

RATINGS BANCARIOS DESPUÉS DE LA CRISIS FINANCIERA: MACHINE LEARNING VS MODELOS TRADICIONALES

Daniel García Cerveró

Trabajo de investigación 019/028

Master en Banca y Finanzas Cuantitativas

Director/a: Dr. Juan Fernández de Guevara

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia

Universidad de Castilla-La Mancha

www.finanzasquantitativas.com

RATINGS BANCARIOS DESPUÉS DE LA CRISIS FINANCIERA: MACHINE LEARNING VS MODELOS TRADICIONALES

Daniel García Cervero

Trabajo de investigación 019/020
Máster en Banca y Finanzas Cuantitativas

Director/a: Juan Fernández de Guevara

Universidad Complutense de Madrid
Universidad del País Vasco
Universidad de Valencia
Universidad de Castilla-La Mancha

www.finanzascuantitativas.com

ÍNDICE

ÍNDICE	2
1. INTRODUCCIÓN	4
2. LITERATURA PREVIA	5
3. DESCRIPCIÓN DE LOS DATOS Y VARIABLES EMPLEADAS	7
4. MÉTODOS DE PREDICCIÓN Y CONTRASTE EMPLEADOS	10
4.1 Métodos de predicción	10
4.1.1 Logit ordenado con efectos fijos.....	10
4.1.2 Random Forest.....	11
4.2 Método de contraste de modelos	14
5. RESULTADOS.....	15
6. CONCLUSIONES.....	18
REFERENCIAS.....	20
CUADROS	22
ANEXO.....	31

RESUMEN

Este trabajo analiza la evolución y los determinantes de los *ratings* bancarios desde el estallido de la crisis financiera, contrastando las mejoras en la predicción de los *ratings* por la utilización de técnicas de *machine learning* en lugar de los modelos logit/probit ordenados habituales en la literatura, analizando los principales determinantes de esta evolución. Mas precisamente, se utiliza un modelo de *machine learning* (*random forest*) para contrastar la robustez de los resultados frente a los modelos ordenados habituales. La muestra está formada los bancos con *rating* corporativos disponibles en 23 países de todo el mundo a lo largo del periodo comprendido entre 2003 y 2016, ambos inclusive. Los resultados obtenidos mediante los modelos ordenados son similares a lo que se suele obtener en la literartura, particularmente en lo que se refiere a la importancia de cada determinante de los *ratings*. Además, se constata que las estimaciones mediante el método de *random forest* proporciona unas predicciones más precisas en las tres agencias que los modelos logit ordenados, particularmente cuando existe un reducido de observaciones, debido al mejor aprovechamiento de la información de este procedimiento en comparación a los modelos ordenados.

1. INTRODUCCIÓN

En 2007 tuvo lugar una de las mayores crisis financieras globales que se recuerdan en las últimas décadas originada, entre otros factores, por una mala evaluación del riesgo por parte de las agencias de calificación que provocó el estallido de la burbuja de las hipotecas subprime. La crisis, que comenzó en EEUU y se propagó al resto del mundo, puso en entredicho los criterios empleados por las principales agencias de calificación (Fitch, Moody's y Standard and Poor's) a la hora de asignar los *ratings* que suelen utilizarse como indicadores de la calidad crediticia de empresas y gobiernos. Fueron acusadas de falta de transparencia en sus criterios de asignación de *ratings*. Además se señaló su modelo de negocio, en el que existía un conflicto de interés ya que el cliente que las contrataba era el que precisamente tenía que recibir el *rating* (Bolton, Freixas & Shapiro, 2012). Este hecho, unido a la gran importancia que los inversores daban a estos sistemas de calificación a la hora de tomar sus decisiones de inversión, provocaron que los organismos reguladores de todo el mundo pusieran el foco tras la crisis en estos *ratings* y comenzasen a regular y supervisar de forma más estricta tanto a las agencias como a los criterios de calificación aplicados por mismas.

