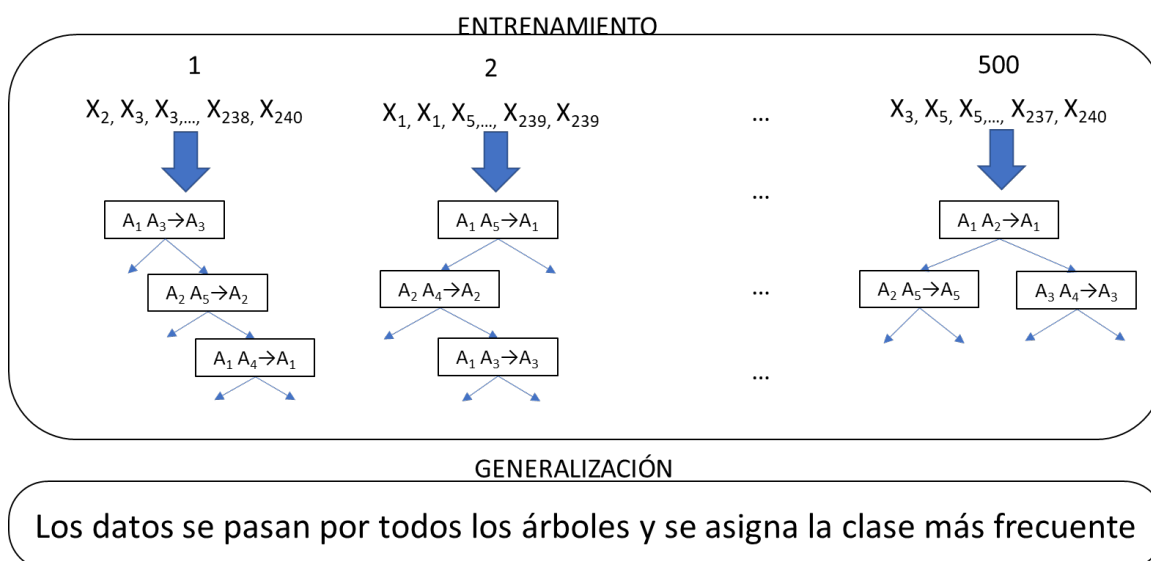
Para introducir una mayor aleatoriedad a la decisión se practica la selección de atributos (Ho, 1998). Para cada nodo sólo se utiliza un subconjunto de atributos.

Aplicando estas técnicas el algoritmo genera un conjunto de árboles que, debido a que utiliza distintos subconjuntos de variables y datos para su construcción, serán distintos entre sí. El conjunto de árboles que se ha obtenido en el proceso de aprendizaje se aplica a los patrones y la decisión que se obtiene es la mayoritaria entre todos los árboles.

CAPÍTULO 4: Los clasificadores

Gráficamente nuestro modelo quedaría representado de la siguiente forma:

Gráfico 8: Random Forest



Fuente: elaboración propia

Como puede observarse los conjuntos de cada árbol son distintos, no están todos los elementos y existen elementos repetidos y en cada nodo sólo se opta entre un conjunto limitado de variables. En la implementación en KNIME el número de variables en cada nodo es \sqrt{n} siendo n el número de atributos.

Random Forest es un método no determinista, cada vez que se ejecuta puede generar resultados distintos.

Este algoritmo se caracteriza por un excelente desempeño; en el trabajo realizado por Manuel Fernández Delgado, Eva Cernada, Senén Barro y Amorim Dinani (2014) donde se compararon 179 métodos de clasificación utilizando 121 bases de datos de distinta naturaleza los algoritmos del tipo "Random Forest" han obtenido los mejores resultados. Este algoritmo ha sido utilizado en los trabajos más recientes por Huseyin Ozturk, Ersin Namlı, y Halil Ibrahim Erdal (2016) y Lieven De Moor, Prabesh Luitel, Piet Sercu y Rosanne Vanpée (2018).

3. CLASIFICADORES ORDINALES

La clasificación ordinal es un tipo de análisis estadístico que se utiliza para ordenar y clasificar variables categóricas que guardan un orden entre sí. A diferencia de la clasificación nominal, donde las categorías son únicamente etiquetas que no tienen ningún orden lógico, en la clasificación ordinal, las categorías tienen un orden predefinido. De tal forma que se establece un orden entre las clases y el error se considera mayor conforme mayor sea la distancia entre la clase marcada en la predicción y la clase (Gutiérrez et al., 2016).

El único trabajo previo que se ha encontrado de estimación de ratings de deuda soberana basados en una aproximación ordinal es el de Fernández Navarro et al. (2013). Un análisis de su funcionamiento fue presentado en el XXXIII Congreso Internacional de Economía Aplicada Asepelt por Galnares Jiménez-Placer Carlos, Martínez Estudillo Alfonso Carlos y Carbonero Ruz Mariano (2019) con la comunicación: “Estimación de ratings de deuda soberana en la UE15 de 2002 a 2017 mediante clasificadores ordinales”.

En su trabajo presentan tres tipos diferenciados de clasificadores ordinales en función de su metodología para afrontar la ordinalidad de las clases:

- Naïve: El problema de clasificación ordinal es simplificado a clasificadores usuales de aprendizaje automático como clasificadores nominales, sensibles a los costes o regresiones. Se incluyen en esta categoría porque suponen alguna limitación en su carácter ordinal.
- Descomposición Ordinal Binaria: que supone descomponer las clases en grupos binarios siguiendo un esquema ordinal, de forma que el grupo binario esté siempre compuesto por clases adyacentes. Utilizando un

CAPÍTULO 4: Los clasificadores

clasificador con salida probabilística, y se determina por diferencia de probabilidades la probabilidad de cada clase. Esta clasificación se basa en el trabajo de Eibe Frank y Mark Hall (2001).

- Modelos de límites: en los que se predice una variable numérica latente y se definen los límites de los valores de esa variable para cada una de las clases.

Siguiendo esa taxonomía se han usado los siguientes clasificadores, que a continuación son descritos brevemente:

3.1 Métodos considerados NAÏVE:

3.1.1 Support Vector Classifier with OneVsOne SVC1V1

(Hsu & Lin, 2002): Máquina de vector soporte en la que se comparan todas las posibles parejas de clases (one vs. one). Se considera Naïve porque realmente es un clasificador nominal en el que no se tiene en cuenta el orden.

3.1.2 Support Vector Classifier with OneVsAll SVC1VA

(Hsu & Lin, 2002): Máquina de vector soporte en la que se compara cada clase con el resto de las clases. Igualmente se considera naïve porque es un clasificador nominal en el que no se tiene en cuenta el orden. Este clasificador es muy similar a la implementación de SVM en KNIME.

3.1.3 Support Vector Machines for regression SVR

(Smola & Schölkopf, 2004): Vectores soportes de regresión, con objetivos normalizados. Este clasificador se considera naïve porque asume que la distancia entre las clases es la misma.

CAPÍTULO 4: Los clasificadores

3.1.4 Cost-Sensitive Support Vector Classifier **CSSVC**

(Hsu & Lin, 2002): Máquina de Vector Soporte donde se compara cada clase con el resto y se incluyen costes absolutos como diferentes pesos de la clase negativa para cada una de las descomposiciones. Este método también se considera naïve por asumir que la distancia entre las clases es la misma.

3.2. Métodos de descomposición binaria:

3.2.1 Neural Network with Ordered Partition **NNOP**

(Cheng et al., 2008). Red Neuronal con descomposición binaria ordinal y una capa oculta. Este modelo incluye una metodología para ignorar predicciones inconsistentes.

3.2.2 Extreme Learning Machine with Ordered Partition **ELMOP**

(Deng et al., 2010). Red Neuronal con una capa oculta y pesos aleatorios en la entrada de la capa oculta. Agiliza sustancialmente el cómputo frente a una red neuronal clásica al no computar los pesos de la capa de entrada.

3.2.3 Hierarchical Partial Order Label Decomposition **HPOLD**

(Sánchez-Monedero et al., 2018). Clasificación ordinal parcial con un modo jerárquico. El método binario es SVC1V1 y el ordinal Support Vector for Ordinal Regression with Implicit constraints SVORIM (Chu & Keerthi, 2007), que se expone más adelante como modelo de límites.

3.3 Modelos de límites

3.3.1 Kernel Discriminant Learner for Ordinal Regression **KDLOR**

(Sun et al., 2010). Aunque el análisis discriminante no se considera una técnica de clasificación se puede usar en el análisis ordinal porque a las proyecciones obtenidas se les pueden aplicar límites para discriminar las diferentes clases. En este caso el análisis discriminante se ha reformulado imponiendo una restricción ordinal en las clases proyectadas.

3.3.2 Ordinal Projection Based Ensemble **OPBE**

(Pérez-Ortíria et al., 2014). Basado en una descomposición en tres clases siguiendo la estructura ordinal utiliza como base SVORIM como clasificador base, pero podría implementarse con cualquier clasificador de ORCA.

3.3.3 Reduction applied to Support Vector Machines **REDSVM**

(Lin & Li, 2012), con una estructura aumentada de clasificadores binarios que lo hace equivalente a modelos de límites.

3.3.4 Boosting ensemble for Ordinal Regression **ORBOOST**

(Lin & Li, 2006). Ensemble de modelos de límites con funciones sigmoideas normalizadas como clasificadores de base.

3.3.5 Neural Network based on Proportional Odd Model **NNPOM**

(Mathieson, 2001; McCullagh, 1980). Implementa un modelo de red neuronal ordinal para regresión ordinal; el modelo tiene una capa oculta y una salida con

CAPÍTULO 4: Los clasificadores

una sola neurona y un conjunto de límites igual al número de clases menos una. El método Proportional Odd Model POM estándar (McCullagh, 1980) se aplica a esta neurona para obtener salidas probabilísticas.

3.3.6 Support Vector for Ordinal Regression with Explicit constraints **SVOREX**

(Chu & Keerthi, 2007). Computa hiperplanos paralelos discriminantes par los datos y un conjunto de límites imponiendo restricciones explícitas en el problema de optimización utilizando solo las clases adyacentes en la determinación de los límites.

3.3.7 Support Vector for Ordinal Regression with Implicit constraints **SVORIM**

(Chu & Keerthi, 2007). Computa hiperplanos paralelos discriminantes par los datos y un conjunto de límites imponiendo restricciones implícitas en el problema de optimización. En este caso todas las clases son consideradas al calcular los errores de cada hiperplano.

3.3.8 Linear version of the SVORIM method **SVORIMLIN**

(Chu & Keerthi, 2007). Implantación de un SVORIM considerando un cómputo lineal en vez de gaussiano.

4 MÉTRICAS DE ERROR

Para evaluar los resultados de las clasificaciones obtenidas y la consecución de los objetivos plantados en la presente tesis se han utilizado distintas métricas.

4.1. Métrica nominal

En estas métricas solo se tiene en cuenta si el patrón está correctamente clasificado o no, con independencia de que, en caso de error, el patrón se clasifique más cerca o más lejos de la clase correcta.

4.1.1 Datos correctamente clasificados **CCR**

Se define como el porcentaje de patrones correctamente clasificados, esto es, en los que el error de clasificación es igual a 0 (Correct Classified Results **CCR**). Matemáticamente se define como la suma del número de casos correctamente clasificados, el error absoluto es 0, entre el total de patrones clasificados.

$$e_{i,t} = |R_{i,t}^* - R_{i,t}| \quad (27)$$

$$CCR = \frac{1}{N} \sum_{n=1}^N \llbracket e_n = 0 \rrbracket \quad (28)$$

Siendo $e_{i,t}$ el error absoluto del país i en el periodo t y $R_{i,t}^*$ y $R_{i,t}$ el rating estimado y el rating real, N el número de patrones y $\llbracket \cdot \rrbracket$ un test booleano que da el valor 1 si la condición se cumple.

Esta métrica no tiene carácter ordinal porque los errores se valoran igual con independencia de la distancia entre el patrón correcto y el patrón estimado.

4.2 Métricas ordinales

Estas métricas si tienen en cuenta el carácter ordinal de la serie tratada y sus resultados dependen del tamaño de los errores cometidos en la clasificación

CAPÍTULO 4: Los clasificadores

4.2.1 Datos aproximadamente clasificados **ACR**

Se define como el porcentaje en que los datos en los que el error de clasificación es menor o igual a una clase, esto es, los ratings estimados son correctos o han sido clasificados en una clase contigua a la correcta.

$$e_{i,t} = |R_{i,t}^* - R_{i,t}| \quad (29)$$

$$ACR = \frac{1}{N} \sum_{n=1}^N \mathbb{I}[e_n \leq 1] \quad (30)$$

Las definiciones son las mismas que para el cálculo de datos correctamente clasificados con la diferencia de que el test booleano en este caso se cumple también cuando el error es igual a 1. El carácter ordinal lo proporciona el valorar de forma distinta los errores que son superiores a una clase de estos.

4.2.2 Error absoluto medio **MAE**

Se define como la desviación media medida, en valor absoluto, del error de clasificación, esto es, la diferencia entre la clase predicha y la clase real.

$$MAE = \frac{1}{N} \sum_{n=1}^N e(x_n) \quad (31)$$

Siendo N es el número de ocurrencias en datos de generalización (Baccianella et al., 2009). Su valor teórico está comprendido entre 0 y N-1 al existir N clases.

4.2.3. Error absoluto medio por clases **AMAE**

Definido como la media del error absoluto medio por clases. En este caso, se calcula en primer lugar el error absoluto por cada clase y luego se promedia. De esta forma se da una mayor valoración que en el caso del MAE a las clases

CAPÍTULO 4: Los clasificadores

menos frecuentes pues los errores de cada clase tienen el mismo peso con independencia de que sean más o menos frecuentes.

$$AMAE = \frac{1}{J} \sum_{j=1}^J MAE_j = \frac{1}{J} \sum_{j=1}^J \frac{1}{n_j} \sum_{n=1}^{n_j} e(x_n) \quad (32)$$

Siendo n_j el número de ocurrencia en cada una de las j clases (Baccianella et al., 2009).

5. SOFTWARE UTILIZADO

Para la clasificación nominal se ha utilizado la aplicación KNIME (Berthold et al., 2009).

- La regresión logística se ha utilizado en su configuración por defecto y una normalización gaussiana sobre el conjunto de entrenamiento. Esta normalización se ha aplicado al conjunto generalización.
- La red neuronal se ha ajustado en el número de neuronas (15) mediante la optimización que incluye el propio software. Para evitar el sobreentrenamiento se han utilizado para la optimización particiones distintas a las utilizadas en la experimentación. Los datos han sido normalizados utilizando una normalización de mínimos y máximos (0,1) sobre el conjunto de entrenamiento y dicha normalización se ha aplicado al conjunto de generalización.
- La máquina vector soporte se ha usado con la función de base radial y optimizándose el parámetro de costes (10^6) con particiones distintas a las utilizadas en la experimentación. Igualmente se ha utilizado una normalización de mínimos y máximos sobre el conjunto de entrenamiento.

CAPÍTULO 4: Los clasificadores

- Random Forest se ha utilizado en sus parámetros por defecto exceptuando el número de árboles que, tras diversas pruebas realizadas, se ha determinado en 500 al comprobarse que cifras superiores no suponían ninguna mejora.

Para los problemas de clasificación ordinal se ha utilizado la implementación sobre Matlab ORCA (Sánchez-Monedero et al., 2019). En este caso todos los algoritmos se han utilizado en su configuración por defecto con las rutinas de optimización definidas en la propia aplicación.

CAPITULO 5

EXPERIMENTACIÓN

1 INTRODUCCIÓN

Para la validación de las hipótesis planteadas se proponen diferentes tipos de experimentación. En un primer lugar se analizan los resultados de dividir la muestra en 30 particiones en las que el 70% se dedica al entrenamiento del modelo y el 30% a la generalización. Sobre estos resultados se hace un análisis de sensibilidad comprobando como varían los resultados en función de que se quiten o añadan nuevas variables. También se hace una comparación de los resultados obtenidos con los que se obtienen en el caso de usar las variables de otros autores con la misma muestra.

Seguidamente se procede a una validación cruzada de 10 subconjunto (10-fold) con los mismos clasificadores.

Con los mismos subconjuntos se han utilizado clasificadores ordinales con el software ORCA (Sánchez-Monedero et al., 2019).

A continuación, se ha realizado un análisis de como varían los resultados en el caso de filtrar los resultados obtenidos para contemplar las asimetrías en los movimientos de rating.

Con todos los resultados obtenidos se efectúa una comparativa entre todos los clasificadores para detectar cuales son los casos más fáciles y difíciles de clasificar.

También se realiza un ensamble de clasificadores en el que se adjudica como valor la media entre todas las clases asignadas. Por último, con los resultados obtenidos con Random Forest y el ensamble de clasificadores y tras el filtrado aplicado se realiza un análisis detallado de los errores que se comenten en cada una de estas clasificaciones en cada país.

2. HOLD-OUT

El Hold-Out se ha realizado mediante las opciones de partición de KNIME utilizando la opción de conjuntos estratificados por clases.

Para evitar el sesgo que pudiera tener un determinado conjunto de generalización se han realizado 30 particiones distintas. Estas 30 particiones se han utilizado en todos los casos los mismos 30 conjuntos de entrenamiento y 30 de generalización para que los resultados sean lo más comparables posible.

En estos casos sólo se han utilizado métricas nominales. No se usan métricas ordinales porque los conjuntos de generalización son variados y las métricas no serían comparables con otros trabajos por estar los patrones repetidos en la generalización con distinta frecuencia.

Se ha comprobado que todos los patrones han entrado en alguno de los conjuntos de generalización por lo que el conjunto completo de patrones ha sido estimado en los 30 experimentos. La siguiente tabla representa el número de veces que cada patrón se ha incluido en el conjunto de generalización, que varía entre 3 repeticiones en los casos de Austria 2002, Francia 2017 y Alemania 2014 y 17 repeticiones en los casos de Grecia 2017 y Suecia 2015.

CAPÍTULO 5: Experimentación

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Austria	3	15	11	11	6	12	10	9	8	6	10	8	14	13	8	7
Belgica	7	8	8	7	10	9	6	13	12	7	8	5	12	5	10	4
Dinamarca	8	4	6	9	8	10	8	10	11	10	5	7	4	12	6	12
Finlandia	10	6	5	11	7	9	10	8	10	4	9	11	11	8	9	9
Francia	9	8	9	11	7	8	11	9	8	12	10	8	12	9	11	3
Alemania	14	7	13	9	7	4	8	8	10	10	7	11	3	8	7	11
Grecia	9	7	8	12	15	9	9	11	7	9	13	9	11	5	13	17
Irlanda	7	9	9	11	11	7	10	5	11	6	8	7	7	6	10	10
Italia	7	11	8	9	11	7	9	13	12	11	12	7	14	6	9	8
Luxemburgo	9	8	9	8	8	14	16	7	6	7	5	5	8	11	10	10
Holanda	10	7	11	9	10	11	7	11	9	9	7	9	7	9	14	6
Portugal	8	11	6	10	8	9	12	10	8	8	5	11	8	8	4	8
España	10	12	10	7	11	5	8	6	9	9	6	11	12	8	7	11
Suecia	6	4	10	9	7	9	11	15	13	10	7	9	8	17	7	8
Reino Unido	9	11	9	14	10	11	13	11	9	9	8	10	11	9	9	11

Fuente: elaboración propia

Para el ajuste de los clasificadores se han dejado las opciones por defecto de KNIME para la Regresión Logística, en Random Forest se ha definido como 500 el número de árboles, en el caso de la Red Neuronal y la Máquina de Vector de Soporte se ha usado la función de optimización de KNIME con diferentes particiones y los parámetros se han mantenido en el resto de los experimentos. En el caso de la Máquina de Vector Soporte se ha utilizado la opción de función de base radial.

Los resultados del Hold-Out con las 30 particiones 70%/30% definidas anteriormente y se exponen a continuación:

Tabla 14: Resultados de 30 hold-out 70/30

	Media	Minimo	Maximo	Desv Típica
Regresión Logística	0,8532	0,7917	0,9028	0,0300
Random Forest	0,9019	0,8333	0,9722	0,0334
Red Neuronal	0,8829	0,8194	0,9583	0,0364
Maquina Vector Soporte	0,8681	0,8056	0,9306	0,0377

Fuente: elaboración propia

CAPÍTULO 5: Experimentación

Como se puede observar los resultados son excelentes con independencia del clasificador que se utilice. La media está en todos los casos por encima del 85% de datos correctamente clasificados y el mínimo, salvo el caso de la Regresión Logística que se queda en el 79,17% por encima del 80%.

También se puede observar que existe una diferencia importante entre los valores máximos y mínimos para todos los clasificadores que fluctúan entre los 11,11 puntos de diferencia de la Regresión Logística hasta los 13,89 de la Red Neuronal y Random Forest quedándose la diferencia en el caso de la Máquina de Vector Soporte en 12,5 puntos. Esto confirma la necesidad de repetir el experimento porque de utilizar cualquiera de las particiones extremas, ya sea máxima o mínima, llevaría a conclusiones erróneas sobre la capacidad de predicción de los clasificadores.

2.1 Consistencia de los resultados

Para comprobar que los datos obtenidos en las 30 particiones utilizadas en el experimento son consistentes se ha repetido el experimento 3 veces adicionales y comprobado mediante test de hipótesis mediante la prueba t de medias emparejadas que las diferencias no son significativas con un intervalo de confianza del 5%.