A pesar de que las agencias de calificación (en adelante CRAs) justifican sus metodologías argumentando que sus *ratings* son emitidos teniendo en cuenta la evolución en el largo plazo (evaluación *through the cycle*) e ignorando hechos coyunturales (evaluación *point in time*), se ha demostrado que tras la crisis estas agencias empeoraron los *ratings* que asignaban. Sin embargo, cabe preguntarse si esta rebaja generalizada de las calificaciones fue debida al endurecimiento de la regulación, al cambio en los criterios de asignación de *ratings* o a un empeoramiento de la situación financiera de los bancos. Salvador et al (2018) analizan precisamente esta cuestión, obteniendo que, como era de esperar, se produjo una caída generalizada en los *ratings* emitidos por las distintas agencias a raíz de la crisis, salvo en el caso de Japón. Concretamente el ajuste fue, en media, de un 8,75% en Fitch, un 8,17% en Standard and Poor's y de un 15,94% en Moody's, para los bancos Europeos y de Estados Unidos. De este ajuste, el 39,96% y el 19,21% fue debido a un cambio en los criterios de evaluación de las agencias en el caso de Fitch y Standard and Poor's respectivamente, mientras que en el caso de Moody's, fue debido casi en su totalidad al empeoramiento de la calidad crediticia de las empresas calificadas. Estos resultados sugieren que, a diferencia de Fitch y Standard and Poor's, la única agencia que de algún modo sigue una perspectiva más a largo plazo a la hora de asignar los *ratings* es Moodys, puesto que ha sido la única que ha demostrado un comportamiento menos procíclico al no castigar demasiado a las entidades financieras en los *ratings* durante el periodo de crisis.

Siguiendo en esta línea de investigación, el objetivo de este trabajo es proponer una metodología alternativa para el análisis realizado por estos autores basadas en modelos logit ordenados, empleando una técnica de *machine learning* conocida como *random forest*. Esta técnica, a diferencia del logit ordenado, es un algoritmo que estima múltiples árboles de decisión, para posteriormente agregarlos obteniendo así una predicción más precisa y estable.

La principal diferencia entre ambos modelos es que, en casos como este, en el que el número de variables es alto y las relaciones entre ellas y con los *ratings* puede no ser lineal, es de esperar que el método de *random forest* ajuste con más precisión. Por otro lado, como se describe en la sección 4.1.2, el *random forest* es un modelo que requiere de la fijación de forma externa de diversos parámetros (hiperparámetros), lo que supone que, aunque permite un mejor ajuste de la predicción, incrementa el riesgo de sobre-ajustar el modelo a una determinada muestra.

El trabajo se ha organizado de la siguiente manera: En primer lugar, en la sección 2 se presenta la literatura previa relacionada con la modelización de *ratings*. Posteriormente, en la sección 3, se describen la muestra y variables utilizadas. En la sección 4 se presentan los dos métodos que serán aplicados para la predicción de los *ratings* bancarios, logit ordenado y *radom forest*. Los resultados obtenidos con cada modelo y la comparación de las predicciones de las distintas metodologías se realiza en la sección 5. Finalmente, en la sección 6, se presentan las conclusiones principales del trabajo.

2. LITERATURA PREVIA

La modelización del proceso que siguen las agencias de calificación a la hora de emitir un *rating* es un tema de estudio sobre el que existe una literatura previa extensa, por lo que se han propuesto numerosos modelos capaces de predecir los *ratings* bancarios (Morgan, 2002; Godlewski, 2007; Iannotta et al, 2008; Peresetsky and Karminsky, 2011; Bellotti et al., 2011; Caporale et al., 2011, entre otros).

Sin embargo, a pesar de que las tres principales agencias proporcionan guías sobre el proceso de emisión de dichos *ratings*, sigue existiendo una gran incertidumbre en torno al mecanismo de asignación de *ratings*. La característica principal de este proceso de emisión de *ratings* es la falta de transparencia (Bellotti et al., 2011; Iannotta, 2006). Esta falta de transparencia es evidente al realizar un análisis del proceso de evaluación que se lleva a cabo previamente a la asignación de *ratings*. Según las CRAs, los *ratings* son emitidos teniendo en cuenta la evolución de la salud crediticia en el largo plazo, como demuestran Altman & Rijken (2004). Sin embargo, en los trabajos de Salvador et al. (2014; 2018) se señala como, a consecuencia de la crisis, tiene lugar un cambio generalizado de los *ratings* tanto de Fitch como de S&P. Moody's en cambio, no parece implementar cambios en su forma de evaluación asociados a variaciones transitorias en la situación patrimonial de las entidades. La falta de transparencia, y la no objetivación de los motivos que producen cambios de *rating* es crucial, ya que como concluye Gärtner et al. (2011), existe la posibilidad de que el empeoramiento de un *rating* desencadene un proceso de autorrealización que puede llevar incluso a países relativamente sanos hacia el *default*.

Numerosos autores han tratado de analizar este proceso investigando las variables más importantes a la hora de predecir los *ratings* bancarios. Por ejemplo, Ögüt et al. (2012) encuentran evidencia de que la rentabilidad (medida por el rendimiento sobre el capital), el uso eficiente de los recursos y financiación de las empresas y los hogares

en lugar del gobierno, son factores importantes a la hora de modelizar los *ratings*. En esta línea, Peresetsky et al. (2011) y Bellotti et al. (2011), encuentran evidencias de que además de los factores financieros, los factores macroeconómicos también son influyentes. Otros autores como Alp (2013), Baghai et al. (2014) y Salvador et al. (2018), tratan de investigar otros motivos subyacentes a los cambios en los *ratings*. Entre estos se incluyen los cambios en la regulación de los países o los *ratings* soberanos.

Además, varios estudios han analizado también la heterogeneidad de los *ratings* entre países de la UE (Caporale et al., 2011), entre distintos países de todo el mundo (Shen et al., 2012; Salvador et al., 2018) o entre las propias agencias de calificación (Morgan, 2002; Iannotta, 2006).

Metodológicamente ha habido una amplia investigación en la predicción de los *ratings* de los bonos utilizando modelos logit y probit ordenados como, por ejemplo: Poon et al., 1999; Morgan, 2002; Altman and Rijken, 2004; Amato and Furfine, 2004; Iannotta, 2006; Bellotti et al., 2011; Gärtner, 2011; Shen et al. 2012; Alp, 2013; Salvador et al., 2014 y 2018. A pesar de ser los modelos mayoritarios en la literatura previa, en los últimos años los modelos de inteligencia artificial se han ido abriendo paso, aunque en los años 90 ya había publicaciones relacionadas con el *machine learning*. Sin embargo, no existe un consenso generalizado sobre qué modelos predicen mejor. Por un lado, tenemos los estudios de Kim (2005) y Lee (2007), que muestran que los modelos de inteligencia artificial no proporcionan mejores predicciones de los *ratings* de los bonos que los métodos estándar de elección ordenada, aunque Kim (1993) o Bellotti et al. (2011) sí que obtienen una mejor predicción.

La mayoría de los trabajos que han empleado inteligencia artificial, han empleado modelos como redes neuronales o los *support vector machines*, sin embargo, en este trabajo se emplea un modelo distinto (*random forest*), empleado por Ozturk et al. (2016) y por De Moor et al. (2017), para el caso de *ratings* soberanos.

La comparación entre regresiones logit ordenadas y los modelos de *random forest* es, obviamente, una forma válida de abordar el principal reto de la modelización de los *ratings*, que consiste en aumentar la probabilidad de conseguir una correcta clasificación de los bancos en un *rating* determinado. Sin embargo, no se conoce estudio previo que compare modelos de elección ordenada y *random forest* en términos de predicción de *issuer ratings* bancarios, que es el objetivo de este trabajo. Como punto de partida se ha tomado el estudio de Salvador et al. (2018), el cuál estima que proporción del cambio en los *ratings* tras la crisis financiera se debió al endurecimiento de los criterios de asignación de los *ratings*, y que parte al empeoramiento de la salud financiera de los bancos, partiendo de una muestra internacional de bancos y empleando un modelo probit ordenado de efectos individuales fijos.

El *random forest* es un procedimiento que obtiene unas predicciones de los *issuer ratings* a partir de múltiples árboles de decisión, calibrando cada uno de ellos con una combinación aleatoria distinta de la muestra original, de forma independiente y con

una combinación de los inputs (las variables explicativas) diferente en cada árbol (Breiman, L. 2001). Es decir, este modelo basa su estimación en los resultados de múltiples combinaciones de la muestra de datos y de las variables independientes del modelo. Podría decirse que estudia el *rating* que se obtiene al combinar de todas las formas posibles las variables del modelo, en muestras distintas, para así poder predecir posteriormente el *rating* mediante la elección del resultado más votado de entre todas estas combinaciones y con la máxima precisión posible.