CAPÍTULO 5: Experimentación

Tabla 15: Media de los resultados en diferentes particiones

	Partición 1	Partición 2	Partición 3	Partición 4*
Regresión Logística	0,8532	0,8546	0,8542	0,8523
Random Forest	0,9019	0,8991	0,8995	0,8991
Red Neuronal	0,8829	0,8861	0,8787	0,8875
Maquina Vector Soporte	0,8681	0,8681	0,8681	0,8708

*Resultados publicados en Applied Economic Letters

Fuente: elaboración propia

Tras analizar las diferencias entre las medias se comprueba mediante test de hipótesis para medias emparejadas con un nivel de significación del 5% que las diferencias no son significativas por lo que podemos concluir que las medias obtenidas representan la capacidad de predicción de nuestro modelo.

2.2 Comparativa con otros trabajos

Para comprobar la representación se han realizado dos análisis diferentes. En primer lugar, se analizan los resultados mínimos y máximos con los de otros autores; posteriormente se han utilizado las variables seleccionadas por algunos de estos autores en las mismas particiones que utilizamos para nuestro trabajo para comprobar si esos conjuntos de variables tienen un mejor o peor desempeño.

CAPÍTULO 5: Experimentación

Tabla 16: Mínimos y máximos de datos correctamente clasificados en diferentes trabajos

		Test set	Accuracy %
Modelo Propuesto	Reg. logística	Hold-out (30%)	79.17 – 90,18
	Random Forest		83,33 – 97,22
	Red Neuronal		81,94 – 95,83
	Máquina Vector Soporte		80,56 – 93,06
Bennell et al. (2006)		Hold-out (years 1998 & 1999)	31.80 - 40.40
Van Gestel et al. (2006)		Leave one out	39.70 – 44.80
Fernández-Navarro et al. (2013)		Hold-out (year 2010)	62.96 - 85.18
Sánchez-Monedero et al. (2014)		Hold-out (year 2010)	37.04 - 85.19
Ozturk et al. (2016)		10-fold cross validation	34.25 – 79.26
De Moor et al. (2018)		Hold-out (30%)	49.09 - 91.53

Fuente: elaboración propia

Como se puede comprobar los valores mínimos obtenidos en nuestro modelo están por encima de todos los valores mínimos del resto de los trabajos. Los valores máximos también están por encima si exceptuamos el valor máximo de la regresión logística que es inferior al valor máximo obtenido por De Moor et al. (2018).

Estos resultados no son del todo comparables pues utilizan distintos conjuntos de países y diferentes periodos. Para realizar una comparativa más completa se han introducido en nuestro conjunto de datos las variables utilizadas por otros autores y, utilizando las mismas particiones y los mismos clasificadores se ha rotado el modelo con estas variables. A las medias obtenidas se les ha realizado la prueba t de medias emparejadas para saber si las diferencias son significativas con un intervalo de confianza del 5%. Como resultado se obtienen los siguientes datos:

CAPÍTULO 5: Experimentación

Tabla 17: Medias de las 30 particiones con conjuntos de variables propuestos por otros autores:

	Modelo Propuesto (7 variables)	Ozturk et al. 2016 (9 variables)	Sánchez-Monedero et al. 2014 (11 variables)	Fernández-Navarro et al. 2013 (11 variables)
Random Forest	90,19	88,56*	88,66*	88,66*
Regresión Logística	85,32	85,04	85,23	85,09
Red Neuronal	88,28	87,41	86,90*	86,34*
Máquina Vector Soporte	86,81	82,69*	85,69	85,69

*Diferencia significativa con prueba t de medias emparejadas

Fuente: elaboración propia

Los resultados de nuestras variables presentan mejores medias siendo las diferencias significativas en todos los casos en el caso de Random Forest y en dos casos en el caso de la Red Neuronal y uno en la Máquina de Vector Soporte.

2.3 Análisis de sensibilidad

Para efectuar el análisis de sensibilidad se ha procedido a evaluar el modelo eliminando una a una las variables seleccionadas para comprobar el efecto de la eliminación. También se han añadido variables usadas por otros autores para comprobar si estas variables implican una mejora del modelo y nuestro modelo no está completo.

Tabla 18: Resultados eliminando una variable

	Modelo Propuesto	Variable eliminada						
		Deuda	Exc.Prim	Interés	PIB	PPC	Crecim.	Resilien.
Regresión Logística	0,8532	0,6912*	0,8532	0,8542	0,8560	0,8153*	0,8389*	0,7931*
Random Forest	0,9019	0,8727*	0,9028	0,8926*	0,8995	0,8773*	0,9023	0,8884*
Neural Network	0,8829	0,7792*	0,8792	0,8787	0,8778	0,8477*	0,8912	0,8509*
Máquina Vector Soporte	0,8681	0,8069*	0,8745	0,8440*	0,8681	0,8500*	0,8944	0,8426*

* Diferencia significativa con prueba t de medias emparejadas

Fuente: elaboración propia

CAPÍTULO 5: Experimentación

Como se puede observar todas las diferencias significativas son negativas, esto es, al eliminar la variable el modelo empeora. También es de destacar que sólo la Deuda, el PIB per cápita y la resiliencia generan diferencias significativas con los cuatro clasificadores, el Interés genera una diferencia con dos de los cuatro clasificares, el crecimiento con uno mientras que el Excedente primario y el PIB no generan diferencias significativas con ningún clasificador.

A continuación, se ha procedido a añadir variables utilizadas por otros investigadores a nuestro modelo para comprobar si esta adición genera diferencias significativas o no. Para ello se han seleccionado las siguientes variables.

Coste laboral unitario: Indicador de la competitividad de un país en el comercio internacional. Unidad de medida: Índice (IDX2015). Fuente OCDE.

Balanza por cuenta corriente: Indica cuánta importación neta de capital requiere un país. Unidad de medida: Porcentaje del PIB. Fuente Eurostat.

Eficacia del gobierno: Indicador de la calidad de la formulación e implementación de políticas, y la credibilidad y compromiso del gobierno con dichas políticas. Unidad de medida: Percentil. Fuente Banco Mundial.

Inflación: Muestra el cambio en el nivel del índice de precios para una cesta de bienes de uso común. Unidad de medida: Tasa de variación anual. Fuente Eurostat.

Desempleo: Indicador del tamaño de la brecha de producción y de la subutilización de los recursos. Unidad de medida: Tasa. Fuente Eurostat y Banco Mundial.

CAPÍTULO 5: Experimentación

Tabla 19: Resultados tras añadir una variable al modelo propuesto.

	Variable añadida					
	Original	Coste Laboral Unitario	Balanza Cuenta Corriente	Eficacia del Gobierno	Inflación	Desempleo
Regresión Logística	0,8532	0,8667	0,8880*	0,8532	0,8565	0,8532
Random Forest	0,9019	0,8991	0,9000	0,8986	0,8963*	0,9019
Red Neuronal	0,8829	0,8889	0,8949	0,8926	0,8708	0,8838
Máquina Vector Soporte	0,8681	0,8644	0,8611	0,8574*	0,8537*	0,8694

* Diferencia significativa con prueba t de medias emparejadas

Fuente: elaboración propia

Sólo en el caso de la Balanza por Cuenta Corriente y la Regresión Logística la adición de una variable genera una diferencia significativa en la capacidad de predicción del modelo.

A la vista de los resultados se puede concluir que el modelo es completo y suficiente pues salvo el caso de la Regresión Logística y la Balanza por Cuenta Corriente el modelo no mejora significativamente al añadir o eliminar alguna variable.

3. 10 FOLD

Se ha analizado un análisis de toda la muestra mediante una validación cruzada estratificada por clases. A diferencia del hold-out en este caso sólo se estima una vez en cada ocasión cada uno de los patrones. Esto nos permite hacer un análisis de la clasificación ordinal comparando toda la serie. Para evitar los efectos que las distintas particiones pueden tener en las estimaciones todas ellas

CAPÍTULO 5: Experimentación

se han realizados con los mismos diez subconjuntos de países y periodos obtenidos mediante una única partición aleatoria estratificada por clases.

3.1 Random Forest

Random Forest no es un método determinista. Esto hace que los resultados de aplicar el método sobre una misma partición no tengan que ser iguales. Por ello se ha repetido el experimento 30 veces con la misma partición.

La media y la desviación típica de las 30 repeticiones se muestra a continuación:

Tabla 20: Media y desviación típica de las 30 repeticiones de Random Forest

	MEDIA	DESVT
CCR	0,911	0,004
ACR	0,996	0,001
MAE	0,093	0,005
AMAE	0,224	0,012

Fuente: elaboración propia

Como se puede observar los resultados son altamente satisfactorios, con una media de clasificaciones correctas superior al 90% y una desviación típica contenida entre las distintas repeticiones del experimento.

Para analizar el efecto de las 30 repeticiones se ha calculado la media de cada uno de los atributos y su frecuencia en las repeticiones del experimento. Hay que destacar que en 228 de los 240 casos analizados el resultado es el mismo en cada una de las repeticiones sólo en 12 ocasiones en alguno de los resultados alguna de las repeticiones da un resultado distinto.

Los resultados correspondientes a la media de las 30 repeticiones difieren ligeramente de la media de los 30 resultados. Para poder hacer un análisis del

CAPÍTULO 5: Experimentación

clasificador se va a tomar como resultados las métricas de la clasificación promedio de las 30 repeticiones.

Tabla 21: Resultados de Random Forest

CCR	0,9167
ACR	0,9958
MAE	0,0875
AMAE	0,2137

Fuente: elaboración propia

Otro análisis que realizar es la consistencia de las 30 repeticiones y en qué casos se producen variaciones en la clasificación. Para ello se ha analizado la frecuencia relativa de la moda en las 30 repeticiones.

Tabla 22: Casos en los que las repeticiones de Random Forest dan resultados diferentes:

PAIS	AÑO	f MODA
Portugal	2011	57%
Portugal	2010	60%
Finlandia	2016	70%
España	2012	73%
Austria	2015	77%
Irlanda	2011	77%
Grecia	2017	87%
Italia	2012	87%
Grecia	2016	93%
Francia	2010	97%
Grecia	2008	97%
Grecia	2010	97%

Fuente: elaboración propia

Es destacable que sólo en 4 de los 12 casos la frecuencia de la moda está por debajo del 75%. Como se puede comprobar todos los casos corresponden a países que han cambiado de rating a lo largo del periodo y que, por tanto, son más difíciles de predecir.

Para realizar un análisis más detallado a continuación se presenta la matriz de confusión correspondiente a la media de las distintas clasificaciones.

CAPÍTULO 5: Experimentación

Tabla 23: Matriz de confusión de Random Forest

CLASE	CLASIFICADO COMO					
	Aaa	Aa	A	Baa	Ba-B	≤Caa
Aaa	142	1				
Aa	6	45				
A		2	8	2		
Baa			3	12	2	
Ba-B			1	2	8	
≤Caa					1	5

Fuente: elaboración propia

Se puede observar como la mayoría de los resultados están en la diagonal, en 5 casos el resultado ha sido una clase inferior a la original, en 14 el resultado ha sido una clase superior y en sólo un caso la diferencia ha sido superior a una clase siendo el resultado dos clases superiores al original.

La tabla que se presenta a continuación presenta el número de errores y su amplitud en función de cada uno de los países:

Tabla 24: Frecuencia de errores absolutos por países en Random Forest

	ERR. ABS.	
	1	2
Irlanda	7	
España	3	
Austria	2	
Grecia	2	1
Italia	2	
Finlandia	1	
Francia	1	
Reino Unido	1	

Fuente: elaboración propia

El análisis de esta tabla pone de manifiesto como los errores se concentran en los países que más han cambiado de rating, no existiendo ningún error en los 6 países que han mantenido la categoría durante todo el periodo (Alemania, Belgica, Dinamarca, Holanda, Luxemburgo, y Suecia) ni en el caso de Portugal.

3.2 Regresión Logística

Al ser la regresión logística un método determinista siempre debería dar el mismo resultado con la misma partición por lo que no se ha procedido a repetir el experimento. Los resultados obtenidos con los mismos 10 subconjuntos de patrones usados en Random Forest se presentan a continuación:

Tabla 25: Resultados Regresión Logística

CCR	0,8583
ACR	0,9833
MAE	0,1583
AMAE	0,3128

Fuente: elaboración propia

Aun siendo inferiores a los obtenidos con Random Forest siguen siendo unos buenos resultados. Seguidamente, se presenta la matriz de confusión correspondiente a la Regresión Logística.

Tabla 26: Matriz de confusión de la Regresión Logística

CLASE	CLASIFICADO COMO					
	Aaa	Aa	A	Baa	Ba-B	≤Caa
Aaa	134	8	1			
Aa	9	41		1		
A	1	3	6	2		
Baa			2	13	2	
Ba-B				2	8	1
≤Caa				1	1	4

Fuente: elaboración propia

En esta ocasión en 15 ocasiones se ha asignado un rating inferior al real y en 19 se ha asignado un rating superior. Como se observaba en el caso de Random Forest existen más ocasiones en la que el modelo asigna ratings superiores que en las que asigna ratings inferiores. En esta ocasión en 4 casos se asignan ratings con dos categorías de diferencia sobre el real.

A continuación, se representan el número de errores por países y amplitud.

CAPÍTULO 5: Experimentación

Tabla 27: Errores absolutos por países en la Regresión Logística

	ERR. ABS.	
	1	2
Austria	6	1
Irlanda	6	2
Reino Unido	5	
Grecia	4	1
Francia	3	
Finlandia	2	
España	2	
Italia	1	
Portugal	1	

Fuente: elaboración propia

En esta ocasión también se han quedado fuera los 6 países que han mantenido la categoría durante todo el periodo.

3.3 Red Neuronal

Al no ser la red neuronal un método determinista y poder dar diferentes resultados con el mismo conjunto de datos se ha repetido el experimento en 30 ocasiones para evitar el sesgo que una sola estimación podría obtener. Se han usado los mismos parámetros de configuración que para el caso del hold-out.

Los resultados representan la media de las 30 repeticiones y se muestran a continuación:

Tabla 28: Media y desviación típica de las 30 repeticiones de la Red Neuronal

	MEDIA	DESVT
CCR	0,879	0,016
ACR	0,988	0,016
MAE	0,139	0,050
AMAE	0,242	0,064

Fuente: elaboración propia

CAPÍTULO 5: Experimentación

Para el análisis del clasificador, de forma similar a lo realizado con Random Forest, se ha procedido al análisis de la clasificación promedio de las 30 repeticiones.

Tabla 29: Resultados de la Red Neuronal

CCR	0,896
ACR	1,000
MAE	0,104
AMAE	0,156

Fuente: elaboración propia

En el caso de la red neuronal existe un mayor componente probabilístico en el ajuste del modelo por lo que existe una mayor variación en los resultados obtenidos en las 30 repeticiones. En 76 ocasiones no coinciden los resultados en las distintas repeticiones; de estos casos 36 corresponden a casos en los que sólo difiere una de las 30 repeticiones. La siguiente tabla presenta los casos en los que la frecuencia de la moda es inferior al 70%.

Tabla 30: Casos en los que la frecuencia de la moda de la Red Neuronal es menor al 70%.

PAIS	AÑO	F Moda
Irlanda	2010	40,0%
Grecia	2017	46,7%
Grecia	2016	50,0%
Francia	2009	53,3%
Irlanda	2011	53,3%
Irlanda	2012	53,3%
Irlanda	2015	53,3%
Italia	2011	60,0%
Portugal	2009	60,0%
Reino Unido	2017	60,0%
Irlanda	2017	63,3%
Italia	2009	63,3%
Italia	2012	63,3%
Portugal	2011	63,3%
Reino Unido	2015	63,3%
Grecia	2009	66,7%
Portugal	2010	66,7%
España	2009	66,7%

Fuente: elaboración propia

CAPÍTULO 5: Experimentación

Hay que destacar que en 2 casos la frecuencia de la moda es inferior al 50%: el caso de Irlanda 2010 en el que el rating obtenido oscila entre 4 categorías distintas y el de Grecia 2010 en el que el rating estimado oscila en tres categorías distintas.

Seguidamente se presenta la matriz de confusión de la media de las 30 repeticiones de la Red Neuronal.

Tabla 31: Matriz de confusión de la Red Neuronal

CLASE	CLASIFICADO COMO					
	Aaa	Aa	A	Baa	Ba-B	≤Caa
Aaa	137	6				
Aa	7	40	4			
A		2	8	2		
Baa			2	14	1	
Ba-B					11	
≤Caa					1	5

Fuente: elaboración propia

Se puede observar que a pesar de la elevada dispersión de resultados de cada una de las repeticiones los resultados son bastante consistentes al clasificarse todos los casos en el entorno de un error absoluto igual o inferior a 1. En 13 ocasiones el modelo ha clasificado a los patrones por debajo de su rating y en 12 por encima, presentando por tanto un comportamiento diferente al de otros clasificadores.

CAPÍTULO 5: Experimentación

Tabla 32: Errores absolutos por países en la Red Neuronal

	ERR. ABS. 1
Irlanda	5
Austria	4
Francia	3
Reino Unido	3
Finlandia	2
Grecia	2
Italia	2
Portugal	2
Bélgica	1
España	1

Fuente: elaboración propia

En este caso, en el caso de Bélgica, país que ha mantenido la clase en todo el periodo, hay un error de predicción en el ejercicio 2007 en el que el modelo le asigna la máxima categoría “Aaa” frente a la “Aa” que le corresponde.

3.4 Máquina de Vector de Soporte

Para la máquina de vector soporte se ha modificado la configuración utilizada para el hold-out para conseguir mejor desempeño en la clasificación ordinal. Al ser la máquina de vector soporte un método basado en la optimización determinista solo se ha realizado una vez el experimento.

Tabla 33: Resultados de la Máquina de Vector Soporte

CCR	0,900
ACR	0,983
MAE	0,117
AMAE	0,203

Fuente: elaboración propia

Seguidamente presentamos la matriz de confusión

Tabla 34: Matriz de confusión de la Máquina de Vector Soporte

CAPÍTULO 5: Experimentación

CLASE	CLASIFICADO COMO					
	Aaa	Aa	A	Baa	Ba-B	≤Caa
Aaa	137	5	1			
Aa	7	44				
A			7	4	1	
Baa		2	2	13		
Ba-B				1	9	1
≤Caa						6

Fuente: elaboración propia

En esta matriz de confusión se observa como con este método existe una mayor dispersión de los resultados, existiendo más datos con diferencias mayores a la unidad. En este caso en 10 ocasiones se asigna una categoría inferior a la que le corresponde y en 2 dos categorías inferiores. En cuanto a los casos clasificados por encima de lo que les corresponde en 10 ocasiones se le asigna una categoría superior a la que le corresponde en y en 2 ocasiones se asigna dos categorías superiores a las que les. En total en 12 ocasiones se clasifica por debajo y en 12 por encima mostrando este clasificador un perfil equilibrado.

En cuanto a los errores absolutos cometidos por países se presentan en la siguiente tabla

Tabla 35: Errores absolutos por países en la Máquina de Vector Soporte

	ERR. ABS.	
	1	2
Austria	4	1
Irlanda	4	1
España	4	
Finlandia	2	
Grecia	2	1
Reino Unido	2	
Francia	1	
Italia	1	
Portugal		1

Fuente: elaboración propia

CAPÍTULO 5: Experimentación

n esta ocasión también se han quedado fuera de los errores los países que no han cambiado de categoría a lo largo del periodo analizado.

3.5 Clasificadores Ordinales

Para verificar de forma más completa la idoneidad de las variables seleccionadas para la determinación del rating de los países estudiados se ha procedido a la estimación de ratings mediante el software de clasificación ordinal ORCA (Ordinal Regression and Classification Algorithms) (Sánchez-Monedero et al., 2019) es un entorno de programación en MATLAB donde se incluyen una serie de clasificadores ordinales. Este software permite una fácil comparación entre el comportamiento de distintos algoritmos por lo que es muy útil para el análisis del comportamiento de las variables seleccionadas en este estudio con diversos clasificadores ordinales.

Para realizar este análisis se ha partido de los mismos 10 subconjuntos que se usaron en los clasificadores de KNIME para que los resultados sean lo más comparables posible. Para la realización de los distintos experimentos se han utilizado los valores por defecto dejando que la aplicación realice su optimización de parámetros. El software proporciona diversas métricas de clasificación de las que se han seleccionado las utilizadas en los experimentos anteriores para la comparación.

CAPÍTULO 5: Experimentación

Tabla 36: Resultados con clasificadores ordinales

	CCR	ACR	MAE	AMAE
CSSVC	0,892	0,983	0,125	0,258
ELMOP	0,858	0,958	0,196	0,451
HPOLD	0,892	0,992	0,117	0,174
KDLOR	0,879	0,988	0,133	0,280
NNOP	0,896	0,992	0,113	0,187
NNPOM	0,871	0,992	0,138	0,280
OPBE	0,879	0,992	0,133	0,205
ORBOOST	0,892	0,996	0,113	0,216
REDSVM	0,888	0,992	0,121	0,232
SVC1V1	0,883	0,979	0,142	0,244
SVC1VA	0,888	0,983	0,129	0,208
SVOREX	0,879	0,988	0,146	0,335
SVOORIMLIN	0,871	0,992	0,138	0,257
SVORIM	0,888	0,988	0,138	0,336
SVR	0,879	0,983	0,146	0,359

Fuente: elaboración propia

Se puede comprobar que las variables seleccionadas tienen un excelente comportamiento con clasificadores ordinales los datos correctamente clasificados (CCR) oscilan entre un 85,8% y un 89,6%, los datos clasificados en el entorno de una clase del correcto (ACR) oscilan entre el 95,8% y el 90,6% , el error absoluto medio (MAE) oscila entre el 19,6 y el 11,3 y el error absoluto medio por clase (AMAE) oscila entre el 0,451 y el 0,174.