3. DESCRIPCIÓN DE LOS DATOS Y VARIABLES EMPLEADAS

Este trabajo utiliza una muestra de más de 18 mil observaciones correspondientes a bancos de 23 países¹ de la OCDE obtenidas de BankScope para el periodo 2003-2016. De la base de datos se seleccionan todos los bancos para los que existe información sobre las variables necesarias para realizar el análisis, siempre que estos no sean considerados como atípicos. Como valores atípicos se han considerado aquellas observaciones, que en determinadas variables clave, toman valores por encima del percentil 75, y/o por debajo del percentil 25, +/- 1,5 veces el rango intercuartílico. Las variables seleccionadas para la detección de atípicos son las ratios de crédito y de depósitos sobre activo, la rentabilidad sobre activos ROA (beneficios antes de impuestos entre activos totales), la ratio de capital (recursos propios entre activos totales) y la ratio de costes medios (gastos de Explotación entre activos totales). Los *ratings* bancarios se obtienen también de BankScope. Aunque existe un número de bancos mayor para los que las agencias de calificación evalúan sus emisiones, el número de bancos con *rating* emisor es relativamente reducido, siendo la variable que más restringe la muestra en relación al contenido de BankScope. Por ejemplo, en España de un total de 94 bancos existentes en el año 2016, únicamente 16 en Fitch, 16 para Moody's y 10 para Standard and Poor's figuraban en BankScope con *rating* emisor.

El **Cuadro 1** muestra la distribución de observaciones por países y años para los cuáles alguna de las tres agencias de calificación tiene información del *rating*. Como se puede observar, hay años para los cuáles algún país no tiene información o la información que se tiene es reducida. En contraposición, en algunos países la muestra es más elevada en cuanto a la información del *rating* se refiere, como es el caso de Alemania, España, Italia, Japón, Francia y Estados Unidos (salvo los últimos dos años). Esta información, en ocasiones escasa, hace que presumiblemente el *random forest* se ajuste mejor a los datos, ya que, mientras que el logit ordenado solo muestra el resultado que proporciona una combinación de dicha base de datos, el *random forest* prueba muchas combinaciones y obtiene como resultado la media de ellas.

Tras la limpieza de *outliers* llevada a cabo y la descripción de los datos, se describe tanto las variables consideradas para dicha predicción, como los *ratings* para los cuales se desea predecir su valor.

¹ La muestra está constituida por bancos de los siguientes países: Austria, Bélgica, Republica Checa, Dinamarca, Finlandia, Francia, Alemania, Grecia, Hungría, Irlanda, Italia, Japón, Lituania, Luxemburgo, Holanda, Noruega, Polonia, Portugal, Eslovenia, España, Suecia, Inglaterra y Estados Unidos.

En cuanto a los *ratings* bancarios empleados, la mayoría de los estudios, excluyendo entre otros Morgan (2002), Iannotta et al. (2008) y Salvador et al. (2014), han empleado únicamente los llamados “*individual ratings*” de las tres principales agencias de calificación, los cuales tan solo tienen en cuenta la situación financiera del banco. Sin embargo, en este trabajo, al igual que en los tres estudios mencionados, se emplean los *issuer ratings*. Los *issuer ratings*, además de tener en cuenta la situación financiera de las entidades, también tienen en cuenta el apoyo que tienen los bancos por parte de sus propietarios o de las autoridades gubernamentales. Puesto que el objetivo de este trabajo es analizar el comportamiento de los *ratings* bancarios durante la crisis, se ha considerado adecuado utilizar este *rating* debido a que durante estos años la importancia del soporte gubernamental ha sido fundamental, ya que los estados se han visto obligados a rescatar numerosas entidades financieras.

Los *issuer ratings*, se clasifican en 22 categorías que oscilan desde los *ratings* de inversión, que van desde AAA/Aaa hasta BBB-/Baa3 hasta los *rating* especulativos, con mayor riesgo crediticio, que van desde BB+/Baa3 hasta D. Dado el reducido número de observaciones en alguna de las categorías de la escala, se han agrupado las categorías en 5 niveles, como ya hicieron en sus estudios previos Poon et al. (1999), Morgan (2002), Caporale et al. (2011), entre otros. Con esta segunda agrupación se busca unificar los *ratings* con menor número de observaciones para obtener así una mayor representatividad de los mismos. La escala está construida de forma que los *ratings* más altos son aquellos con una mayor calidad crediticia. El resultado de esta transformación ha sido ilustrado en el **Cuadro 2**.