3.6 Análisis conjunto

Tras haber utilizado 19 clasificadores, 4 de KNIME y 15 de ORCA se va a analizar por una parte el desempeño de usar un ensamble de todos los clasificadores usados, asignando a cada país y periodo la clase media asignada por los 19 clasificadores. Posteriormente se analizan cómo se comportan los patrones en los distintos clasificadores.

CAPÍTULO 5: Experimentación

Tabla 37: Resultados del ensamble de todos los clasificadores

CCR	0,9125
ACR	0,9958
MAE	0,0917
AMAE	0,1496

Fuente: elaboración propia

Los resultados se pueden considerar como muy buenos pero salvo en el caso del AMAE, en el que el resultado del ensamble es mejor al mejor obtenido individualmente por todos los clasificadores, son inferiores a los obtenidos por el mejor clasificador Random Forest. A continuación, se presenta la matriz de confusión.

Tabla 38: Matriz de confusión del ensamble de clasificadores

CLASE	CLASIFICADO COMO					
	Aaa	Aa	A	Baa	Ba-B	≤Caa
Aaa	138	4	1			
Aa	5	44	2			
A		1	10	1		
Baa			4	12	1	
Ba-B				1	10	
≤Caa					1	5

Fuente: elaboración propia

En este caso en 9 ocasiones se han clasificado los patrones por debajo del rating asignado y en 12 por encima.

Otro aspecto interesante para estudiar en la clasificación es los errores que se comenten en cada clase. Para ello se ha calculado el promedio del error absoluto de cada clase.

CAPÍTULO 5: Experimentación

Tabla 39: Promedio de error absoluto por clase

CLASE	MEDIA	DESV. TIP.
Aaa	0,046	0,0142
Aa	0,193	0,0473
A	0,346	0,1950
Baa	0,341	0,0932
Ba-B	0,273	0,1249
≤Caa	0,351	0,3138

Fuente: elaboración propia

Como se puede comprobar existe una significativa diferencia entre la clase Aaa y el resto. Esto es debido a que al ser la clase más numerosa el modelo tiene mayor capacidad de aprendizaje.

Por último, para analizar el comportamiento de los distintos clasificadores en nuestra muestra se ha comprobado cómo se comporta cada patrón con los 19 clasificadores y se han dividido los patrones en cuatro categorías en función del número de clasificadores que generan una correcta clasificación:

- Patrones que son correctamente clasificados por todos los algoritmos.
- Patrones que son bien clasificados por la mayoría de los algoritmos, esto es en 10 de los 19 algoritmos presentan una correcta clasificación.
- Patrones que son correctamente clasificados por una minoría de algoritmos, es decir, solo son correctamente clasificados entre 1 y 9 algoritmos.
- Patrones incorrectamente clasificados por todos los algoritmos.

A continuación, se presenta el resultado del análisis realizado:

CAPÍTULO 5: Experimentación

Tabla 40: Clasificación de los patrones en los distintos clasificadores.

CORR. CLASIF.	CASOS	%	% ACUM.
19	176	73,33%	73,33%
10-18	40	16,67%	90,00%
1-9	21	8,75%	98,75%
0	3	1,25%	100,00%

Fuente: elaboración propia

Es de destacar la gran cantidad de patrones que son correctamente clasificados por la totalidad de los algoritmos y como el 90% de los patrones son correctamente clasificados por la mayoría de los algoritmos.

En cuanto a los tres casos en los que ningún algoritmo ha sido capaz de clasificar correctamente el patrón estos son los siguientes:

- Austria 2016
- Finlandia 2017
- Irlanda 2015

En el caso de Austria todos los clasificadores dan la máxima categoría Aaa mientras la agencia baja de rating la deuda de Austria a la categoría Aa en ese año. Sin embargo, en el ejercicio anterior 15 clasificadores asignaron esa clase a la deuda de Austria y 7 dos ejercicios antes.

El caso de Finlandia es similar, aunque en este caso la agencia bajó la categoría en el ejercicio anterior en el que 10 clasificadores reconocieron esa categoría e incluso 2 asignaron la categoría inferior A.

En Irlanda en 2015 mientras la Agencia da una categoría Baa, nuestros clasificadores asignan en 17 casos la categoría superior A y en 2 casos asignan dos categorías por encima de la asignada por la Agencia Aa. La categoría A fue reconocida por la agencia en el ejercicio 2016.

CAPÍTULO 5: Experimentación

Estos casos ponen de manifiesto las dificultades que existen a la hora de estimar los ratings en casos en los que se produce una recuperación de categoría por lo que procede analizar los efectos de un filtrado de datos en los que se incluya un retraso a la hora de permitir que un país suba de categoría.

4. FILTRADO DE DATOS

Para el filtrado de datos se parte de la hipótesis de que en todos los casos los ratings bajan nada más se conoce el empeoramiento de un país, pero tardan en subir más por la asimetría en los movimientos de rating. Se analizado como varían las métricas si se retrasa la subida uno y dos años siendo sensiblemente mejores las métricas en el caso de que se demoren las subidas un solo ejercicio por lo que se ha estudiado este supuesto.

CAPÍTULO 5: Experimentación

Tabla 41: Datos correctamente clasificados tras el filtrado

	ORIG.	FILTRADA	Δ	$\Delta\%$
RF	0,917	0,925	0,008	0,909%
SVM	0,900	0,917	0,017	1,852%
MLP	0,896	0,917	0,021	2,326%
REGLOG	0,858	0,883	0,025	2,913%
CSSVC	0,892	0,925	0,033	3,738%
ELMOP	0,858	0,904	0,046	5,340%
HPOLD	0,892	0,900	0,008	0,935%
KDLOR	0,879	0,888	0,008	0,948%
NNOP	0,896	0,921	0,025	2,791%
NNPOM	0,871	0,883	0,013	1,435%
OPBE	0,879	0,896	0,017	1,896%
ORBOOST	0,892	0,904	0,013	1,402%
REDSVM	0,888	0,900	0,013	1,408%
SVC1V1	0,883	0,900	0,017	1,887%
SVC1VA	0,888	0,908	0,021	2,347%
SVOREX	0,879	0,896	0,017	1,896%
SVOORIMLIN	0,871	0,892	0,021	2,392%
SVORIM	0,888	0,896	0,008	0,939%
SVR	0,879	0,896	0,017	1,896%

Fuente: elaboración propia

Como se puede observar en todos los clasificadores el filtrado de los datos clasificados supone una mejora apreciable. Estas mejoras oscilan entre los 0,08 de Random Forest, que supone clasificar bien dos patrones más en la serie filtrada de los 20 incorrectamente clasificados en la serie original (10%), a los 0,046 de Extrme Learning Machine with Ordered Partition que supone clasificar correctamente 11 patrones más de los 34 incorrectamente clasificados en la serie original (32%). El filtrado, por tanto, tiene mayor capacidad de mejora en las series peor clasificadas.

En cuanto al dato de los patrones clasificados en el entorno de una clase de su clase real existen muy pocas posibilidades de mejora.

CAPÍTULO 5: Experimentación

Tabla 42: Datos clasificados en el entorno de una clase tras el filtrado

	ORIG.	FILTRADA	Δ	$\Delta\%$
RF	0,996	1,000	0,004	0,418%
SVM	0,983	0,975	-0,008	-0,847%
MLP	1,000	1,000	0,000	0,000%
REGLOG	0,983	0,988	0,004	0,424%
CSSVC	0,983	0,979	-0,004	-0,424%
ELMOP	0,958	0,967	0,008	0,870%
HPOLD	0,992	0,988	-0,004	-0,420%
KDLOR	0,988	0,996	0,008	0,844%
NNOP	0,992	0,983	-0,008	-0,840%
NNPOM	0,992	0,996	0,004	0,420%
OPBE	0,992	0,988	-0,004	-0,420%
ORBOOST	0,996	0,992	-0,004	-0,418%
REDSVM	0,992	0,992	0,000	0,000%
SVC1V1	0,979	0,983	0,004	0,426%
SVC1VA	0,983	0,975	-0,008	-0,847%
SVOREX	0,988	0,988	0,000	0,000%
SVOORIMLIN	0,992	1,000	0,008	0,840%
SVORIM	0,988	0,988	0,000	0,000%
SVR	0,983	0,988	0,004	0,424%

Fuente: elaboración propia

En este caso las mejoras son muy pocas. Los patrones en esta categoría oscilan entre 0 y 10 por lo que las posibilidades de mejorar en este aspecto son muy limitadas. El efecto del filtrado sobre esta métrica oscila entre -0,008 que supone que el filtrado empeora la métrica en 2 patrones y 0,008 que supone que el filtrado mejora la métrica en 2 patrones.

En el caso del error absoluto medio si se produce una mejora en todos los clasificadores.

CAPÍTULO 5: Experimentación

Tabla 43: Error Absoluto Medio tras el Filtrado.

	ORIG.	FILTRADA	Δ	$\Delta\%$
RF	0,088	0,075	-0,013	-14,286%
SVM	0,117	0,108	-0,008	-7,143%
MLP	0,104	0,083	-0,021	-20,000%
REGLOG	0,158	0,129	-0,029	-18,421%
CSSVC	0,125	0,096	-0,029	-23,333%
ELMOP	0,196	0,133	-0,063	-31,915%
HPOLD	0,117	0,113	-0,004	-3,571%
KDLOR	0,133	0,117	-0,017	-12,500%
NNOP	0,113	0,096	-0,017	-14,815%
NNPOM	0,138	0,121	-0,017	-12,121%
OPBE	0,133	0,125	-0,008	-6,250%
ORBOOST	0,113	0,104	-0,008	-7,407%
REDSVM	0,121	0,108	-0,013	-10,345%
SVC1V1	0,142	0,117	-0,025	-17,647%
SVC1VA	0,129	0,117	-0,013	-9,677%
SVOREX	0,146	0,121	-0,025	-17,143%
SVOORIMLIN	0,138	0,108	-0,029	-21,212%
SVORIM	0,138	0,121	-0,017	-12,121%
SVR	0,146	0,121	-0,025	-17,143%

Fuente: elaboración propia

La mejora, disminución del valor del error absoluto medio, oscila entre el 0,004 de Hierarchical Partial Order Label Decomposition, que supone una mejora porcentual del 3,6% sobre un error absoluto medio particularmente bajo, y el 0,063 de Extreme Learning Machine with Ordered Partition que supone una mejora del 31,9% respecto al error absoluto medio de la serie sin filtrar.

CAPÍTULO 5: Experimentación

Tabla 44: Error Absoluto Medio por Ponderado por Clase tras el Filtrado

	ORIG.	FILTRADA	Δ	$\Delta\%$
RF	0,214	0,166	-0,047	-22,2%
SVM	0,203	0,189	-0,014	-7,1%
MLP	0,156	0,115	-0,041	-26,2%
REGLOG	0,313	0,184	-0,128	-41,0%
CSSVC	0,258	0,122	-0,136	-52,8%
ELMOP	0,451	0,194	-0,258	-57,0%
HPOLD	0,174	0,155	-0,019	-11,1%
KDLOR	0,280	0,190	-0,090	-32,2%
NNOP	0,187	0,130	-0,057	-30,4%
NNPOM	0,280	0,221	-0,059	-21,0%
OPBE	0,205	0,182	-0,023	-11,4%
ORBOOST	0,216	0,168	-0,048	-22,2%
REDSVM	0,232	0,135	-0,096	-41,5%
SVC1V1	0,244	0,200	-0,043	-17,8%
SVC1VA	0,208	0,144	-0,064	-30,9%
SVOREX	0,335	0,202	-0,132	-39,6%
SVOORIMLIN	0,257	0,181	-0,075	-29,3%
SVORIM	0,336	0,222	-0,114	-34,0%
SVR	0,359	0,222	-0,138	-38,3%

Fuente: elaboración propia

En el caso del AMAE también se producen mejoras en todos los clasificadores y estas son más significativas al ser este parámetro mayor que el MAE y, por tanto, más susceptible de mejora. Las mejoras en este caso oscilan entre el 0,014 de la Máquina de Vector Soporte y el 0,258 de Extreme Learning Machine with Ordered Partition que suponen una disminución del 7,1% y el 57% respecto al AMAE de la serie original.

Tras el análisis, se comprueba que el filtrado supone una mejora de la clasificación en todos los algoritmos, mejorando más cuanto peor sea la clasificación previa al filtrado.

5. ANÁLISIS POR PAÍSES

Para analizar los resultados de la clasificación por países se ha dividido el conjunto en tres grupos:

- Países referentes: esto es, países que han mantenido el rating Aaa en todos los ejercicios.
- Países estables: países que han variado dos o menos clases.
- Países inestables: países que han variado más de dos clases.

Sobre estos grupos de países se ha analizado cómo funciona las previsiones en el conjunto de los clasificadores, cómo funciona el ensamble de clasificadores tras el filtrado y posteriormente se ha realizado un análisis del comportamiento de Random Forest en las 30 repeticiones después del filtrado.

5.1 Países referentes

Estos países son Dinamarca, Holanda, Alemania, Luxemburgo y Suecia. Estos países han sido correctamente clasificados en la clase Aaa en todos los periodos por todos los clasificadores exceptuando los dos casos siguientes:

- Alemania 2008 en el que el clasificador SVC1VA ha asignado la categoría inmediatamente inferior.
- Alemania 2011 en el que el clasificador ORBOOST también ha asignado la categoría Aa.

Al filtrar los datos correspondientes a estos países los errores pasan a ser 2 en cada uno de estos clasificadores pues la categoría inferior se mantiene por un segundo año.

CAPÍTULO 5: Experimentación

Se puede considerar que en este conjunto de países el modelo es totalmente preciso pues de los 1520 casos analizados (5 países por 16 periodos por 19 clasificadores) solo en dos ocasiones se ha asignado la clase inmediatamente inferior.

Esta conclusión es extrapolable al ensamble de clasificadores que asigna correctamente la categoría en todas las ocasiones.

En el caso de Random Forest las 30 repeticiones del experimento asignan a estos países la máxima categoría no existiendo ningún error en ninguna de las repeticiones.

5.2 Países estables

En este conjunto de países se pueden distinguir tres subtipos:

- Bélgica, que ha permanecido de forma estable en la segunda categoría durante todo el periodo.
- Austria, Finlandia, Francia y Reino Unido, que partiendo de la categoría superior bajan a la siguiente a lo largo del periodo.
- Italia que, partiendo de la clase Aa baja dos categorías durante el periodo hasta la clase Baa.

5.2.1 Bélgica

Si analizamos otros clasificadores se han producido los siguientes errores en la clasificación:

CAPÍTULO 5: Experimentación

- 2002 7 clasificadores (HPOLD, KDLOR, NNPO, REDSVM, SVOREX, SVORIM y SVR) proponen la categoría inmediatamente inferior generando un error absoluto de 1 en cada caso.
- 2007 3 clasificadores en la proponen la categoría superior (MLP, OPBE y ORBOOST).
- 2008 5 clasificadores proponen la categoría superior (CSSVC, OPBE, REDSVM, SVC1V1 y SVC1VA). Sólo OPBE
- 2011 ORBOOST genera una categoría por debajo.

El Ensamble de clasificadores proporciona en todos los casos la categoría correcta.

también Random Forest, en todas sus iteraciones, clasifica su deuda como Aa por lo que el error atribuible a este país es 0 en todas las repeticiones del experimento.

Cabe destacar que Bélgica es un país con un alto índice de endeudamiento desde el inicio de la serie. Esto es lo que ha generado que algunos clasificadores interpreten una categoría inferior en los primeros ejercicios. Este endeudamiento se redujo hasta el año 2007, que algunos algoritmos interpretan como una mejora de su categoría para volverla a subir. Su rápida salida de la crisis hace que su resiliencia sea particularmente baja y, por tanto, no afecte a la clasificación.

Nos encontramos por tanto en un País con una clasificación fácil de entender para la mayoría de los algoritmos utilizados.

CAPÍTULO 5: Experimentación

5.2.2 Austria

Este país ha permanecido en la máxima categoría Aaa hasta el ejercicio 2016 y 2017 en los que la agencia Moody's rebaja el rating a la segunda categoría Aa.

Hasta el ejercicio 2008 todos los clasificadores dan resultados correctos para Austria. A partir de 2009 se han producido los siguientes errores:

- 2009, la Regresión Logística, REDSVM y SVOORIMLIM asignan una categoría inferior.
- 2010, la Regresión Logística, NNPO, y SVOORIMLIM asignan una categoría inferior
- 2011, KDLOR, NNPO, SVOORIMLIN y SVR asignan una categoría por debajo; la Máquina de Vector Soporte, la Regresión Logística, CSSVC, ELMOP, HPOLD, NNOP, ORBOOST, REDSVM, SVC1V1, SVC1VA, SVOREX y SVORIM asignan dos categorías por debajo y OPBE asigna tres categorías por debajo. Por tanto, de los 19 clasificadores 4 asignan una categoría menos, 12 asignan dos categorías menos y uno tres categorías menos y solo Random Forest y la Red Neuronal asignan la categoría correcta.
- 2014, la Máquina de Vector Soporte, la Red Neuronal, La Regresión Logística, CSSVC, NNOP, NNPO y SVC1VA asignan una categoría inferior.
- 2015, todos los clasificadores menos Random Forest, OPBE, SVOORIMLIM y SVR asignan una categoría inferior.
- 2016, este es el año en que Moody's rebaja la categoría, pero todos los clasificadores asignan la máxima categoría. Al aplicar el filtrado se

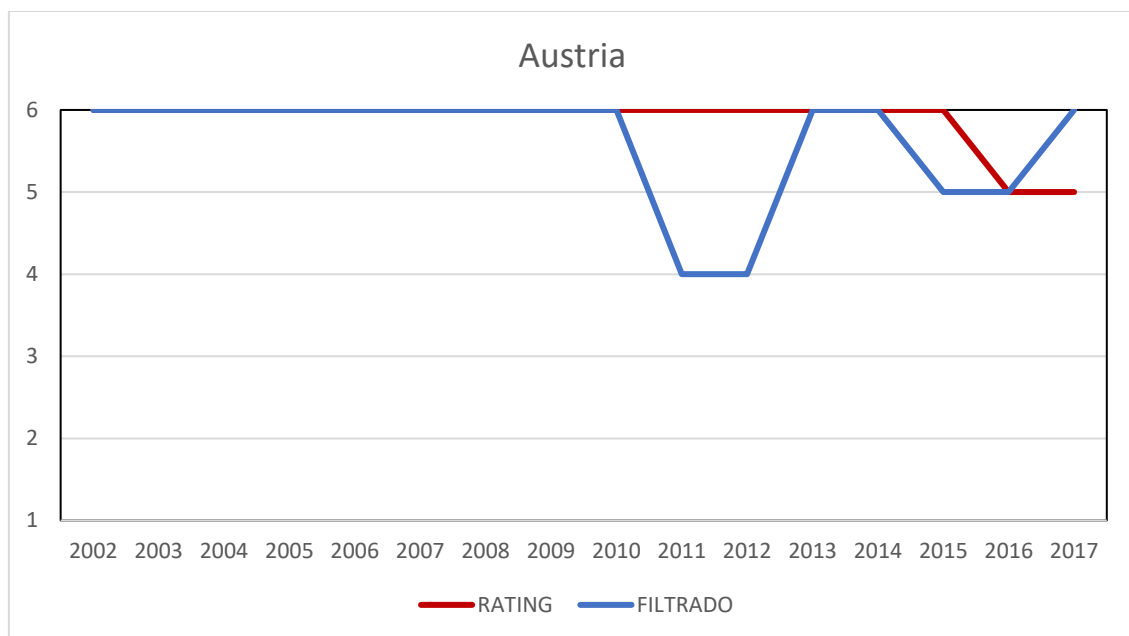
CAPÍTULO 5: Experimentación

corregirían los errores de los 15 clasificadores que asignaron la categoría inferior en 2015.

- 2017, todos los clasificadores menos ORBOOST asignan la máxima categoría.

Estas discrepancias hacen que mayoritariamente los clasificadores, con los datos referidos al ejercicio 2012, asignan dos categorías menos. Tras el filtrado, esta diferencia se mantiene en el ejercicio siguiente.

Gráfico 9: Resultados del ensamble de clasificadores para Austria

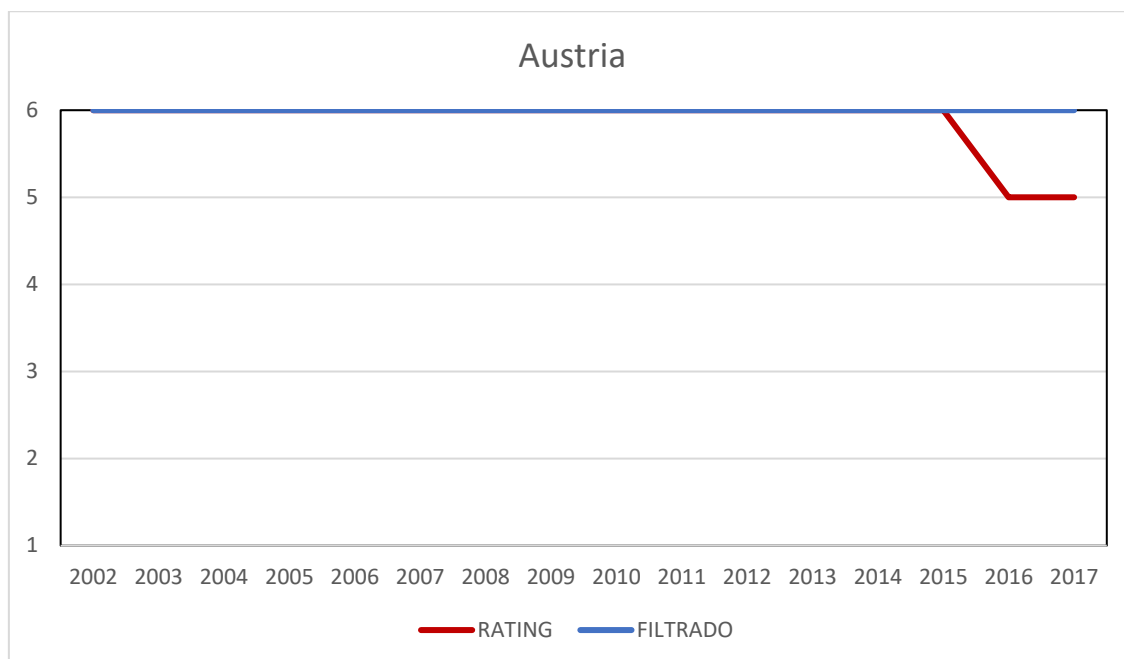


Fuente: elaboración propia

Si analizamos las 30 repeticiones de Random Forest observamos como en todos los años menos en el 2015 todas las iteraciones coinciden. En este año en 7 de las 30 repeticiones el modelo asigna una categoría inferior.

CAPÍTULO 5: Experimentación

Gráfico 10: Resultados de Random Forest para Austria



Fuente: elaboración propia

Nos encontramos con un país con dos episodios en los que empeoraron sus condiciones, un primer episodio tras la crisis financiera de 2008 que no tuvo reflejo en el rating, pero con un momento crítico en el año 2011 en el que 17 clasificadores asignaron categorías inferiores y un segundo episodio en el año 2015, anticipado por tres clasificadores en 2014, y que ha supuesto una bajada de rating en 2016, año en el que nuestros clasificadores dan por superado el periodo no así la agencia de rating. Nuestros clasificadores detectan una anomalía en el ejercicio 2011, adelantan de forma clara la bajada y proponen una posterior subida que no se ha producido.

CAPÍTULO 5: Experimentación

5.2.3 Finlandia

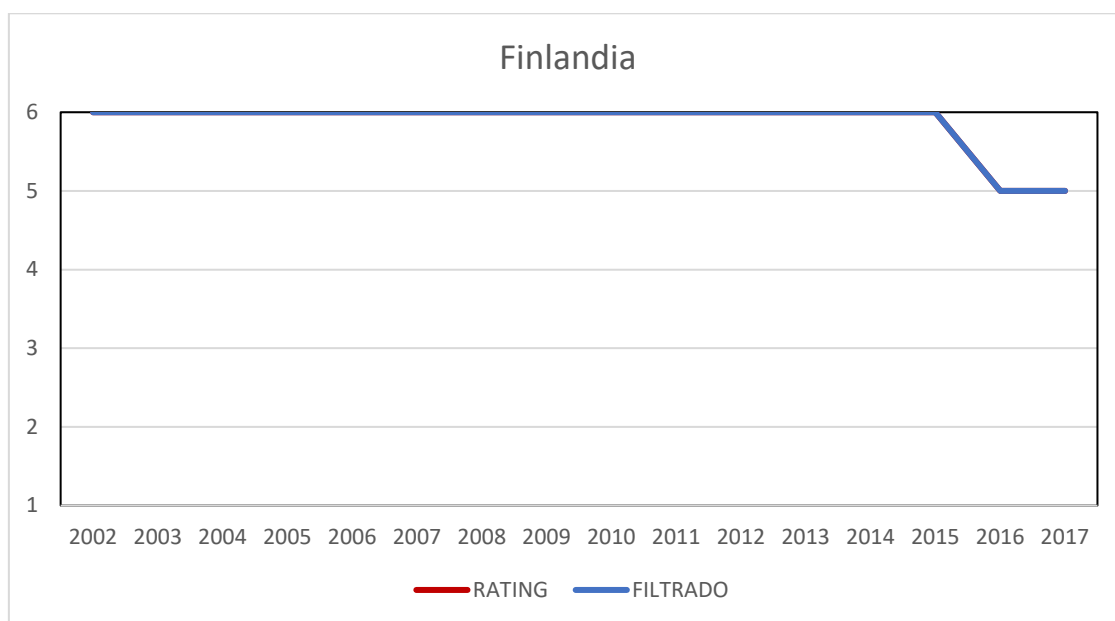
Finlandia se ha mantenido en la máxima categoría hasta el ejercicio 2016 en el que la agencia Moody's baja su categoría a Aa. Los clasificadores han asignado la categoría correcta salvo las siguientes excepciones:

- 2010, KDLOR asigna una categoría menos
- 2015, NNPOM asigna una categoría menos
- 2016, año en el que baja de rating, la Máquina de Vector Soporte, la Red Neuronal, la Regresión Logística, CSSVC, ELMOP y SVC1VA asignan una categoría más y HPOLD, KDLOR y NNPOM asignan una categoría menos.
- 2017, todos los clasificadores asignan la máxima categoría, pero la agencia mantiene la categoría inferior. El filtrado hace que en la mayor parte de los casos este error se corrija.

El ensamble de clasificadores recoge la bajada en 2016 y la mantiene en 2017 por el filtrado de los datos.

CAPÍTULO 5: Experimentación

Gráfico 11: Resultado del ensamble de clasificadores para Finlandia

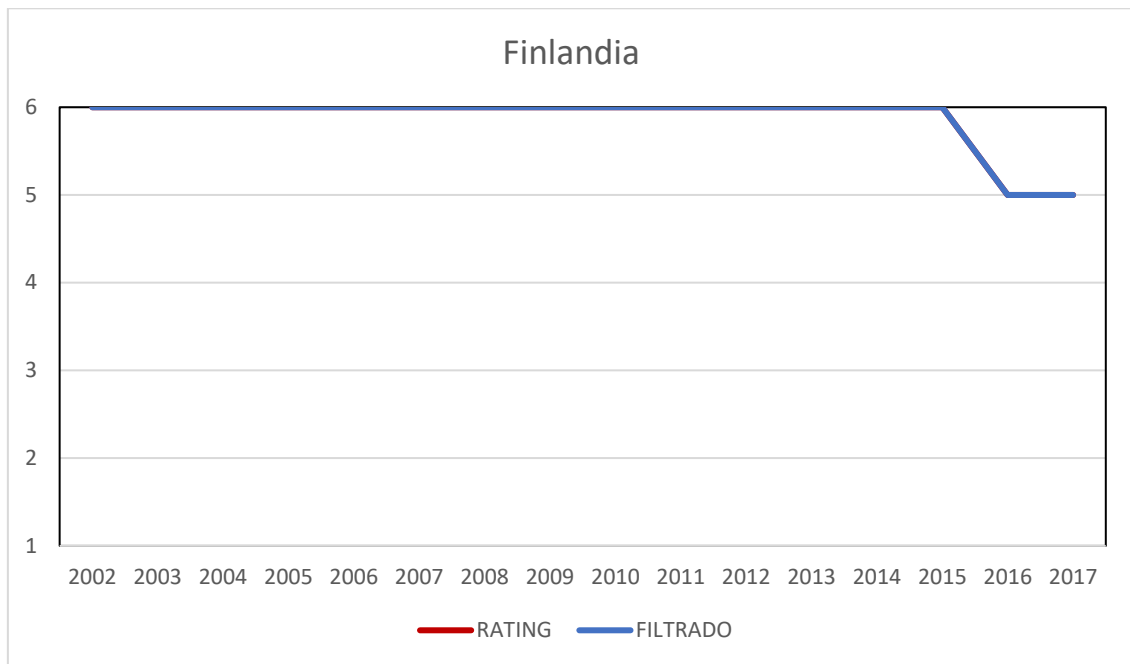


Fuente: elaboración propia

En el caso de Random Forest los resultados coinciden en todas las repeticiones salvo en el ejercicio 2016 en el que en 9 de las 30 repeticiones se mantiene la máxima categoría. En el 2017 el filtrado de los datos hace que los resultados coincidan en 21 de las 30 repeticiones con los de la agencia Moody's.

CAPÍTULO 5: Experimentación

Gráfico 12: Resultados de Random Forest para Finlandia



Fuente: elaboración propia

En este caso coinciden el ejercicio en el que se detecta la bajada de rating por el modelo propuesto y el asignado por la agencia de rating y, aunque se detecta una mejora de las condiciones en 2017 la agencia no vuelve a subir el rating tras haberlo bajado en el ejercicio anterior.

5.2.4 Francia

Francia se mantiene en la máxima categoría hasta el ejercicio 2012 en el que la agencia Moody's rebaja el rating a una categoría inferior. Esta bajada es anticipada por distintos clasificadores como se evidencia de las diferencias entre el rating asignado por la agencia y el estimado por los clasificadores:

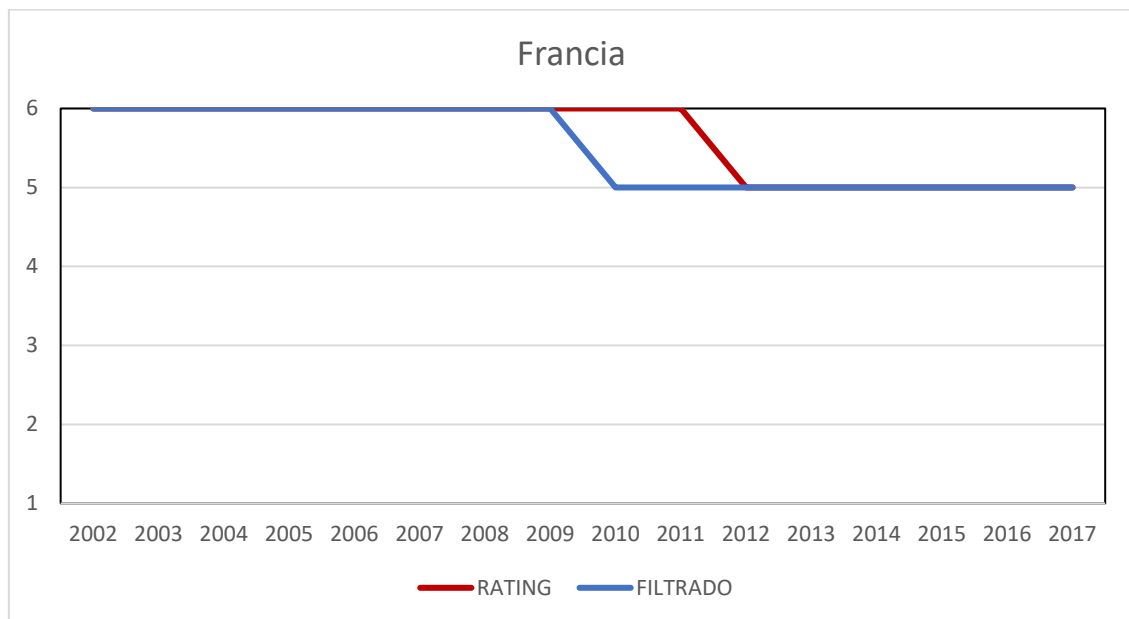
- 2009, la Red Neuronal, la Regresión Logística, REDSVM, SVOREX, SVOORIMLIN y SVR asignan una categoría inferior.

CAPÍTULO 5: Experimentación

- 2010, la Red Neuronal, la Regresión Logística, ELMOP, HPOLD, KDLOR, SVC1V1, SVOREX, SVOORIMLIN, SVORIM y SVR asignan una categoría inferior.
- 2011, todos los clasificadores menos OPBE asignan una categoría inferior.
- 2012, CSSVC, ELMOP, KDLOR, NNOP, NNPOM, OPBE, SVC1VA y SVR asignan una categoría superior.

El ensamble de clasificadores, por tanto, anticipa en dos ejercicios la bajada de rating que realiza la agencia.

Gráfico 13: Resultados del ensamble de clasificadores para Francia

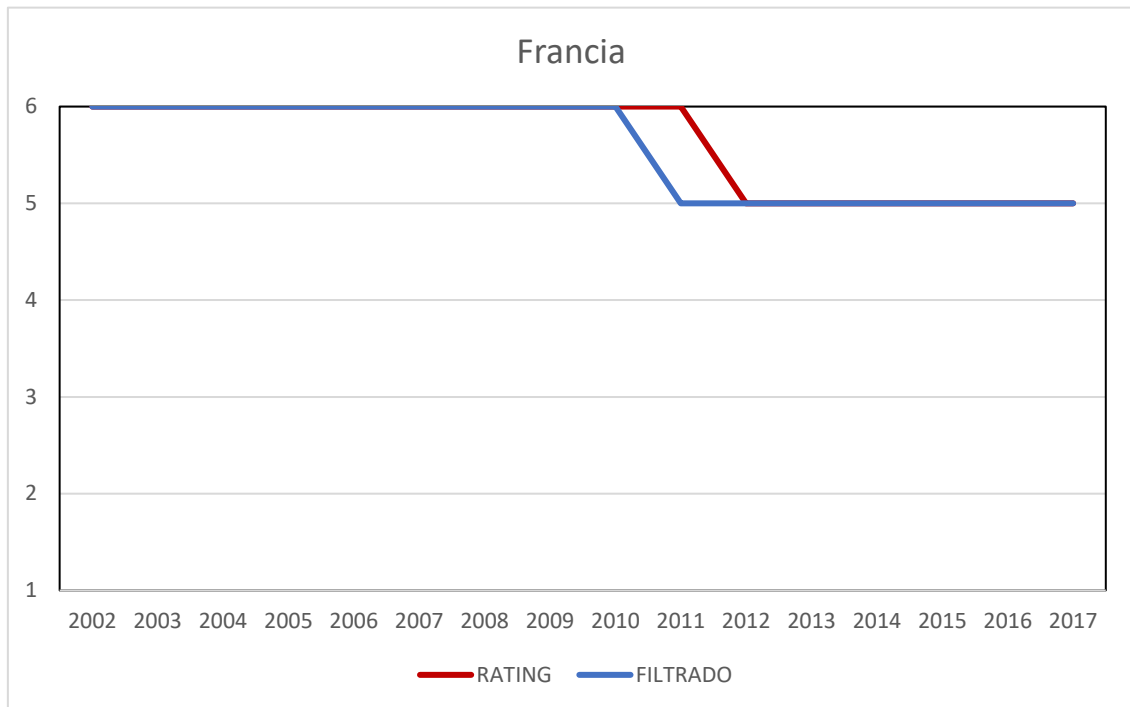


Fuente: elaboración propia

En el caso de Random Forest, en una sola de las 30 repeticiones se anticipa la bajada de rating en dos ejercicios y en el resto del periodo coinciden las 30 repeticiones anticipando en un ejercicio la bajada de rating de la agencia.

CAPÍTULO 5: Experimentación

Gráfico 14: Resultados de Random Forest para Francia



Fuente: elaboración propia

Se trata de un caso en el que hay una evidente anticipación de la bajada de rating anticipándose en 6 estimaciones en 3 años, en 10 en dos años y en 18 en un año. Aparentemente la agencia demoró la bajada de rating esperando una recuperación de las magnitudes macroeconómicas que no se produjeron.

5.2.5 Reino Unido

El caso de Reino Unido es similar al de Francia; mantiene la máxima categoría hasta que en 2012 la agencia rebaja una categoría. Las diferencias en las clasificaciones son las siguientes:

- 2011, la Máquina de Vector Soporte, la Red Neuronal, la Regresión Logística, ELMOP, HPOLD, KDLOR, REDSVM, SVC1V1, SVC1VA,

CAPÍTULO 5: Experimentación

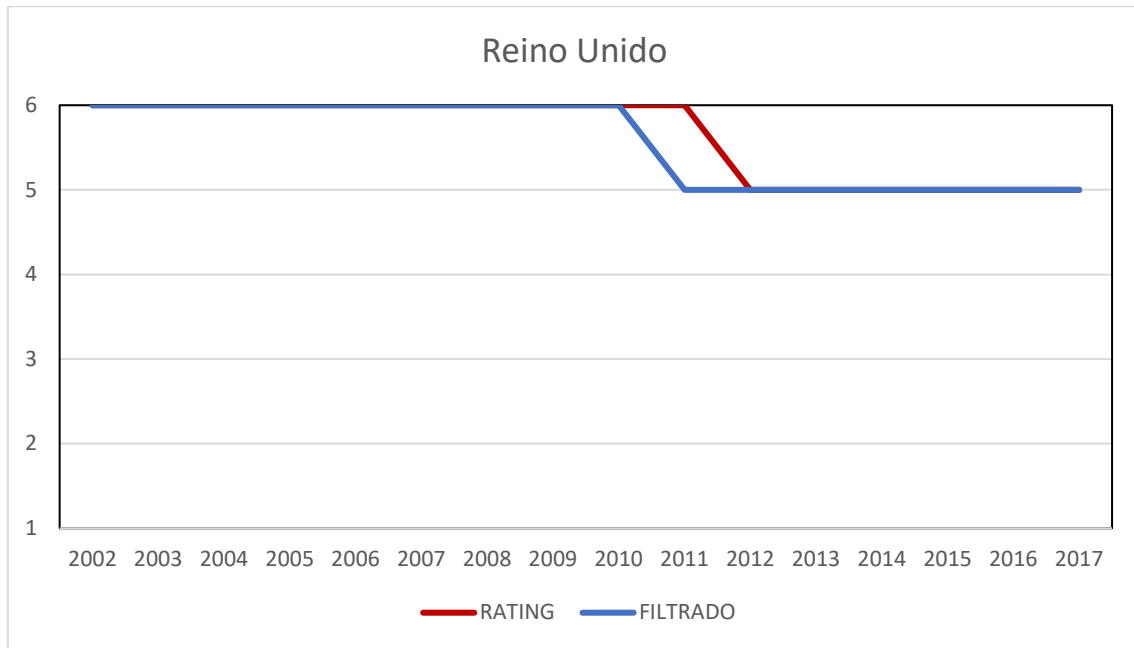
SVOREX, SVOORIMLIN y SVORIM asignan una categoría menos anticipándose por tanto 12 clasificadores a la decisión de la agencia.

- 2012, Random Forest, la Máquina de Vector Soporte, la Regresión Logística, ELMOP, KDLOR, NNOP, NNPO, OPBE, ORBOOST, SVOORIMLIN y SVR asignan la máxima categoría a pesar de que la agencia rebaja el rating relacionado con este ejercicio.
- 2013, HPOLD, OPBE, SVOREX, SVOORIMLIN y SVR asignan la máxima categoría a pesar de mantenerse la categoría Aa por la agencia.
- 2014, OPBE mantiene la máxima categoría Aaa.
- 2015, la Red Neuronal, la Regresión Logística y SVOORIMLIN mantienen la máxima categoría.
- 2016, la Regresión Logística, ELMOP, OPBE, SVOORIMLIN, mantienen la máxima categoría.
- 2017, la Red Neuronal, la Regresión Logística, HPOLD, OPBE, SVC1V1, SVOREX, SVOORIMLIN y SVORIM también mantienen la máxima categoría.

El ensamble de clasificadores es capaz de anticipar en un ejercicio la bajada de rating de la agencia.

CAPÍTULO 5: Experimentación

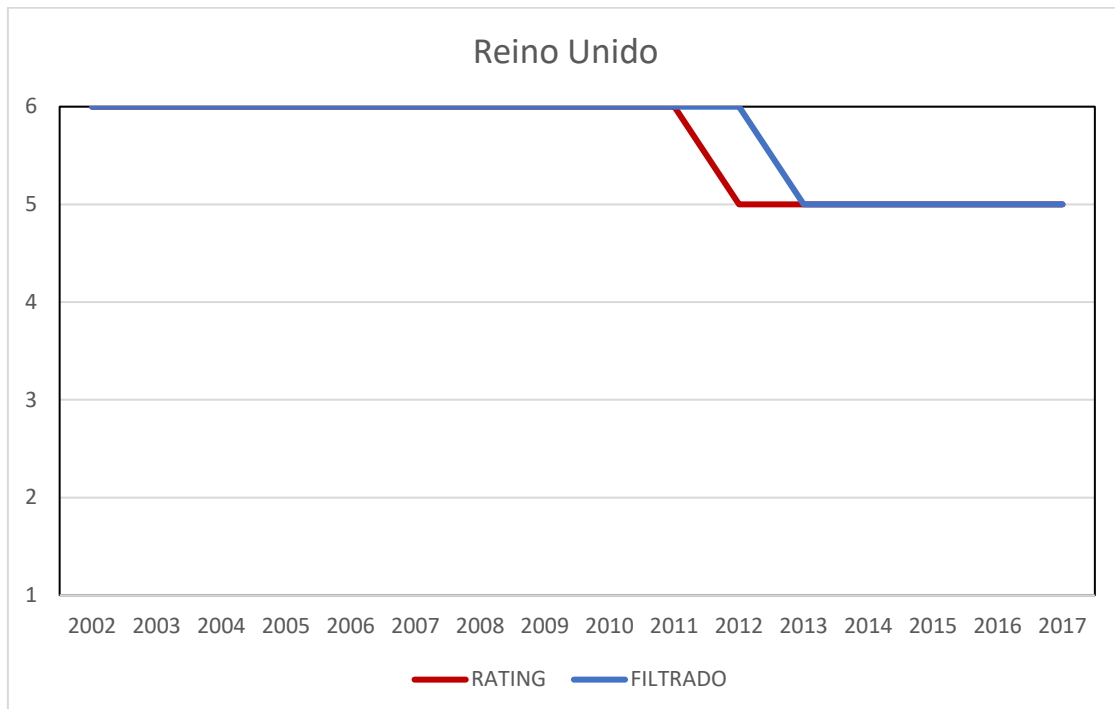
Gráfico 15: Resultado del ensamble de clasificadores para Reino Unido



Fuente: elaboración propia

En el caso de Random Forest todas las iteraciones coinciden y el modelo retrasa en un ejercicio la bajada de rating que realiza Moody's.