Los factores considerados en este trabajo como predictores de los *issuer ratings*, han sido seleccionados en función de los que la literatura previa ha demostrado su relevancia para la predicción de *ratings* bancarios. En el **Anexo** puede encontrarse la definición precisa de las variables, incluyendo también los códigos de línea de Bankscope empleados para el cálculo.

De estos factores, tanto cualitativos como cuantitativos, empezaremos por el factor que muchos estudios metodológicos consideran clave en la evaluación de la calidad crediticia de los bancos, la rentabilidad. Al igual que Altman y Rijken (2004, 2006), se considera la **ROA** como medida de rentabilidad, puesto que es la relación entre el beneficio logrado en un determinado período y los activos totales de una empresa. Es un factor fundamental ya que, como señaló Moody's (2007a), determina la capacidad de un banco de generar valor económico y de absorber las pérdidas derivadas tanto de los riesgos asumidos por el banco, como de la situación económica. Como es lógico, se espera que el signo del coeficiente que acompañe a este factor sea positivo, aunque al tratar con datos de un período histórico tan amplio, esto podría no ser así en todos los casos.

Otro factor habitualmente utilizado como determinante del *rating* es el **Tamaño** del banco, obtenido a partir del logaritmo de los activos totales. Como se demuestra en los estudios de Caporale et al. (2011) y Shen et al. (2012), es de esperar que cuanto más grande sea el banco, más esfuerzos harán los gobiernos por impedir que este quiebre, debido al gran impacto que tendría su caída (hipótesis del Too Big to Fail). En base a esta hipótesis, se espera que el signo de este coeficiente sea positivo. Además, si esta percepción de menor riesgo la tienen también los clientes, el banco podrá financiarse a menor coste, por lo que se situará en una situación ventajosa frente a los rivales de menor tamaño.

A la hora de determinar un *rating* bancario, es también relevante conocer el nivel de solvencia del mismo, ya que, al igual que la rentabilidad, mayor volumen de capital supone un colchón a la hora de absorber pérdidas en escenarios desfavorables. Para medir este factor se ha elegido la ratio de **Capital**, construida como el cociente entre recursos propios entre activos totales del banco. Desafortunadamente no es posible obtener para la totalidad de bancos incluidos en la muestra la ratio de capital regulatoria, que sería un mejor indicador de la solvencia de la entidad. También se espera que el signo del coeficiente vinculado a este factor sea positivo, ya que la menor solvencia de un banco debería repercutir en términos de un menor *rating*.

La ratio de **Liquidez** se define como activos líquidos entre activos totales y se asume, que la capacidad de un banco para hacer frente a sus pagos de forma inmediata en un escenario de crisis, es fundamental a la hora de evitar su quiebra (Fitch 2011; Moody's 2007a). En consecuencia, es de esperar que el signo sea positivo también.

En los últimos años, hemos sido testigos de un aumento de la competencia bancaria. Las entidades financieras se han visto obligadas a cerrar numerosas sucursales a fin de reducir costes y ser así más eficientes. Por este motivo, otra ratio a considerar es la de **Eficiencia**, que se define por la ratio de eficiencia operativa (gastos de explotación y el margen bruto), como ya hicieran previamente Shen et al. (2012). El signo esperado para este coeficiente es negativo, ya que un alto valor de esta ratio indica que la actividad que desarrolla el banco conlleva más gastos por unidad de ingresos, por lo tanto, esto penalizará el *rating* emitido para dicha entidad.

Observando la estructura del balance de un banco, puede concluirse que otro factor relevante a la hora de determinar un *rating* es la ratio de **Crédito** sobre los activos totales, que representa la cantidad de activo susceptible de verse afectada por el riesgo de impago y alerta de desequilibrios en el balance debido a una concentración excesiva de riesgos en la actividad crediticia. En principio, cuanto más alto sea el valor de esta ratio, mayor será el riesgo de crédito y peor será el *rating*. Sin embargo, como ya señalaran Iannotta et al. (2008), los préstamos de un banco también pueden tener una contribución positiva al *rating* asignado, puesto que son más estables que otros activos como, por ejemplo, las acciones. Teniendo en cuenta ambas consideraciones, la influencia de esta variable tiene que ser establecida empíricamente.

Para medir la calidad crediticia de las carteras del balance, y puesto que en Bankscope no hay información relacionada con activos dudosos, se ha utilizado como proxy la ratio de dotaciones por **Pérdidas activos dudosos** (como porcentaje de los préstamos totales), la cual se espera que influya negativamente en el *rating*.

D'Apice et al. (2014) encontró evidencia de que las tres agencias de calificación tenidas en cuenta para este trabajo asignaban mejores *ratings* a aquellos bancos que desarrollan un modelo de negocio tradicional. Este estudio, junto con el llevado a cabo por De Young and Torna (2013), que encontró evidencias de que los bancos con un modelo de negocio no tradicional fueron más castigados durante la crisis, indican sin ninguna duda que otro factor a tener en cuenta a la hora de obtener el *rating* es el modelo de negocio que lleve a cabo cada entidad. Por ello, se define la ratio **Tradinc** como el margen de intereses entre el margen de intereses más los ingresos distintos de intereses. Es decir, Tradinc indica el porcentaje que los ingresos tradicionales de intermediación representan sobre el total de ingresos de explotación. El signo en este caso dependerá de cómo consideren las agencias de calificación la

diversificación de ingresos. Si consideran que una estructura diversificada de ingresos fortalece al banco, la influencia será negativa, mientras que, si consideran que el mayor peso de los ingresos tradicionales tiene ventaja frente a los no tradicionales el efecto sobre el *rating* será positivo.

Como ya se indicó al comienzo del apartado, los *ratings* bancaros, además de depender de factores cuantitativos, también dependen de factores cualitativos. Los factores cualitativos son de difícil medición por su propia naturaleza, sin embargo, la literatura suele incluirlas.

Como las propias agencias indican, el entorno económico en que operan los bancos influye en su calificación, así pues, los bancos que operan en economías más prósperas o con menos conflictos es de esperar que reciban una calificación mayor. Con el fin de captar estas características del entorno económico en las que opera cada banco, se ha elegido la tasa real de crecimiento del PIB (**Ciclo**). Como ya se ha mencionado, a mayor PIB de un país, mejor es su entorno económico, por lo que se espera un signo positivo del coeficiente que acompaña a esta variable.

Por último, al igual que Salvador et al. (2018), se han incluido tres variables ficticias con el fin de capturar el efecto que los cambios en la regulación europea, americana y japonesa han tenido sobre los *ratings* emitidos. Así pues, en 2010, entró en vigor la Ley de reforma de Dodd-Frank (2010). Para capturar este cambio de regulación, consideramos la variable ficticia **Reg_EEUU**, que toma valor 1 para los años 2010-2013 y cero para los demás. En Europa, el cambio de regulación tuvo lugar principalmente en julio de 2011 con el "Reglamento (UE) n° 513/2011", que asignaba la supervisión directa de las agencias de calificación crediticia a la Autoridad Europea de Valores y Mercados (ESMA) y aumentaba las responsabilidades de las agencias de calificación. Por lo tanto, se ha creado la variable dummy **Reg_EURO**, que toma un valor igual a la unidad para las calificaciones emitidas en Europa desde 2011, y cero en caso contrario. En el caso de Japón, se introduce en el modelo la variable ficticia **Reg_JAPAN**, la cual trata de captar los cambios regulatorios que se produjeron cuando los reguladores intervinieron el 17 de junio de 2009 con varias disposiciones que regulaban las agencias de calificación crediticia en virtud de la modificación de la Ley de Instrumentos Financieros y Bolsas. Estos cambios entraron en vigor el 1 de abril de 2010. Por ello, la variable toma un valor igual a la unidad para las calificaciones emitidas en Japón desde 2010, y cero en caso contrario.

4. MÉTODOS DE PREDICCIÓN Y CONTRASTE EMPLEADOS

4.1 Métodos de predicción

4.1.1 Logit ordenado con efectos fijos

El modelo de referencia empleado es el logit ordenado, que es el más utilizado en la literatura sobre los determinantes de los *ratings* bancarios. La finalidad de este modelo es estimar la probabilidad de que una observación con características particulares sea atribuida a una categoría específica de entre aquellas que componen la variable dependiente. El modelo logit ordenado asume que las distintas categorías que

conforman la variable independiente tienen un orden determinado, implicando que hay categorías mejores que otras. En definitiva, los modelos logit ordenados empleados, estiman la probabilidad de obtener un *rating* determinado en función de las variables consideradas a la hora de emitir un determinado *rating*. En este trabajo se han estimado dos modelos logit distintos. El primero es utilizado para predecir el *rating* según las características de los bancos y del entorno en el que operan:

$$Y_{it}^* = \beta' X_{it-1} + \varepsilon_i + v_{it} \quad (1)$$

donde Y_{it}^* es una variable latente, es decir, se utiliza para expresar el modelo en términos de probabilidad y expresa la *utilidad* o el *pago* que la agencia de calificación asigna a que un banco pertenezca a un *rating* concreto y no a otro. X_{it-1} , son las variables que definen los *ratings* bancarios objeto de este trabajo. Al igual que Altman y Rijken (2004), Caporale et al. (2011) y Salvador et al. (2018), los valores de los factores que determinan los *ratings* se han introducido con un período de retardo, $t-1$. La razón para proceder de esta manera es que los datos contables se emiten anualmente y se desconocen a la hora de emitir el *rating* del primer trimestre del año. Por ejemplo, el *rating* de un banco emitido para el primer trimestre de 2008, t , se refiere a la situación de los activos bancarios en el último año, 2007, $t-1$, ya que en ese momento se desconoce el valor de los factores referentes a 2008. ε_i son los efectos fijos asociados a cada banco. El término error, v_{it} , es un término de error estocástico. Con el fin de prevenir potenciales problemas de heteroscedasticidad y/o autocorrelación en los residuos, las ecuaciones han sido estimadas considerando errores estándar agrupados a nivel de banco.

La segunda ecuación nos permite estudiar si el posible cambio en *ratings* que haya podido tener lugar, ha sido causado por la crisis de 2008. Para ello se define una variable ficticia (EC), la cual tomará un valor igual a uno en los años posteriores al origen de la crisis (de 2008 a 2016 concretamente) y cero en los años previos. Esta variable interactúa con el resto de las variables explicativas del modelo.

$$Y_{it}^* = \beta' X_{it-1} + \beta' EC + EC \cdot \beta' X_{it-1} + \varepsilon_i + v_{it} \quad (2)$$

A través de esta interacción, se puede comprobar si se produjo un cambio en la asignación de los *ratings* debido a la crisis o no, lo cual desmentiría el supuesto de que las agencias emiten sus *ratings* teniendo en cuenta solo el largo plazo. Por lo tanto, si los coeficientes (β') que acompañan a la interacción de esta variable EC con cualquiera de las variables del modelo son estadísticamente distintos de cero, esto indicará que efectivamente, el efecto de esta variable cambió con la crisis y que, por lo tanto, las agencias ajustaron sus criterios de calificación. De ocurrir esto, se confirmaría que los *ratings* emitidos por las agencias de calificación tienen una relevancia limitada, ya que los criterios empleados para su cálculo siguen sin ser transparentes a día de hoy.

4.1.2 Random Forest

La aportación de este trabajo se centra en valorar la posible mejora que conllevaría utilizar una aproximación distinta a la tradicional, basada en los modelos ordenados descritos. Para ello, se siguen los trabajos de Ozturk et al. 2016; De Moor et al. 2017 y se obtienen predicciones de los *ratings* bancarios mediante procedimientos de

machine learning. Por lo tanto, además de utilizar el logit ordenado, también se tratan de predecir los *ratings* bancarios utilizando el modelo de *random forest*.

Los modelos de *random forest* son modelos de aprendizaje supervisados utilizados tanto en problemas de clasificación como de regresión. Para ilustrar el procedimiento de los modelos de aprendizaje, consideremos el modelo en su versión más simple, los llamados árboles de decisión (*Decision trees*). Los árboles de decisión son también modelos de aprendizaje supervisados utilizados tanto en problemas de clasificación como de regresión.

Como puede observarse en el ejemplo descrito en el **Cuadro 3**, se considera que existen 3 variables independientes, el tamaño de la entidad, el ciclo económico en el que se encuentra el país en el que opera y su negocio. Se construye un árbol de decisión en función de las distintas categorías de las variables y se observa en promedio qué *rating* es el que le corresponde a los bancos incluidos en el extremo de cada rama del árbol. En este ejemplo solo se controlan las tres variables utilizadas para la toma de decisión y en base a lo que nos diga un solo árbol. El procedimiento empleado por el *random forest* se basa en la generación de un número elevado de árboles, por ejemplo 100, con muestras distintas en cada uno de ellos, controlando el número de variables, la profundidad del árbol (número de nodos o puntos dónde se toma una decisión), la información mínima en cada nodo, etc. La generalización de estos árboles de decisión por procedimientos de aleatorización garantiza que la agregación (por ejemplo, la media) de cada uno de estos árboles tenga buenas propiedades en términos de la capacidad predictiva