Gráfico 16: Resultados de Random Forest para Reino Unido



Fuente: elaboración propia

CAPÍTULO 5: Experimentación

Aunque el caso de Reino Unido tiene similitudes con el de Francia en este caso la anticipación de la bajada de rating no está tan generalizada y existen más excepciones a la clasificación de los últimos años en la categoría Aa.

5.2.6 Italia

El caso de Italia es el más extremo de los países que se han considerado como estables, mientras estos países han bajado como mucho una categoría y se mantienen todos en la categoría Aa parte de esta categoría y desciende a la categoría A en 2011 y a la categoría Baa en 2012, categoría que mantiene hasta el final del periodo estudiado.

Seguidamente se indican las diferencias entre el rating asignado y el estimado por los distintos clasificadores utilizados:

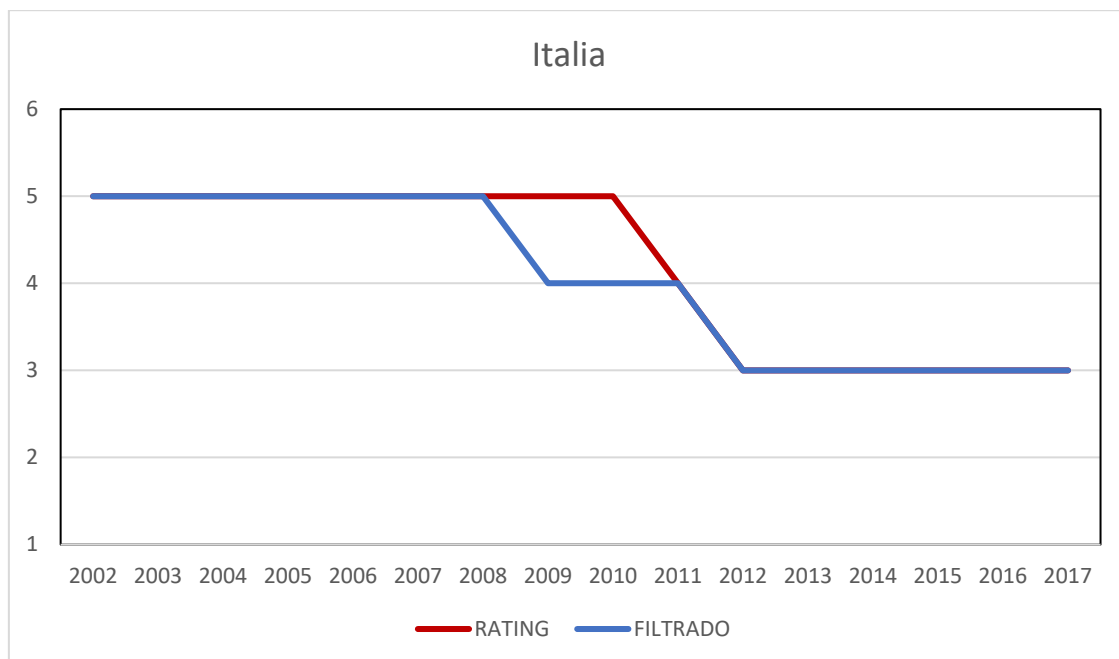
- 2007, OPBE asigna la máxima categoría en lugar de AA
- 2008, CSSVC, ELMOP y SVC1VA asignan dos categorías por debajo de la categoría asignada por la agencia. No existe ningún clasificador con diferencia de una categoría.
- 2009, la Red Neuronal, HPOLD, KDLOR, OPBE, REDSVM, SVOREX, SVOORIMLIN, SVORIM y SVR asignan una categoría por debajo de la que asigna Moody's.
- 2010, la Red Neuronal, HPOLD, KDLOR, NNPOM, OPBE, ORBOOST, REDSVM, SVOREX, SVORIMLIN, SVORIM y SVR asignan una categoría inferior. Esto implica que la mayoría de los clasificadores, 11 en este caso, anticipan en un año al menos la bajada que realiza la agencia en 2011.

CAPÍTULO 5: Experimentación

- 2011, Random Forest, la Regresión Logística y ELMOP asignan una categoría por encima al no estimar la bajada de rating que hace la agencia para ese ejercicio la Máquina de Vector Soporte, CSVVC, NNOP, SVC1V1, SVC1VA, SVORIMLIN asignan una categoría por debajo anticipando la bajada que realiza Moody's en el ejercicio siguiente.
- 2012, Random Forest, ELMOP, HPOLD, NNOP, NNPOM, ORBOOST, REDSVM, SVOREX y SVORIM mantienen la categoría A mientras la agencia rebaja la categoría a Baa.

El ensamble de todos estos clasificadores anticipa en dos ejercicios la primera bajada y a partir de 2011 coincide con los movimientos que realiza la agencia.

Gráfico 17: Resultados del ensamble de clasificadores para Italia



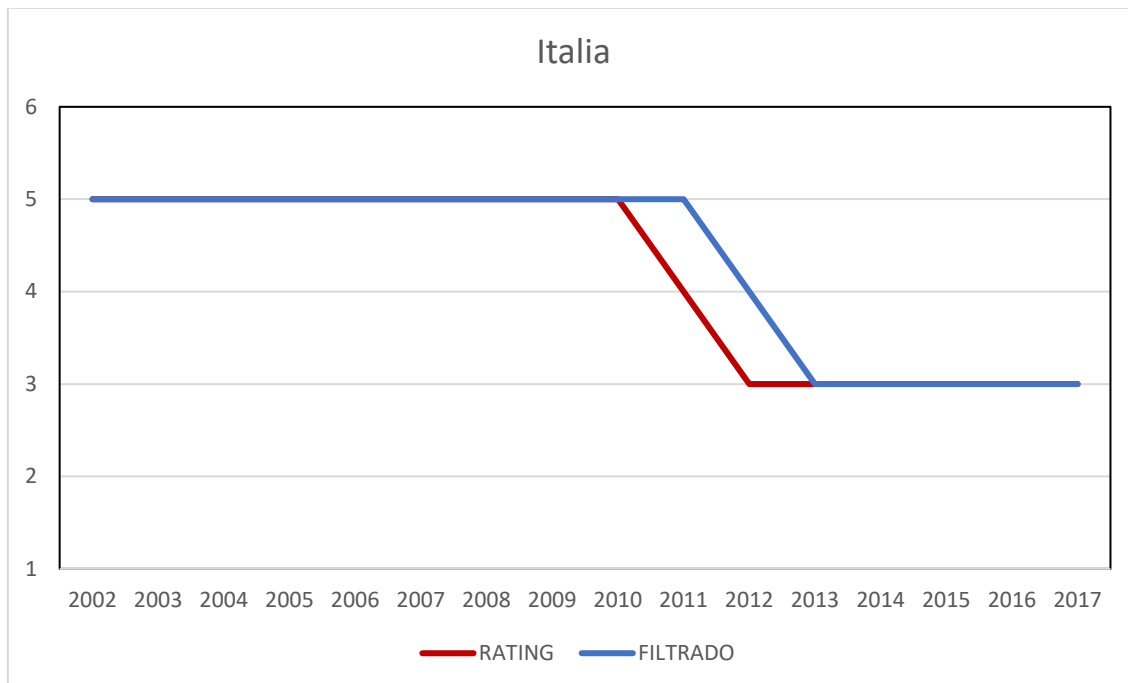
Fuente: elaboración propia

En el caso de Random Forest ninguna de las 30 repeticiones detecta la bajada de rating del ejercicio 2011 y en 4 de las 30 repeticiones se detecta la bajada a la categoría Baa en 2012 sin que exista otra diferencia en las distintas repeticiones. Random Forest se retrasa en un ejercicio en la estimación de las

CAPÍTULO 5: Experimentación

bajadas de rating de Italia siguiendo la misma senda, pero con un año de retraso.

Gráfico 18: Resultados de Random Forest para Italia



Fuente: elaboración propia

Se presenta en el caso de Italia una clara tendencia descendente que es perfectamente identificada por todos los clasificadores a partir de 2013 en el que asignan la misma categoría que Moody's. Existe discrepancia entre en cuanto al momento en que se inicia la bajada de categoría, aunque ya en 2010 la mayoría de los clasificadores, once de los diecinueve, anticipan la primera bajada de rating como se recoge en el gráfico del ensamble.

5.3 países inestables

Este es, sin duda, el conjunto de países más dispar. Dos países España e Irlanda, partían de la máxima categoría Aaa; en el caso de Irlanda llega a descender 4 categorías en tres ejercicios, aunque al final del periodo recupera 1

CAPÍTULO 5: Experimentación

y España desciende 3 en el mismo periodo. Portugal desciende 3 categorías en 2 ejercicios. Por último, Grecia llega a tener que reestructurar su deuda incurriendo en default en el ejercicio 2012.

5.3.1 España

Tras partir de una situación de elevada solvencia, con un bajo nivel de deuda que parte de un 51,3% del PIB en el ejercicio 2002 y va disminuyendo hasta alcanzar un mínimo del 35,6% del PIB en el ejercicio 2007, la crisis afecta significativamente a al equilibrio presupuestario partiendo de una situación de equilibrio presupuestario que llega a generar superávits presupuestarios en 2005, 2006 y 2007. A partir de 2008 se empiezan a generar déficits presupuestarios significativos que llegan a un máximo del 11% del PIB en el ejercicio 2008; estos déficits se mantienen por encima del 4% hasta el ejercicio 2016 con el consiguiente empeoramiento del nivel de endeudamiento que en 2010 supera el 60% del PIB y que sigue subiendo hasta alcanzar el 95% en 2013 y que se mantiene por encima de ese importe hasta el final del periodo.

Este comportamiento ha tenido su reflejo en los ratings asignados por Moody's que partiendo de la máxima categoría Aaa baja una categoría hasta Aa con la información relativa al ejercicio 2010 y seguidamente otra categoría en 2011 y otra en 2012 pasando a ser el rating A y Baa respectivamente.

Los clasificadores utilizados asignan la categoría correcta con las siguientes excepciones:

- 2009, la Máquina de Vector Soporte, CSSVC, ELMOP, KDLOR, NNPOM, ORBOOST, SVC1V1, SVC1VA y SVR asignan una categoría menos que

CAPÍTULO 5: Experimentación

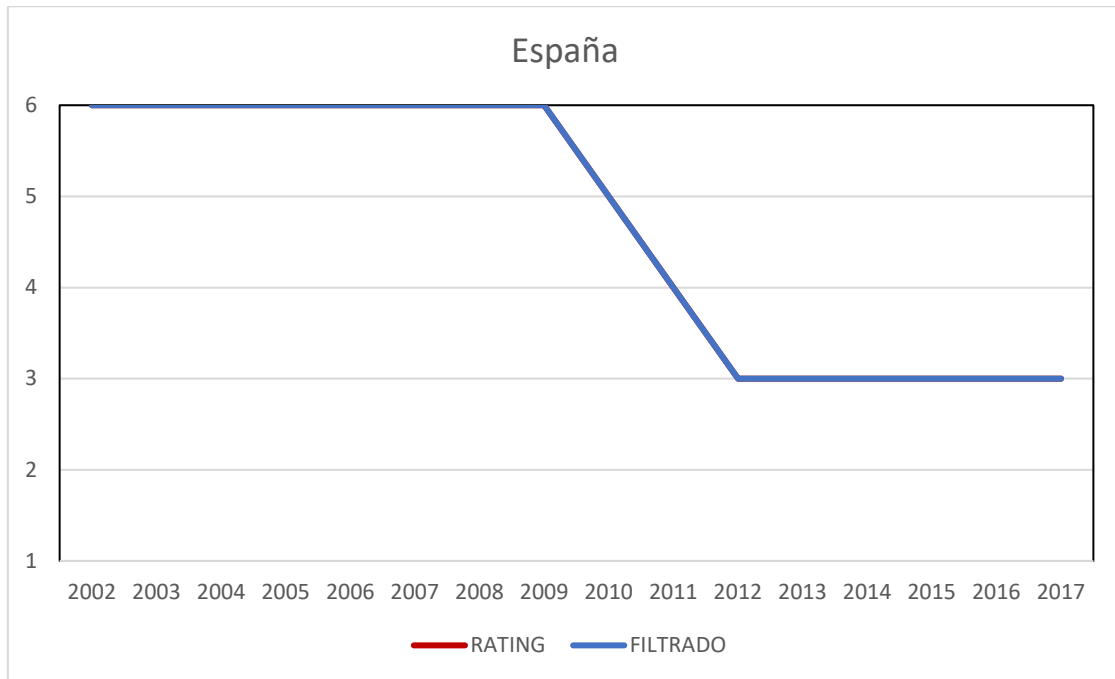
la agencia. Esto es, nueve de los 19 clasificadores anticipan en un ejercicio la bajada de rating que se produce en 2010.

- 2010, Random Forest, la Máquina de Vector Soporte, la Regresión Logística, CSSVC, ELMOP, NNOP, NNPOM, OPBE, ORBOOST, SVC1V1 y SVC1VA no detectan la bajada que practica la agencia por lo que mantienen la máxima categoría 11 clasificadores.
- 2011, Random Forest, la Red Neuronal, la Regresión Logística, HPOLD, NNOP, OPBE, ORBOOST, SVC1V1, SVOREX, SVOORIMLIM, SVORIM y SVR asignan una categoría superior al no recoger la bajada de rating que hace la agencia; la Máquina de Vector Soporte, CSSVC, ELMOP y SVC1VA asignan un rating inferior anticipando el movimiento de la agencia en 2012.
- 2012, la Red Neuronal, la Regresión Logística, NNOP, NNPOM, ORBOOST y SVOORIMLIM recogen la categoría Baa que asigna la agencia y el resto de los clasificadores (13) asignan una categoría superior.
- 2013, NNPOM asigna una categoría superior siendo el resto de las clasificaciones correctas.

Todos los clasificadores detectan la bajada de rating, discrepando en el momento temporal en el que la inician e incluso interrumpiendo la misma algún ejercicio. Todos estos efectos se compensan y tras el filtrado los ratings asignados por el ensamble de clasificadores para España coinciden con los de la agencia.

CAPÍTULO 5: Experimentación

Gráfico 19: Resultado del ensamble de clasificadores para España

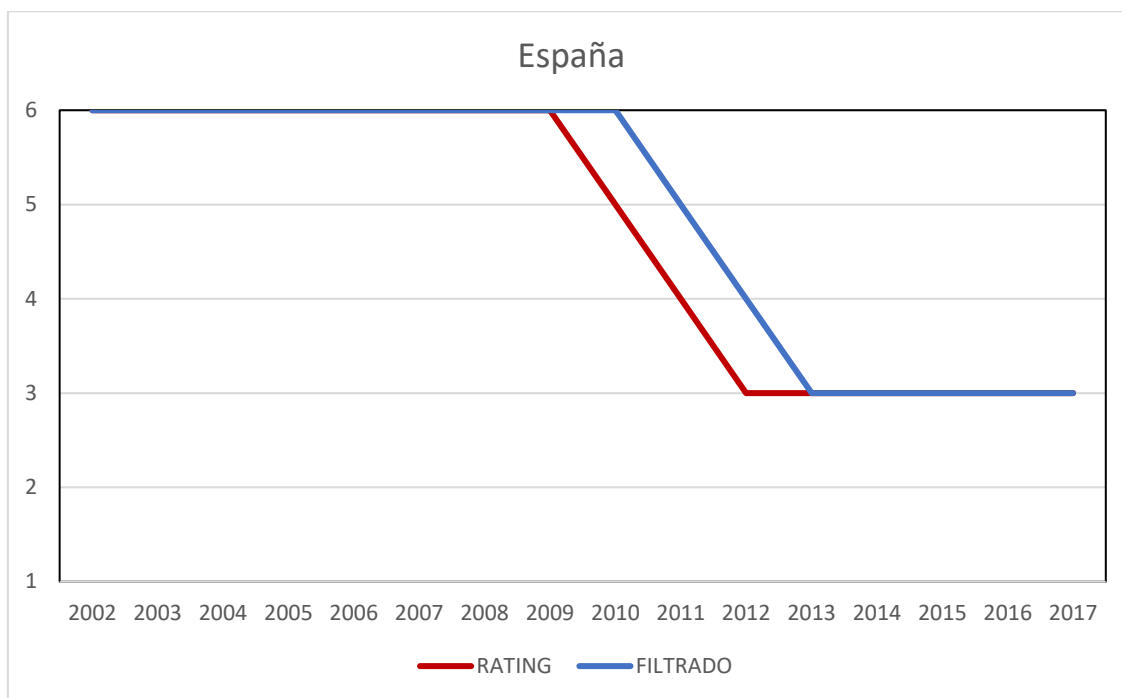


Fuente: elaboración propia

En el caso de Random Forest el modelo propuesto recoge las bajadas con un año de retraso respecto a las realizadas por la agencia. Todas las repeticiones coinciden salvo en el caso del ejercicio 2012 en el que en 9 de las 30 iteraciones el modelo recoge la bajada de categoría en el año correcto mientras en las 21 iteraciones restantes mantiene una categoría por encima.

CAPÍTULO 5: Experimentación

Gráfico 20: Predicción de Random Forest para España



Fuente: elaboración propia

Se presenta en este caso un país con una clara trayectoria descendente en su solvencia. Todos los clasificadores son capaces de detectar esta tendencia, aunque no coinciden en muchos casos con el momento temporal que aplica la agencia Moody's.

5.3.2 Irlanda

La trayectoria de la deuda de Irlanda en comparación con la de España tiene diferencias significativas:

- Mejor situación de partida, el déficit de Irlanda en 2007 era sólo de un 23,9% del PIB.
- Incremento más brusco de la deuda, en 2009 alcanza el 61,5% del PIB y en 2012 el 119,9% del PIB.

CAPÍTULO 5: Experimentación

- Clara disminución de la deuda al final del periodo, de 2013 a 2017 la deuda baja del 119,7% del PIB al 68,4% en 2017.

Este comportamiento se refleja en los ratings asignados por la agencia que pasan de la mayor categoría Aaa a principio del periodo a bajar una categoría en 2009, dos categorías en 2010, un más en 2011 y a partir de 2013 empieza una recuperación con una categoría en ese ejercicio y otra en 2016.

Las excepciones a la asignación de la categoría propuesta por la agencia son las siguientes:

- 2008, HPOLD, ORBOOST y SVR asignan una categoría por debajo de la asignada por la agencia y ELMOP y SVC1V1 asigna dos categorías por debajo. Estos clasificadores anticipan el comportamiento de la agencia en 2009.
- 2009, Random Forest, la Máquina de Vector Soporte, CSSVC y OPBE, asignan una categoría por encima; la Red Neuronal, NNOP y SVC1V1, asignan una categoría por debajo y la Regresión Logística, ELMOP y SVC1VA asignan dos categorías por debajo.
- 2010, SVC1V1 y SVR asignan la categoría Aaa, la Máquina de Vector Soporte, CSSVC, ELMOP, HPOLD, KDLOR, NNPOM y OPBE asignan la categoría Aa, la Regresión Logística, NNOP, ORBOOST, REDSVM y SVOREX y SVORIM, asignan la categoría A, la Red Neuronal y SVOORIMLIN son los dos únicos clasificadores que asignan la misma categoría que la agencia Baa, Random Forest asigna la categoría Ba-B y SVC1VA asigna la categoría \leq Caa. Nos encontramos sin duda ante el patrón más difícil de clasificar de toda la serie y el único que ha sido asignado a todas las categorías por el conjunto de clasificadores.

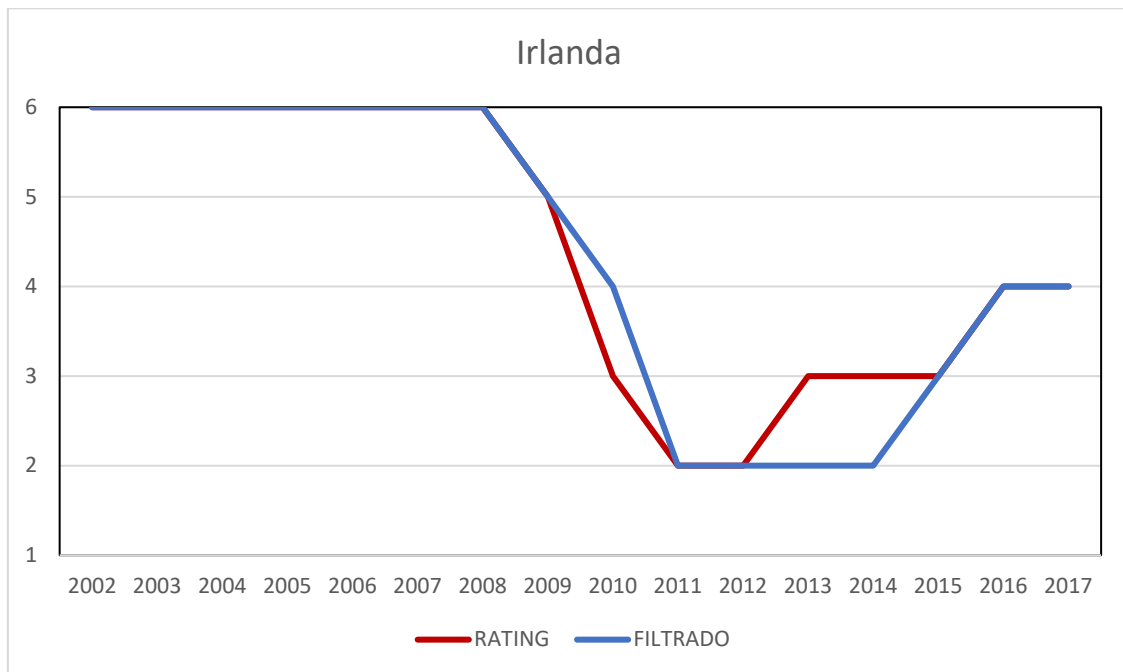
CAPÍTULO 5: Experimentación

- 2011, Random Forest, ORBOOST, SVC1V1 y SVR asignan una categoría por encima y la Regresión Logística, NNPOM, OPBE y SVORIM asignan una categoría por debajo.
- 2012, solamente la Red Neuronal, NNPOM, ORBOOST y SVORIMLIM asignan la categoría correcta Ba-B y el resto asigna una categoría por encima.
- 2013, la Máquina de Vector Soporte, la Regresión Logística, CSSVC, ELMOP y SVC1VA son los únicos que recogen la categoría que asigna Moody's tras subir el rating a Baa con los datos relativos a este ejercicio.
- 2014, la Regresión Logística, HPOLD, KDLOR, REDSVM, SVOREX, SVORIM y SVR asignan una categoría por debajo.
- 2015, REDSVM y SVOORIMLIM asignan dos categorías por encima y el resto una categoría por encima. Ningún clasificador asigna la categoría que asigna la agencia.
- 2016, Random Forest, la Máquina de Vector Soporte, la Red Neuronal, la Regresión Logística, ELMOP, NNPOM y SVC1V1 asignan una categoría por debajo y SVORIMLIN una categoría por encima.
- 2017, la Red Neuronal, NNOP, NNPOM, OPBE, ORBOOST, REDSVM, SVOREX y SVORIM asignan una categoría por encima y la Regresión Logística, ELMOP y SVORIMLIN asignan dos categorías por encima.

El ensamble de clasificadores recoge la misma tendencia que la agencia, pero con discrepancias temporales en la bajada y subida de categoría.

CAPÍTULO 5: Experimentación

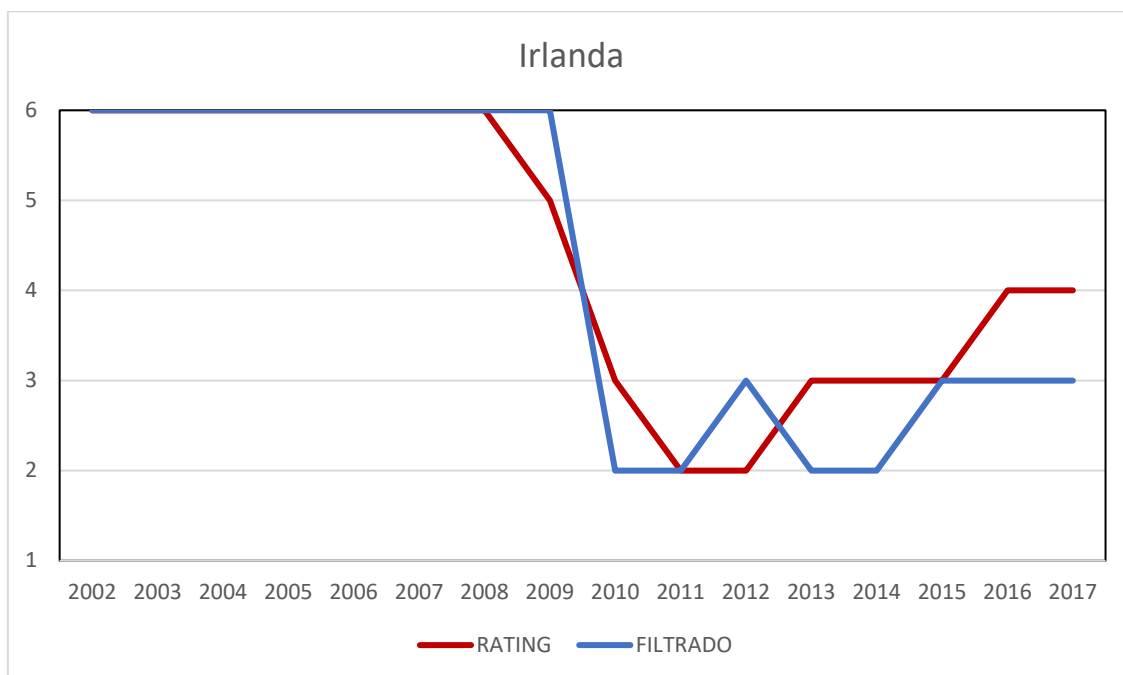
Gráfico 21: Resultado del ensamble de clasificadores para Irlanda



Fuente: elaboración propia

En el caso de Random Forest solo existen discrepancias en el año 2012 que, tras el filtrado, en 7 de las 30 repeticiones el algoritmo predice la clase correcta en vez de la clase superior que asignan las 23 repeticiones restantes.

Gráfico 22: Resultados de Random Forest para Irlanda



Fuente: elaboración propia

CAPÍTULO 5: Experimentación

Nos encontramos sin duda ante el caso más complejo y difícil de estimar, el rápido descenso de cuatro categorías y la posterior recuperación de dos dificultades de gran manera la tarea de los clasificadores utilizados llegando a pronosticar para el ejercicio 2010 la totalidad de las categorías en función de los clasificadores utilizados.

5.3.3 Portugal

Portugal presenta un comportamiento descendente muy fácil de entender para los clasificadores. Empieza en la categoría Aa y desciende dos categorías con los datos relativos a 2010 hasta la categoría Baa en relación al ejercicio 2011 descender una categoría más hasta Ba-B, categoría que mantiene hasta el final del periodo analizado.

Las excepciones a la correcta clasificación son pocas y se concentran en los años 2008 a 2011 coincidiendo con el impacto de la crisis en las cifras macroeconómicas utilizadas.

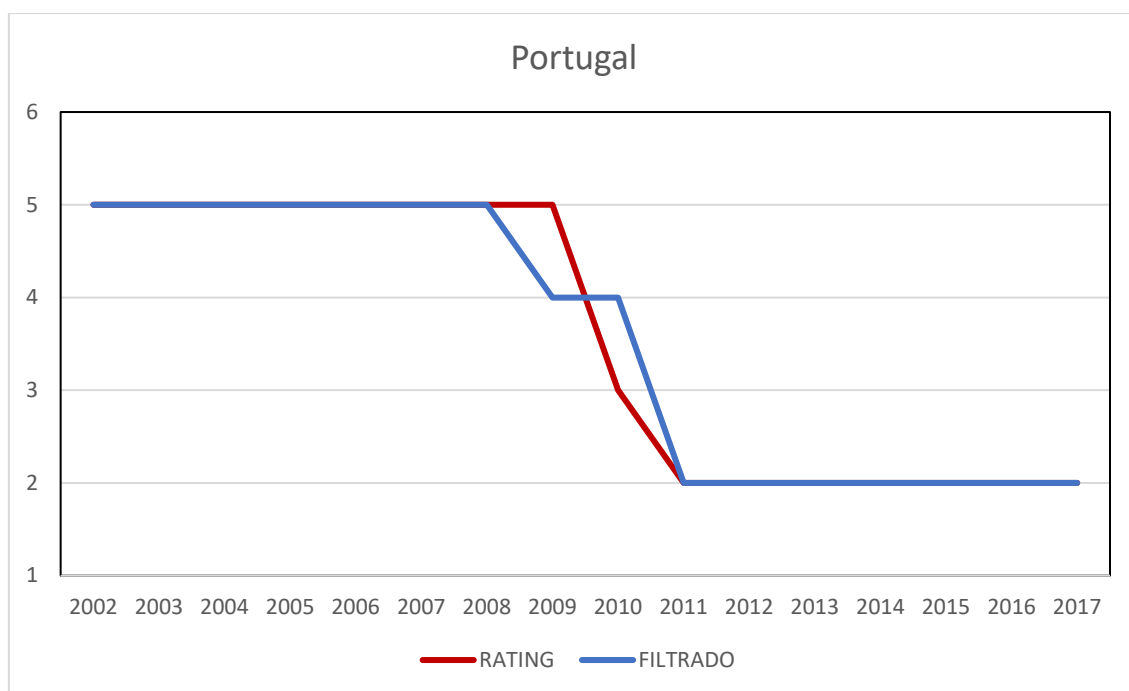
- 2008, NNOP asigna dos categorías por debajo anticipándose en dos ejercicios a la bajada que realiza Moody's.
- 2009, Random Forest, la Máquina de Vector Soporte, la Regresión Logística, CSSVC, HPOLD y SVC1VA asignan la misma categoría que la agencia y el resto una categoría menos.
- 2010, Random Forest, ORBOOST, REDSVM, SVOORIMLIM y SVORIM asignan la misma categoría que la agencia, la Máquina de Vector Soporte y SVC1V1 asignan dos categorías por encima y el resto una.

CAPÍTULO 5: Experimentación

- 2011, NNPOM y ORBOOST asignan una categoría por encima y REDSVM, SVOREX y SVORIM una categoría por debajo.

El ensamble de clasificadores recoge la tendencia de bajada, coincidiendo en los años iniciales y finales y discrepando únicamente en los momentos en los que se produce la bajada.

Gráfico 23: Resultados del ensamble de clasificadores para Portugal

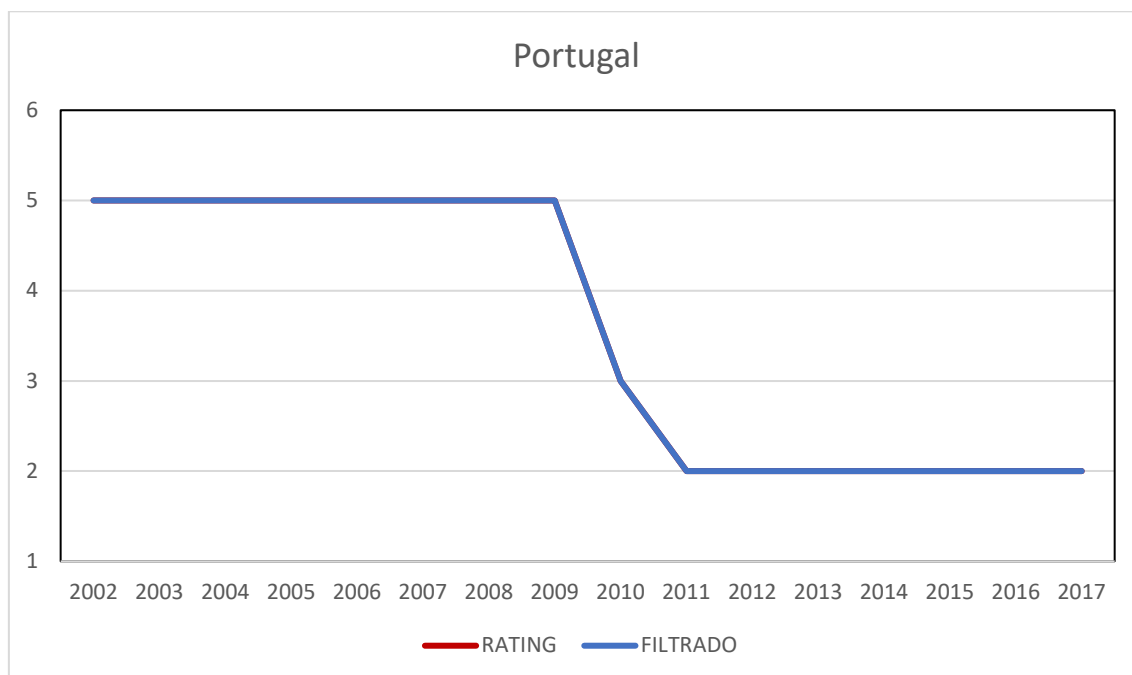


Fuente: elaboración propia

En el caso de Random Forest existen diferencias en las predicciones de las distintas repeticiones en los años 2010 y 2011. En 2010 el 60% de los clasificadores asigna la categoría Ba-B y el 40% A, esto es, dos categorías por encima; al aplicar la media a las predicciones el modelo da como salida la categoría intermedia Baa que es la misma que asigna la agencia de rating. En 2011 el 80% de las iteraciones asigna la categoría Ba-B que es la correcta y un 20% asigna una categoría superior.

CAPÍTULO 5: Experimentación

Gráfico 24: Resultados de Random Forest para Portugal



Fuente: elaboración propia

En el caso de Portugal existe un claro empeoramiento de su situación financiera que lleva a una bajada de rating que Random Forest es capaz de predecir correctamente y el ensamble de clasificadores predice con ligeras diferencias al anticiparse un ejercicio en la primera bajada, aunque con menor intensidad al bajar una categoría en vez de las dos que baja la agencia y retrasándose un ejercicio en la segunda bajada, aunque esta vez con mayor intensidad, para coincidir en 2011 con las predicciones de la agencia.

5.3.4 Grecia

En el caso de Grecia la crisis afecta significativamente a su solvencia, bajando en dos ejercicios a la peor categoría, ratings inferiores a la categoría Caa de Moody's, para recuperar una categoría al final del periodo.

Las discrepancias en las predicciones son las siguientes:

CAPÍTULO 5: Experimentación

- 2003, SVOORIMLIN asigna una categoría por encima.
- 2006, SVOORIMLIN asigna una categoría por encima
- 2008, la Máquina de Vector Soporte asigna dos categorías por debajo, SVORIMLIN asigna una categoría por debajo, la Regresión Logística, CSSVC, NNOP, SVC1V1 y SVC1VA asignan una categoría por encima y ELMOP dos categorías por encima
- 2009, solamente SVR asigna la categoría correcta, ELMOP asigna dos categorías por encima, SVC1V1 asigna dos categorías por debajo y el resto asigna una categoría por debajo.
- 2010, Random Forest asigna dos categorías por encima y ORBOOST asigna una categoría por encima.
- 2011, KDLOR, NNPO, ORBOOST, SVOORIMLIN y SVR asignan una categoría por encima.
- 2013, la Regresión Logística, CSSVC, ELMOP, REDSVM, SVC1VA, asignan una categoría por encima.
- 2015, KDLOR asigna una categoría por encima y SVR dos categorías por encima.
- 2016, Random Forest, la Red Neuronal, ORBOOST, REDSVM, asignan una categoría por encima, la Regresión Logística, CSSVC, KDLOR y NNPO, asignan dos categorías por encima, SVR asigna tres categorías por encima y ELMOP y SVOREX y SVORIM asignan cuatro categorías por encima
- 2017, la Máquina de Vector Soporte, CSSVC, HPOLD, NNOP, OPBE, SVC1V1, SVC1VA y SVOORIMLIN asignan una categoría por debajo, la Regresión Logística, NNPO y REDSVM, asignan una categoría por

CAPÍTULO 5: Experimentación

encima KDLOR y SVR asignan dos categorías por encima y ELMOP, SVOREX y SVORIM, asignan tres categorías por encima.

En este caso en dos ejercicios y 2016 y 2017 los clasificadores asignan 5 de las seis categorías posibles. Estas discrepancias en la clasificación ponen de manifiesto la dificultad de los clasificadores para entrenar casos muy particulares. Al realizar la media de todos los clasificadores mediante el ensamble de clasificadores estas diferencias se diluyen proporcionando unas predicciones muy próximas a las de la agencia y con el anticipo de un ejercicio en la bajada de rating, que se hace de forma más gradual al bajar el ensamble de clasificadores tres categorías en tres ejercicios mientras la agencia baja dos categorías un año más tarde para coincidir en la siguiente bajada con los clasificadores analizados.

Gráfico 25: Resultados del ensamble de clasificadores para Grecia

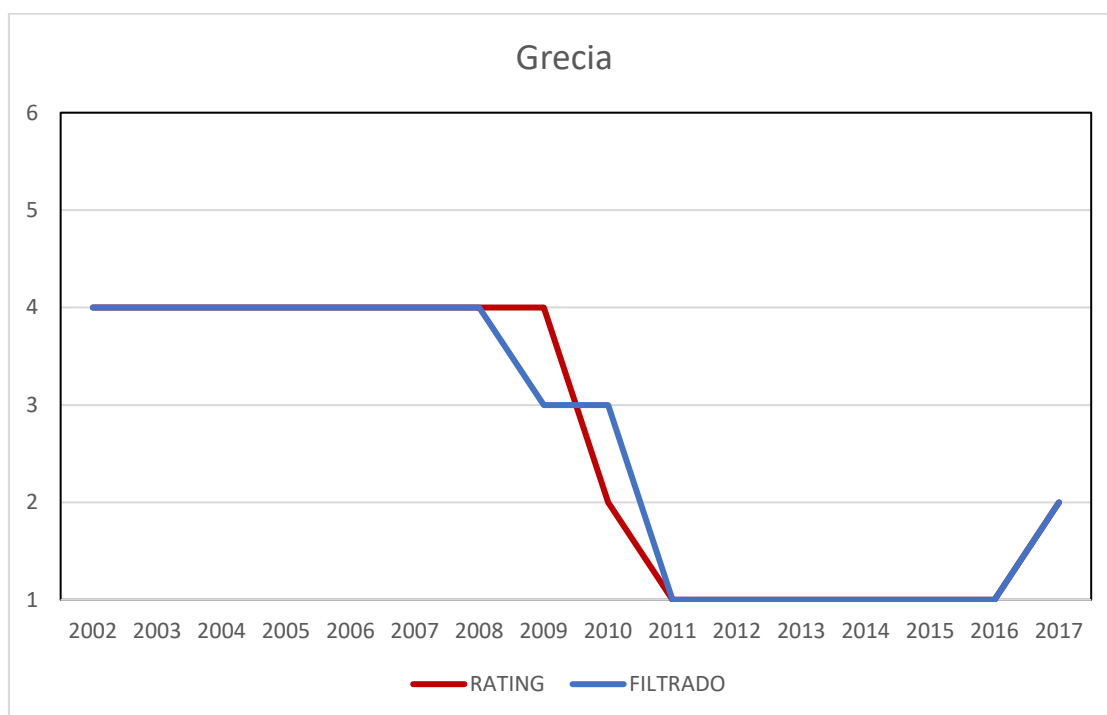


Fuente: elaboración propia

CAPÍTULO 5: Experimentación

Random Forest presenta unos resultados muy similares y bastante estables. Solo en 2010 una de las iteraciones propone dos categorías menos que el resto y en 2017 en cuatro de las treinta repeticiones propone mantener la categoría inferior en vez de subir una categoría como hace el resto de las iteraciones y la agencia Moody's.

Gráfico 26: Resultados de Random Forest para Grecia



Fuente: elaboración propia

En este caso la bajada que realiza la agencia tiene una pendiente distinta a la de Random Forest, anticipándose en un ejercicio la primera bajada, que es de una categoría en vez de dos y retrasándose en un ejercicio la segunda, que es de dos categorías en vez de una.

Se puede comprobar que, aunque con discrepancias puntuales, el caso de Grecia es bastante claro, parte de una situación regular para después de la crisis descender a la peor categoría de la que no saldrá hasta el final del periodo.

6. RESULTADOS POR PAÍSES Y POR AÑOS

Como resumen del análisis realizado por países se puede analizar cómo se comportan los clasificadores por países y por año. La primera tabla presenta la frecuencia de los valores absolutos en cada uno de los países. Del total de 304 clasificaciones por país (16 años por 19 clasificadores) se ordenan por columnas en función de los distintos valores que toman los valores absolutos, por último, se indica el valor total de la suma de los errores absolutos (equivale a multiplicar la frecuencia de cada error por su valor).

Tabla 45: Errores de clasificación por países de la serie original

	Frecuencia de Error Absoluto					Suma Error
	0	1	2	3	4	
Austria	222	69	12	1	0	96
Bélgica	288	16	0	0	0	16
Dinamarca	304	0	0	0	0	0
Finlandia	274	30	0	0	0	30
Francia	262	42	0	0	0	42
Alemania	302	2	0	0	0	2
Grecia	234	51	12	4	3	99
Irlanda	190	94	18	2	0	136
Italia	262	39	3	0	0	45
Luxemburgo	304	0	0	0	0	0
Holanda	304	0	0	0	0	0
Portugal	269	32	3	0	0	38
España	254	50	0	0	0	50
Suecia	304	0	0	0	0	0
Reino Unido	260	44	0	0	0	44
Total	4033	469	48	7	3	598

Fuente: elaboración propia

Como se puede observar la mayor parte de los errores corresponden a una desviación de una categoría respecto al asignado por la agencia siendo más escasos los que corresponden a dos, tres y cuatro categorías. Del total de la suma de los errores la mayor parte corresponde a Irlanda, Grecia y Austria que

CAPÍTULO 5: Experimentación

son los países en los que los algoritmos propuestos han tenido más dificultad de clasificación.

Si se comprueban los mismos datos sobre la serie filtrada se obtiene la siguiente tabla.

Tabla 46: Errores de clasificación por Países sobre la serie filtrada

	Frecuencia de Error Absoluto				Suma Error
	0	1	2	3	
Austria	219	57	26	2	115
Bélgica	287	17	0	0	17
Dinamarca	304	0	0	0	0
Finlandia	284	20	0	0	20
Francia	269	35	0	0	35
Alemania	300	4	0	0	4
Grecia	261	35	4	4	55
Irlanda	208	84	12	0	108
Italia	260	38	6	0	50
Luxemburgo	304	0	0	0	0
Holanda	304	0	0	0	0
Portugal	267	34	3	0	40
España	267	37	0	0	37
Suecia	304	0	0	0	0
Reino Unido	278	26	0	0	26
Total	4116	387	51	6	507

Fuente: elaboración propia

En este caso se observa claramente como el filtrado mejora sensiblemente la clasificación. El total de la suma de errores baja de 598 a 507 y, salvo en el caso de Austria donde el importe total de la suma de errores absolutos aumenta de forma apreciable en la mayor parte de los casos se descende, se mantiene o aumenta débilmente. Destaca la bajada de Grecia en la que pasa de una suma de errores de 99 a 55, seguida de Irlanda que pasa de 136 a 108 y de Reino Unido que pasa de 44 a 26.

El mismo análisis realizado por año nos muestra en que años se concentran los errores.

CAPÍTULO 5: Experimentación

Tabla 47: Frecuencia de errores por año

	Frecuencia de error absoluto					Total Error
	0	1	2	3	4	
2002	278	7	0	0	0	7
2003	284	1	0	0	0	1
2004	285	0	0	0	0	0
2005	285	0	0	0	0	0
2006	284	1	0	0	0	1
2007	281	4	0	0	0	4
2008	262	15	8	0	0	31
2009	217	63	5	0	0	73
2010	215	57	11	2	0	85
2011	194	78	12	1	0	105
2012	229	56	0	0	0	56
2013	260	25	0	0	0	25
2014	270	15	0	0	0	15
2015	245	37	3	0	0	43
2016	233	44	4	1	3	67
2017	211	66	5	3	0	85
Total	4033	469	48	7	3	598

Fuente: elaboración propia

Se puede observar que los errores aumentan significativamente a partir de la crisis de 2008, aumentan significativamente hasta alcanzar un máximo en 2011 para ir disminuyendo hasta 2014 para luego aumentar con las subidas de categorías de algunos países.

La serie transformada presenta un comportamiento similar, pero con mejoras significativas en la cuantía de los errores.

CAPÍTULO 5: Experimentación

Tabla 48: Errores absolutos por año de la serie transformada

	Frecuencia Error Absoluto				Suma Error
	0	1	2	3	
2002	278	7	0	0	7
2003	278	7	0	0	7
2004	285	0	0	0	0
2005	285	0	0	0	0
2006	285	0	0	0	0
2007	285	0	0	0	0
2008	273	5	7	0	19
2009	214	61	10	0	81
2010	223	54	8	0	70
2011	191	80	13	1	109
2012	232	39	13	1	68
2013	269	16	0	0	16
2014	262	23	0	0	23
2015	261	24	0	0	24
2016	263	22	0	0	22
2017	232	49	0	4	61
Total	4116	387	51	6	507

Fuente: elaboración propia

En esta ocasión hay un empeoramiento significativo en el ejercicio 2012 y mejora en los ejercicios 2016, 2017 y 2015 años en los que se producen algunas subidas de rating por parte de los clasificadores que no son asumidas por la Agencia.

7. TIPIFICACIÓN DE LOS ERRORES

Por último, se propone tipificar los errores cometidos por el ensamble de clasificadores y Random Forest para entender cómo se han comportado las variables seleccionadas en la determinación de los ratings.

Para ello se han dividido los errores en tres tipos:

- ANTICIPO, la clasificación se anticipa en uno o dos ejercicios a los movimientos de la agencia. Esta anticipación puede ser parcial o total,

CAPÍTULO 5: Experimentación

esto es, si la agencia baja dos ejercicios en un año y la clasificación propuesta baja un solo ejercicio el año anterior se considera anticipo.

- DEMORA, la clasificación se demora en uno o dos ejercicios a los movimientos que realiza la agencia. Igualmente puede ser total o parcial.
- ERROR, se produce un error cuando no hay un movimiento posterior de la Agencia en el mismo sentido o este se demora más de dos ejercicios en el tiempo.

Tabla 49: Frecuencia de errores por tipo del ensamble de clasificadores

	Austria	Francia	Grecia	Irlanda	Italia	Portugal	Reino Unido
ANTICIPO	1	2	1		2	1	1
DEMORA				3		1	
ERROR	3						

Fuente: elaboración propia

De los 15 errores que se producen en la clasificación 8 se deben a que el algoritmo ha clasificado un movimiento con anterioridad a lo que lo ha hecho la agencia, en 4 ocasiones hay un retraso y en 3 ocasiones es un claro error. Estos errores corresponden en su totalidad a Austria que por una parte en el año 2011 es clasificada como A en vez de Aaa y por el filtrado se traslada a 2012 y que en 2017 es clasificada por el ensamble como Aaa mientras la agencia mantiene la clase Aa.

En cuanto al valor absoluto de los errores, exceptuando los casos comentados de Austria en 2011 y 2012 todos los demás errores suponen una diferencia de una sola categoría por lo que la suma de los errores absolutos para el ensamble de clasificadores asciende a 17.

Si analizamos los resultados de Random Forest obtenemos la siguiente tabla:

CAPÍTULO 5: Experimentación

Tabla 50: Frecuencia de errores por tipo en Random Forest

	Austria	Francia	Grecia	Irlanda	Italia	España	Reino Unido
ANTICIPO		1	1	2			
DEMORA			1	2	2	3	1
ERROR	2			3			

Fuente: elaboración propia

En este caso la mayor parte de los errores corresponden a demoras en reconocer la clasificación de la agencia. Todos los errores corresponden a una diferencia de una sola clase por lo que la suma de errores, 18, es igual a la frecuencia de estos.

CAPITULO 6

CONCLUSIONES

1. CONCLUSIONES DE LA INVESTIGACIÓN

La calificación de deuda soberana mediante ratings es de gran importancia para los gestores públicos y para los inversores. La mejor o peor clasificación puede tener incidencia en los mercados financieros y por tanto influir en el coste de la deuda de un país con los evidentes efectos presupuestarios.

Como se ha comprobado en la revisión de la literatura a la hora de abordar este problema existe una gran diferencia entre las variables a seleccionar para hacer estimaciones de ratings de deuda soberana. El plantear una selección de variables basadas en un modelo de dinámica de la deuda es un planteamiento novedoso en la literatura y ha proporcionado un conjunto de variables que permiten unas predicciones que mejoran a las de otros autores.

Para realizar las predicciones se han utilizado distintas metodologías, tanto pertenecientes a la estadística clásica como a técnicas más modernas que se engloban en el denominado “aprendizaje automático”.

Los ratings son etiquetas nominales que solo proporcionan una información ordinal, si se es más o menos solvente pero no cuanto mejor o peor se es, por lo que se ha optado por incluir clasificadores ordinales además de los exclusivamente nominales.

Para evitar los efectos del sobreentrenamiento se han diseñado los experimentos separando los datos en conjuntos de entrenamiento y de generalización, ya sea con técnicas de hold-out o k-fold.

Por último, se ha aplicado la técnica del filtrado de datos a las clasificaciones obtenidas consiguiéndose mejoras en todos los casos a pesar de partir de unos resultados de clasificación muy elevados.

CAPÍTULO 6: Conclusiones

Podemos concluir que se han logrado los objetivos de investigación planteados, el objetivo número 1 se ha conseguido al proponerse un modelo con el que se consiguen unos resultados que superan a los conseguidos por otros autores. Se han realizado solo para el 10 fold un total de 4.560 estimaciones de rating (15 países x 16 años x 19 clasificadores) con unos resultados muy positivos. Los datos correctamente clasificados están comprendidos entre un 85,8% y un 91,7%, los datos clasificados en el entorno de una categoría fluctúan entre un 95,8% y un 100% el error absoluto medio está comprendido entre el 0,088 y el 0,196 y el promediado por clases entre el 0,156 y el 0,451 lo que supone unos resultados superiores a los obtenidos por otros autores en todas las métricas utilizadas. Todo ello en durante un periodo largo de tiempo que abarca los años previos a la crisis, la crisis financiera y la posterior recuperación.

El objetivo específico 1.1 se ha logrado al definir un conjunto de variables basadas en las teorías sobre la dinámica de la deuda y la solvencia que han demostrado un mejor desempeño que otros conjuntos de variables con una menor justificación macroeconómica. Se ha definido un conjunto limitado de variables inferior a la media usada por otros autores. De estas 7 variables, 4 de ellas se han obtenido sin ningún tipo de transformación y 3 son el resultado de un cálculo simple de datos fácilmente disponibles. Estas variables han permitido una estimación de los ratings con resultados significativos.

El objetivo específico 1.2 también se ha conseguido pues se ha planteado un conjunto de 19 clasificadores de distinta naturaleza consiguiéndose muy buenos resultados de clasificación en todos los casos tanto individualmente, como en el ensamble de todos ellos. Esto demuestra que el modelo obtenido es totalmente independiente del clasificador que se utilice pues funciona con buenos

CAPÍTULO 6: Conclusiones

resultados con clasificadores de naturaleza muy distintas. Además la mayoría de los errores detectados se corresponden con una anticipación del modelo propuesto sobre las acciones de las agencias de rating lo que hace que estos errores generen información positiva más que confusión.

También se ha alcanzado el objetivo número 2 al haber detectado que la estimación mediante variables tiene una clara tendencia de anticipación sobre las acciones de las agencias de rating y que esta anticipación es mayor en el caso de las subidas que en de las bajadas por lo que aplicando un retraso a la subida se mejora la clasificación. Podemos concluir que el filtrado de datos es una técnica que mejora la clasificación original obtenida. Esto ha ocurrido con todos los clasificadores por lo que esta técnica es efectiva e independiente del clasificador utilizado.

Las mejoras en datos correctamente clasificados oscilan entre el 0,9% y 2,9% y se producen en todos los clasificadores. No se producen mejoras en los resultados del entorno de una clase pues el filtrado generalmente tiene efecto de una clase y por tanto no aporta nada a esta métrica. Sí son más significativas las mejoras en el error absoluto medio, entre un 3,6% y un 31,9%, y en el error absoluto medio ponderado, entre un 7,1% y un 51,7%. Estas mejoras suponen una mejora de la clasificación en los 19 clasificadores utilizados por lo que también se puede concluir que el filtrado es una técnica que mejora la clasificación con independencia del clasificador utilizado.

2 CONTRIBUCIÓN DE LA TESIS

El estimar de forma objetiva el rating de la deuda soberana con un conjunto limitado de variables permite tener un conocimiento de la solvencia de un país totalmente independiente de los sesgos que una agencia de rating pueda generar.

El haber determinado las variables a seleccionar partiendo de un modelo macroeconómico de dinámica de la deuda es un planteamiento novedoso que plantea un limitado conjunto de indicadores que desde un punto de vista teórico deberían de funcionar en cualquier situación.

Estas variables son solo 7 son fáciles de obtener y están disponibles para la mayoría de los países por lo que el modelo es fácilmente replicable en otras situaciones.

3. LINEAS FUTURAS DE INVESTIGACIÓN

Es evidente que en esta tesis sólo se ha demostrado que el modelo propuesto funciona adecuadamente en el conjunto de países seleccionados y durante el periodo analizado.

Una primera línea de investigación consistiría en analizar si el modelo funciona también en países con características similares al conjunto analizado en esta tesis. Esto permitiría contrastar que el modelo es adecuado para un tipo concreto de país.

CAPÍTULO 6: Conclusiones

En segundo lugar, habría que analizar cómo se comporta el modelo en países con características distintas al analizado y valorar si es preciso de incluir alguna variable adicional a las planteadas para conseguir predicciones acertadas.

BIBLIOGRAFIA

Bibliografía

- Acharya, V., Davydenko, S. A., & Strebulaev, I. A. (2012). Cash holdings and credit risk. *Review of Financial Studies*, 25(12).
<https://doi.org/10.1093/rfs/hhs106>
- Afonso, A. (2003). Understanding the Determinants of Sovereign Debt Ratings: Evidence for the Two Leading Agencies. *Journal of Economics & Finance*, 27(1), 56.
- Afonso, A., & Gomes, P. (2011). Do Fiscal Imbalances Deteriorate Sovereign Debt Ratings? *Revue Économique*, 6, 1123.
<http://recursos.uloysola.es/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsrep&AN=edsrep.a.cai.recosp.reco.626.1123&lang=es&site=eds-live>
- Afonso, A., Gomes, P., & Rother, P. (2007). What “hides” behind Sovereign Debt Ratings? *European Central Bank Working Paper Series*, 711.
<http://10.0.3.234/ijfe.416>
- Afonso, A., Gomes, P., & Rother, P. (2009). Ordered response models for sovereign debt ratings. *Applied Economics Letters*, 16(8), 769–773.
<https://doi.org/10.1080/13504850701221931>
- Afonso, A., Gomes, P., & Rother, P. (2011). Short and long-run determinants of sovereign debt credit ratings. *International Journal of Finance & Economics*, 16(1), 1–15. <http://10.0.3.234/ijfe.416>
- Agénor, P.-R., & Montiel, P. (2008). *Development macroeconomics / Pierre-Richard Agénor and Peter J. Montiel*. Princeton : Princeton University Press, c2008; 3rd ed.

Bibliografía

- Almeida, H., Cunha, I., Ferreira, M. A., & Restrepo, F. (2017). The Real Effects of Credit Ratings: The Sovereign Ceiling Channel. *Journal of Finance*, 72(1). <https://doi.org/10.1111/jofi.12434>
- Al-Sakka, R., & ap Gwilym, O. (2009). Heterogeneity of sovereign rating migrations in emerging countries. *Emerging Markets Review*, 10(2). <https://doi.org/10.1016/j.ememar.2008.12.001>
- Altman, E. I., & Rijken, H. A. (2005). The Impact of the Rating Agencies' Through-the-cycle Methodology on Rating Dynamics. *Economic Notes*, 34(2), 127-127–154. <https://doi.org/10.1111/j.0391-5026.2005.00147.x>
- Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Evaluation Measures for Ordinal Regression. *Ninth International Conference on Intelligent Systems Design and Applications*. <https://doi.org/10.1109/ISDA.2009.230>
- Banco Central Europeo. (2008). *Technical specifications for the temporary expansion of the collateral framework*.
- Bank For International Settlements. (2011). The impact of sovereign credit risk on bank funding conditions. In *CGFS Papers* (Issue 43).
- Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework Comprehensive Version*.
- Basu, K., De, S., Ratha, D., & Timmer, H. (2013). Sovereign ratings in the post-crisis world : an analysis of actual, shadow and relative risk ratings. *Policy Research Working Paper Series*.

Bibliografía

- Beers, D., Cavanaugh, M., & Takahira, O. (2008). Sovereign Credit Ratings: A Primer. *Standard & Poor's Ratings Direct*, 44.
- Bennell, J. A., Crabbe, D., Thomas, S., & Gwilym, O. A. (2006). Modelling sovereign credit ratings: Neural networks versus ordered probit. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2005.10.002>
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., & Wiswedel, B. (2009). KNIME-the Konstanz information miner: version 2.0 and beyond. *AcM SIGKDD Explorations Newsletter*, 11(1), 26–31.
- Binici, M., Hutchison, M., & Miao, E. W. (2018). Are credit rating agencies discredited? Measuring market price effects from agency sovereign debt announcements Are Credit Rating Agencies Discredited? Measuring Market Price Effects from Agency Sovereign Debt Announcements *. *BIS Working Papers*, 704.
- Bissoondoyal-Bheenick, E. (2005). An analysis of the determinants of sovereign ratings. *Global Finance Journal*, 15, 251–280. <https://doi.org/10.1016/j.gfj.2004.03.004>
- Bissoondoyal-Bheenick, E., Brooks, R., & Yip, A. Y. N. (2006). Determinants of sovereign ratings: A comparison of case-based reasoning and ordered probit approaches. *Global Finance Journal*, 17, 136–154. <https://doi.org/10.1016/j.gfj.2006.06.001>
- Blanchard, O. (1990). Suggestions for a new set of fiscal indicators. In *OECD Economics Department Working Papers*, No. 79. OECD Publishing. <https://doi.org/10.1787/435618162862>

Bibliografía

- Blanchard, O., Chouraqui, J.-C., Hagemann, R. P., & Sartor, N. (1990). The Sustainability Fiscal Policy : New Answers To An Old Questions. In *OECD Economic Studies* (Vol. 15, pp. 1–36).
- Borio, C., & Packer, F. (2004). Assessing new perspectives on country risk. *BIS Quarterly Review*.
- Boumparis, P., Milas, C., & Panagiotidis, T. (2015). Has the Crisis Affected the Behavior of the Rating Agencies? Panel Evidence from the Eurozone. *Economics Letters*, 136, 118–124.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.
- Breiman, L. (2001). BRandom Forests. *Machine Learning*, 45.
- Briguglio, L., Cordina, G., Farrugia, N., & Vella, S. (2009). Economic vulnerability and resilience: Concepts and measurements. *Oxford Development Studies*, 37(3). <https://doi.org/10.1080/13600810903089893>
- Broto, C., & Molina, L. (2016). Sovereign ratings and their asymmetric response to fundamentals. *Journal of Economic Behavior and Organization*, 130, 206–224. <https://doi.org/10.1016/j.jebo.2016.07.017>
- Butler, A. W., & Fauver, L. (2006). Institutional environment and sovereign credit ratings. *Financial Management*, 35(3). <https://doi.org/10.1111/j.1755-053X.2006.tb00147.x>
- Cantor, R. M., & Mann, C. (2003). Measuring the performance of Corporate Bond Ratings. In *Moody's Investors Service, Global Credit Research* (Issue April).

Bibliografía

- Cantor, R., & Mann, C. (2007). Analyzing the Tradeoff Between Ratings Accuracy and Stability. *The Journal of Fixed Income*, 16(4), 60–68. <https://doi.org/10.3905/jfi.2007.683318>
- Cantor, R., & Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Economic Policy Review (19320426)*,.
- Canuto, O., dos Santos, P. F. P., & de Sá Porto, P. C. (2012). MACROECONOMICS AND SOVEREIGN RISK RATINGS. *Journal of International Commerce, Economics and Policy*, 03(02), 1250011. <https://doi.org/10.1142/S1793993312500111>
- Caunedo J, Dicecio R, Komunjer I, & Owyang M.T. (2020). Asymmetry, Complementarities, and State Dependence in Federal Reserve Forecasts. *Journal of Money, Credit & Banking (John Wiley & Sons, Inc.)*, 52(1), 205–228. <https://doi.org/10.1111/jmcb.12590>
- Cheng, J., Wang, Z., & Pollastri, G. (2008). A neural network approach to ordinal regression. *Proceedings of the International Joint Conference on Neural Networks*. <https://doi.org/10.1109/IJCNN.2008.4633963>
- Chih-Wei Hsu, Chih-Chung Chang, & Chih-Jen Lin. (2008). A Practical Guide to Support Vector Classification. *BJU International*, 101(1).
- Choy, S. Y., Chit, M. M., & Teo, W. L. (2021). Sovereign credit ratings: Discovering unorthodox factors and variables. *Global Finance Journal*, 48. <https://doi.org/10.1016/j.gfj.2020.100548>
- Chu, W., & Keerthi, S. S. (2007). Support Vector Ordinal Regression. *Neural Computation*, 19(3), 792–815. <https://doi.org/10.1162/neco.2007.19.3.792>

Bibliografía

- Clatworthy, M. A., Peel, D. A., & Pope, P. F. (2012). Are analysts' loss functions asymmetric? *Journal of Forecasting*, 31(8). <https://doi.org/10.1002/for.1253>
- De Moor, L., Luitel, P., Sercu, P., & Vanpée, R. (2018). Subjectivity in sovereign credit ratings. *Journal of Banking and Finance*. <https://doi.org/10.1016/j.jbankfin.2017.12.014>
- De Pascalis, F. (2016). Reducing regulatory reliance on credit ratings to address investors' over-reliance: Some thoughts in light of the US experience. *Capital Markets Law Journal*, 11(4). <https://doi.org/10.1093/cmlj/kmw020>
- Deng, W. Y., Zheng, Q. H., Lian, S., Chen, L., & Wang, X. (2010). Ordinal extreme learning machine. *Neurocomputing*, 74(1–3). <https://doi.org/10.1016/j.neucom.2010.08.022>
- Dimitriadou, A., Agrapetidou, A., Gogas, P., & Papadimitriou, T. (2021). Credit Rating Agencies: Evolution or Extinction? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3935710>
- Directorate-General for Internal Market, I. E. and Sme. (2014). *Response to the Financial Stability Board request for action plans to reduce reliance on CRA Ratings*.
- Dornbusch, R., Fischer, S., & Startz, R. (1991). *Macroeconomía*. McGraw-Hill; 5a ed.
- Economic andl Financial Affairs*. (2023). https://commission.europa.eu/about-european-commission/departments-and-executive-agencies/economic-and-financial-affairs_e
- Eliasson, A.-C. (2002). Sovereign credit ratings. *Research Notes*.

Bibliografía

Erdem, O., & Varli, Y. (2014). Understanding the sovereign credit ratings of emerging markets. *Emerging Markets Review*, 20. <https://doi.org/10.1016/j.ememar.2014.05.004>

ESPON. (2014). *ECR2 Economic Crisis: Resilience of Regions*.

European Central Bank Asset purchase programmes. (2023). <https://www.ecb.europa.eu/mopo/implement/app/html/index.en.html>

European Commission. (2018). *Eurostat*. <http://ec.europa.eu/eurostat>

European Commission Directorate-General for Economic and Financial Affairs. (n.d.). *European Economic Forecast Spring 2003 to 2018: Vol. EUROPEAN E*. Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs. (2004). *European Economic Forecast Spring 2004: Vol. EUROPEAN E*. Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs. (2005). *European Economic Forecast Spring 2005: Vol. EUROPEAN E*. Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs. (2006). *European Economic Forecast Spring 2006: Vol. EUROPEAN E*. Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs. (2007). *European Economic Forecast Spring 2007: Vol. EUROPEAN E*. Publications Office of the European Union.

Bibliografía

European Commission Directorate-General for Economic and Financial Affairs.
(2008). *European Economic Forecast Spring 2008: Vol. EUROPEAN E*.
Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs.
(2009). *European Economic Forecast Spring 2009: Vol. EUROPEAN E*.
Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs.
(2010). *European Economic Forecast Spring 2010: Vol. EUROPEAN E*.
Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs.
(2011). *European Economic Forecast Spring 2011: Vol. EUROPEAN E*.
Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs.
(2012). *European Economic Forecast Spring 2012: Vol. EUROPEAN E*.
Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs.
(2013). *European Economic Forecast Spring 2013: Vol. EUROPEAN E*.
Publications Office of the European Union.

European Commission Directorate-General for Economic and Financial Affairs.
(2014). *European Economic Forecast Spring 2014: Vol. EUROPEAN E*.
Publications Office of the European Union. <https://doi.org/10.2765/76645>
(print)

Bibliografía

- European Commission Directorate-General for Economic and Financial Affairs. (2015). *European Economic Forecast Spring 2015: Vol. EUROPEAN E*. Publications Office of the European Union. <https://doi.org/10.2765/176015> (print)
- European Commission Directorate-General for Economic and Financial Affairs. (2016). *European Economic Forecast Spring 2016: Vol. INSTITUTIO*. Publications Office of the European Union. <https://doi.org/10.2765/345429> (print)
- Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, 15, 3133–3181.
- Fernández-Navarro, F., Campoy-Muñoz, P., La Paz-Marín, M. De, Hervás-Martínez, C., & Yao, X. (2013). Addressing the EU sovereign ratings using an ordinal regression approach. *IEEE Transactions on Cybernetics*. <https://doi.org/10.1109/TSMCC.2013.2247595>
- Ferri, G., Liu, L.-G., & Stiglitz, J. E. (1999). The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis. *Economic Notes*, 28(3). <https://doi.org/10.1111/1468-0300.00016>
- Financial Stability Board. (2014). *Thematic review of the FSB principles for reducing reliance on CRA ratings*. <http://vlex.com/vid/thematic-review-of-the-816147689>
- Fitch Ratings. (2010). *Sovereign Rating Methodology*.
- Fitch Ratings. (2012). *Sovereign Rating Criteria*. *Fitch Ratings, March*.

Bibliografía

Fitch Ratings. (2017). *Sovereign Rating Criteria*.

Fitch Ratings. (2022). <http://www.fitchratings.com>

FMI. (2010). Sovereigns, Funding and Systemic Liquidity. *Global Financial Stability Report, October*.

Frank, E., & Hall, M. (2001). A simple approach to ordinal classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2167, 145–156. https://doi.org/10.1007/3-540-44795-4_13

Gade, T., Salines, M., Glöckler, G., & Strodthoff, S. (2021). “Loose Lips Sinking Markets?”: The Impact of Political Communication on Sovereign Bond Spreads. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2269624>

Gültekin-Karakaş, D., Hisarciklilar, M., & Öztürk, H. (2011). Sovereign risk ratings: Biased toward developed countries? *Emerging Markets Finance and Trade*, 47(SUPPL. 2). <https://doi.org/10.2753/REE1540-496X4703S204>

Gutiérrez, P. A., Pérez-Ortiz, M., Sánchez-Monedero, J., Fernández-Navarro, F., & Hervás-Martínez, C. (2016). Ordinal Regression Methods: Survey and Experimental Study. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2015.2457911>

Hernández Orallo, J., Ramírez Quintana, M. J., & Ferri Ramírez, C. (2005). *Introducción a la Minería de Datos* (S. A. Pearson Educación, Ed.; 2004th ed.). Pearson Educación, S.A.

Bibliografía

- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 20(8), 832–844.
- Holthausen, R. W., & Leftwich, R. W. (1986). The effect of bond rating changes on common stock prices. *Journal of Financial Economics*, 17(1).
[https://doi.org/10.1016/0304-405X\(86\)90006-1](https://doi.org/10.1016/0304-405X(86)90006-1)
- Hosmer, D. W., & Lemeshow, S. (2000). Applied logistic regression. 2nd Edition. In *John Wiley & Sons, Inc.*
- HOUSE OF LORDS European Union Committee. (2011). *Sovereign Credit Ratings: Shooting the Messenger? Vol. 21 de las*. The Stationery Office Limited.
- Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415–425.
<https://doi.org/10.1109/72.991427>
- Hu, Y.-T., Kiesel, R., & Perraudin, W. (2002). The estimation of transition matrices for sovereign credit ratings. *Journal of Banking and Finance*, 26, 1383–1406.
[https://doi.org/10.1016/S0378-4266\(02\)00268-6](https://doi.org/10.1016/S0378-4266(02)00268-6)
- International Monetary Fund. (2013). *Global Financial Stability Report, April 2013*.
- John Kiff and Liliana Schumacher, M. K. (2013). Rating Through-The-Cycle: What Does the Concept Imply for Rating Stability and Accuracy? *IMF Research Bulletin*. <http://international.vlex.com/vid/rating-does-imply-stability-accuracy-484785046>

Bibliografía

- Jones, L., Alsakka, R., ap Gwilym, O., & Mantovan, N. (2022). Regulating rating agencies: A conservative behavioural change. *Journal of Financial Stability*, 60. <https://doi.org/10.1016/j.jfs.2022.100999>
- Koopman, S. J., Kräussl, R., Lucas, A., & Monteiro, A. B. (2009). Credit cycles and macro fundamentals. *Journal of Empirical Finance*, 16(1). <https://doi.org/10.1016/j.jempfin.2008.07.002>
- Lee, H., Kim, S. G., Park, H. woo, & Kang, P. (2014). Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forecasting and Social Change*, 86. <https://doi.org/10.1016/j.techfore.2013.08.020>
- Li, J. P., Mirza, N., Rahat, B., & Xiong, D. (2020). Machine learning and credit ratings prediction in the age of fourth industrial revolution. *Technological Forecasting and Social Change*, 161. <https://doi.org/10.1016/j.techfore.2020.120309>
- Lin, H. T., & Li, ng. (2012). Reduction from cost-sensitive ordinal ranking to weighted binary classification. *Neural Computation*, 24(5). https://doi.org/10.1162/NECO_a_00265
- Lin, H. T., & Li, L. (2006). Large-margin thresholded ensembles for ordinal regression: Theory and practice. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4264 LNAI. https://doi.org/10.1007/11894841_26
- Löffler, G. (2004). An anatomy of rating through the cycle. *Journal of Banking and Finance*, 28(3), 695–720. [https://doi.org/10.1016/S0378-4266\(03\)00041-4](https://doi.org/10.1016/S0378-4266(03)00041-4)

Bibliografía

- Martín-García, M., Téllez-Valle, C., & Martín-Marín, J. L. (2014). Evolution of Sovereign Rating Models in the Current Crisis. *GCG Georgetown University - Universia*, 8(1), 16–33. <https://doi.org/0.3232/GCG.2014.V8.N1.01>
- Mathieson, M. (2001). Ordinal Models for Neural Networks. *PROGRESS IN NEURAL PROCESSING*, E 2, 523–536. <https://recursos.ulozola.es/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsbl&AN=CN015874462&lang=es&site=eds-live&scope=site>
- McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2). <https://doi.org/10.1111/j.2517-6161.1980.tb01109.x>
- Miricescu, C. (2014). INVESTIGATING THE DETERMINANTS OF LONG-RUN SOVEREIGN RATING. *Financial and Monetary Economics*.
- Miricescu, E.-C., Tâtu, L., & Cornea, D. (2016). The Determinants of the Sovereign Debt Rating: Evidence for the European Union Countries. *Economic Computation & Economic Cybernetics Studies & Research*, 50(1), 175–188.
- Moody's. (2013). *Rating methodology: Sovereign bond ratings*.
- Moody's. (2022). *Moody's*. <http://www.moodys.com>
- Moody's Investor Service. (2008). *Sovereign Bond Ratings*.
- Moody's Investor Service. (2009). *Rating Symbols and Definitions*.
- Moody's Investor Service. (2016a). *Rating Symbols and Definitions*.

Bibliografía

- Moody's Investor Service. (2016b). *Sovereign bond ratings*.
- Moody's Investors Service. (2010). *Sovereign Methodology Update: Narrowing the Gap – A Clarification of Moody's Approach to Local vs. Foreign Currency Government Bond Ratings*.
- Mora, N. (2006). Sovereign credit ratings: Guilty beyond reasonable doubt? *Journal of Banking and Finance*, 30, 2041–2062. <https://doi.org/10.1016/j.jbankfin.2005.05.023>
- Mulder, C. B., & Monfort, B. (2000). Using Credit Ratings for Capital Requirements on Lending to Emerging Market Economies: Possible Impact of a New Basel Accord. *IMF Working Papers*.
- Oprea, F., Onofrei, M., Lupu, D., Vintila, G., & Paraschiv, G. (2020). The determinants of economic resilience. The case of Eastern European regions. *Sustainability (Switzerland)*, 12(10). <https://doi.org/10.3390/su12104228>
- Overes, B. H. L., & van der Wel, M. (2022). Modelling Sovereign Credit Ratings: Evaluating the Accuracy and Driving Factors using Machine Learning Techniques. *Computational Economics*. <https://doi.org/10.1007/s10614-022-10245-7>
- Ozturk, H. (2014). The Origin of Bias in Sovereign Credit Ratings: Reconciling Agency Views with Institutional Quality. *The Journal of Developing Areas*, 48(4), 161–188. <https://doi.org/10.1353/jda.2014.0066>
- Ozturk, H., Namli, E., & Erdal, H. I. (2016). Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. *Economic Modelling*. <https://doi.org/10.1016/j.econmod.2016.01.012>

Bibliografía

- Pérez-Ortíria, M., Gutiérrez, P. A., & Hervás-Martínez, C. (2014). Projection-based ensemble learning for ordinal regression. *IEEE Transactions on Cybernetics*, 44(5). <https://doi.org/10.1109/TCYB.2013.2266336>
- Pontarollo, N., & Serpieri, C. (2020). A composite policy tool to measure territorial resilience capacity. *Socio-Economic Planning Sciences*, 70. <https://doi.org/10.1016/j.seps.2018.11.006>
- Quinlan, J. (1996). *Bagging, boosting, and C4.5*.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning / J. Ross Quinlan*. San Mateo, Calif. : Morgan Kaufmann Publishers, c1993.
- por el que se modifica el Reglamento (CE) n ° 1060/2009 sobre las agencias de calificación crediticia*, (2013) (testimony of Reglamento (UE) n ° 462/2013 del Parlamento Europeo y del Consejo).
- Reusens, P., & Croux, C. (2017). Sovereign credit rating determinants: A comparison before and after the European debt crisis. *Journal of Banking and Finance*, 77. <https://doi.org/10.1016/j.jbankfin.2017.01.006>
- Riedmiller, M., & Braun, H. (1993). Direct adaptive method for faster backpropagation learning: The RPROP algorithm. *1993 IEEE International Conference on Neural Networks*, 586–591. <https://doi.org/10.1109/icnn.1993.298623>
- Rodríguez de Codes Elorriaga, E., Marcelo Antuña, A., Blanco Escolar, R., Mayordomo Gómez, S., Arrizabalaga Uriarte, F., & Stupariu, P. F. (2020). RETOS ASOCIADOS AL USO DE LAS CALIFICACIONES CREDITICIAS

Bibliografía

DE LAS AGENCIAS EN EL CONTEXTO DE LA CRISIS DEL COVID-19 .
Revista de Estabilidad Financiera.

Sánchez-Monedero, J., Campoy-Muñoz, P., Gutiérrez, P. A., & Hervás-Martínez, C. (2014). A guided data projection technique for classification of sovereign ratings: The case of European Union 27. *Applied Soft Computing Journal*, 22, 339–350. <https://doi.org/10.1016/j.asoc.2014.05.008>

Sánchez-Monedero, J., Gutiérrez, P. A., & Pérez-Ortiz, M. (2019). ORCA: A matlab/octave toolbox for ordinal regression. *Journal of Machine Learning Research*, 20.

Sánchez-Monedero, J., Pérez-Ortiz, M., Sáez, A., Gutiérrez, P. A., & Hervás-Martínez, C. (2018). Partial order label decomposition approaches for melanoma diagnosis. *Applied Soft Computing Journal*, 64. <https://doi.org/10.1016/j.asoc.2017.11.042>

Singh, M. K., Gómez-Puig, M., & Sosvilla-Rivero, S. (2021). Quantifying sovereign risk in the euro area. *Economic Modelling*, 95. <https://doi.org/10.1016/j.econmod.2020.12.010>

Slapnik, U. (2019). Understanding Sovereign Credit Ratings: Text-Based Evidence From the Credit Rating Reports. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3372270>

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. In *Statistics and Computing* (Vol. 14, Issue 3). <https://doi.org/10.1023/B:STCO.0000035301.49549.88>

Bibliografía

- Spaventa, L. (1987). The Growth of Public Debt: Sustainability, Fiscal Rules, and Monetary Rules. *IMF Staff Papers*, 34(2), 374–399. <https://doi.org/10.2307/3867139>
- Standard & Poor's. (2013). *Sovereigns: Sovereign government rating methodology and assumptions*.
- Standard & Poor's. (2017). S&P Sovereign Rating Methodology. In *S&P Global Ratings* (Issue 1).
- Standard & Poor's RatingsDirect. (2009). *Understanding Standard & Poor's Rating Definitions*. www2.standardandpoors.com/spf/pdf/fixedincome/Understanding_Rating_Definitions.pdf.
- Sun, B. Y., Li, J., Wu, D. D., Zhang, X. M., & Li, W. B. (2010). Kernel discriminant learning for ordinal regression. *IEEE Transactions on Knowledge and Data Engineering*, 22(6). <https://doi.org/10.1109/TKDE.2009.170>
- Sy, A. N. R. (2002). Emerging market bond spreads and sovereign credit ratings: Reconciling market views with economic fundamentals. *Emerging Markets Review*, 3(4). [https://doi.org/10.1016/S1566-0141\(02\)00043-2](https://doi.org/10.1016/S1566-0141(02)00043-2)
- Sylla, R. (2002). *An Historical Primer on the Business of Credit Rating*. https://doi.org/10.1007/978-1-4615-0999-8_2
- Haque, N., Mathieson, D., & Mark, N. (1997). Rating the raters of country creditworthiness. *Finance and Development*, 34(1).
- Versión Consolidada del Tratado de Funcionamiento de la Unión Europea, Diario Oficial de la Unión Europea C 326/47 (2012).

Bibliografía

- Unión Europea. (2017). *Unión Europea*. Web Oficial de La Unión Europea. https://europa.eu/european-union/index_es%0A
- U.S. Securities and Exchange Commission. (2023). *U.S. Securities and Exchange Commission*. <https://www.sec.gov/about/divisions-offices/office-credit-ratings/current-nrsros>
- Van Gestel, T., Baesens, B., Van Dijcke, P., Garcia, J., Suykens, J. A. K., & Vanthienen, J. (2006). A process model to develop an internal rating system: Sovereign credit ratings. *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2005.10.001>
- Vu, H., Alsakka, R., & ap Gwilym, O. (2022). Does competition improve sovereign credit rating quality? *Journal of International Financial Markets, Institutions and Money*, 76. <https://doi.org/10.1016/j.intfin.2021.101478>
- Vugrin, E. D., Warren, D. E., Ehlen, M. A., & Camhouse, R. C. (2010). A framework for assessing the resilience of infrastructure and economic systems. In *Sustainable and Resilient Critical Infrastructure Systems: Simulation, Modeling, and Intelligent Engineering*. https://doi.org/10.1007/978-3-642-11405-2_3
- Wang, X., Zeng, D., Dai, H., & Zhu, Y. (2020). Making the right business decision: Forecasting the binary NPD strategy in Chinese automotive industry with machine learning methods. *Technological Forecasting and Social Change*, 155. <https://doi.org/10.1016/j.techfore.2020.120032>

ANEXOS

ANEXOS

1 Artículo JCR

2 Participación en congreso 1

3 Participación en congreso 2



Revisiting the determinants of sovereign debt ratings in Europe through artificial intelligence techniques

Carlos Galnares ^a, Alfonso Carlos Martínez-Estudillo ^a, Mariano Carbonero-Ruz ^a and Pilar Campoy-Muñoz ^b

^aDepartment of Quantitative Methods, Universidad Loyola Andalucía, Cordoba, Spain; ^bDepartment of Economics, Universidad Loyola Andalucía, Cordoba, Spain

ABSTRACT

In papers using artificial intelligence (AI) techniques, little attention has been paid to the determinants of sovereign debt ratings. We propose a reduced set of variables regarding the economic performance of a country that are consistent with the idea of debt sustainability. The robustness of this set is supported by the results obtained with different well-known AI techniques using data from EU-15 countries during the 2002–2017 period as the experimental setting. The variables are publicly available, allowing a quick and reliable assessment of the creditworthiness of a sovereign and providing useful information for decision-makers and investors.

KEYWORDS

Sovereign credit ratings; credit rating agencies; EU-15; artificial intelligence

JEL CLASSIFICATION

C38; C45; G15; G24

I. Introduction



The coronavirus outbreak has already brought major economic disruption, and financial markets reflect this. As a result, sovereigns may face negative outlooks or downgrades of their creditworthiness, affecting their funding cost. However, these impacts are very costly in the long run, since sovereign credit rates (SCRs) are benchmark indicators for credit risk assessment and investment mandates of many other financial assets (Borensztein, Cowan, and Valenzuela 2013). The over-reliance of national financial systems on agency ratings (CRAs) has led to a vivid debate on the opaqueness of the rating process (Ozturk, Namli, and Erdal 2016). In response, a new strand of the literature has focused on the prediction of ratings by using artificial intelligence (AI), aiming to demonstrate that these techniques perform better than statistical techniques or even other AI techniques. Unfortunately, less attention has been given to the indicators or determinants employed within these works, and there are a lack of one-to-one relationships between the number of variables and the characteristics of the empirical design (Table 1). In this note, we explore the prediction performance of a reduced set of variables for sovereign creditworthiness across four different AI

approaches. We illustrate the proposal with data from the EU-15 Member States¹ for the period 2002–2017.

II. Data

To assess the creditworthiness of a given sovereign, 7 determinants have been chosen to be consistent with the concept of debt sustainability pointed out by Blanchard et al. (1990), according to which the ratio of government debt to nominal GDP eventually converges at a reasonable level (Table 2). This approach has been enriched by considering forecast of GDP, following the idea raised by Polito and Wickens (2015) and, also GDP per capita and Resilience, as a measure of GDP growth, given its increasing relevance for CRAs after European debt crisis (Reusens and Croux 2017).

The variable *Resilience_{it}* (Manca, Benczur, and Giovannini 2017) is calculated as the accumulated differences between the level of income prior to the downturn (GDP_i^*) in year t_0 and the level of income for each year (GDP_{it}) until such difference is equal to zero in year t_1 . The accumulated value for each year is expressed as a share of the level of income for the corresponding year:

CONTACT Carlos Galnares  cgalnaresjp@gmail.com  Department of Quantitative Methods, Universidad Loyola Andalucía, Escritor Castilla Aguayo 4 Cordoba 14004, Spain

¹Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom.



EcoMod 2018 International Conference on Economic Modeling

CERTIFICATE OF PARTICIPATION

We hereby certify that

Carlos Galnares Jiménez-Placer; Universidad Loyola Anadalucía, has presented the paper entitled:

Estimating the sovereign ratings in the UE-15 2002-2015

at EcoMod 2018 International Conference on Economic Modeling, held at Ca' Foscari University of Venice

from 4 – 6, July 2018.

Sincerely,

A handwritten signature in black ink that reads "Theresa Leary".

Theresa Leary
EcoMod, Manager
7/3/2018





Vigo 19 a 23
junio

XXXIII CONGRESO INTERNACIONAL
DE ECONOMÍA APLICADA

Asepelt 2019
economía azul

CERTIFICADO
DE COMUNICACIÓN

Certificamos que el trabajo titulado

Estimación de ratings de deuda soberana en la ue15 de 2002 a 2017 mediante clasificadores ordinales

de **Carlos Galnares Jiménez-Placer**

y que tiene como coautores/as a

Alfonso Carlos Martínez Estudillo y Mariano Carbonero Ruz

ha sido presentado como comunicación en el XXXIII Congreso Internacional de Economía Aplicada Asepelt 2019: Economía Azul, celebrado en Vigo del 19 al 23 de junio de 2019.

Y para que así conste y surta los efectos oportunos, expiden y firman el presente certificado en Vigo, a 23 de junio 2019

Prof. Dr. Carlos María Fernández-Jardón
Presidente Comité Organizador

Dr. José María Montero Lorenzo
Presidente de Asepelt

Organiza

 **Asepelt**
Asociación Internacional de Economía Aplicada

Universida de Vigo

Colabora

 **ECOBAS**
Economics and Business Administration for Society

CONCELLO
DE VIGO



ZONA FRANCA
VIGO

 **Puerto de Vigo**
Autoridad Portuaria de Vigo



Tesis depositada en Universidad Loyola
Sevilla, 2023