Para formalizar el procedimiento, el *random forest* es un algoritmo de *machine learning* que emplea un conjunto de árboles de decisión incorrelados (basándose en la aleatoriedad) y predice una variable promediando los resultados de dichos árboles independientes (Breiman, 2001). Pero, ¿cómo asegura el bosque aleatorio que el comportamiento de cada árbol individual no esté demasiado correlacionado con el comportamiento de cualquiera de los otros árboles del modelo?, aquí es donde entra en juego la aleatoriedad en el modelo, por partida doble además, ya que no solo es aleatoria la muestra que se emplea para calibrar cada árbol, sino que también lo es el número de variables del modelo que cada árbol tiene en cuenta en dicha calibración (a diferencia del árbol de decisión simple, que tiene en cuenta todas las variables del modelo).

El primer paso para aplicar el procedimiento *del random forest* es dividir la base de datos en dos de forma aleatoria, una que se utilizará para calibrar el algoritmo y otra para contrastarlo o validarlo (en este trabajo la proporción ha sido un 75% de la muestra para entrenar y un 25% para probarlo). Posteriormente, el proceso se divide básicamente en tres partes (Trevor Hastie, 2009, p.588): en primer lugar, se obtienen distintas submuestras a partir de la muestra de entrenamiento original, el tamaño de la submuestra es siempre el mismo que el tamaño de la muestra de entrada original, pero las submuestras se toman mediante la técnica de bootstrapping (con reemplazo), por lo que la muestra que entrena cada árbol será distinta. Una vez obtenidas esas muestras de datos aleatorias, se procede a crear un árbol utilizando cada una de esas muestras y seleccionando, también de manera aleatoria, una serie de variables (x) del conjunto de variables (d) disponibles. Este segundo paso es repetido para cada uno de los N árboles aleatorios que hayamos decidido utilizar. Cada árbol obtendrá unos resultados (*rating* predicho para cada banco en el árbol) en función de los inputs

elegidos de manera aleatoria (muestra y variables elegidas) que se haya empleado en su construcción. Así pues, si tratamos de forma aislada cada uno de los resultados obtenidos por dichos árboles, no obtendremos ninguna solución clara, pero si se juntan todos los árboles teniendo en cuenta cada una de sus reglas de decisión, obtenemos una distribución de probabilidad cuya moda nos proporcionará el *rating* propuesto por el *random forest*. En resumen, el método de *random forest* consiste en calibrar un algoritmo mediante diversos árboles de decisión independientes, cada uno de ellos calibrado con una muestra y unas variables distintas. Por último, el algoritmo calibrado se pone a prueba con la muestra de prueba obtenida al inicio de todo el proceso, y se evalúa la calidad del modelo en función del número de aciertos que este obtenga.

En el algoritmo de *random forest*, como en la mayoría de los algoritmos de *machine learning*, hay una serie de parámetros cuyo valor se obtiene a partir de la muestra de datos y otros cuyo valor ha de ser elegido según el criterio del investigador. programador con el fin de obtener un mejor ajuste del modelo. Estos parámetros son conocidos como hiperparámetros y el algoritmo de Python empleado en este trabajo (*RandomForestClassifier*) permite controlar un total de 17. Algunos de estos hiperparámetros son: el número de árboles, la función utilizada para medir la calidad de las divisiones dentro de cada árbol (entropía o Gini), máxima profundidad de cada árbol (número de niveles), número de variables consideradas a la hora de dividir cada árbol (la raíz cuadrada del total de variables o \log_2 del total de variables, son algunos de los métodos posibles para calcularlo), mínimo número de datos en cada nodo antes de una división, número mínimo de datos en las hojas del árbol (último nodo que ya no se divide más), etc. El algoritmo de *random forest* fija un valor por defecto para cada uno de estos parámetros², pero en nuestro caso se ha optado por elegir el valor de 3 de ellas con el fin de mejorar el ajuste: el número de árboles, el número de variables consideradas y método para evaluar la calidad de cada división del árbol.

Con el fin de encontrar el valor óptimo de estos tres hiperparámetros, se ha empleado otro algoritmo de Python llamado *GridsearchCV*. Este algoritmo permite evaluar y seleccionar de forma sistemática los parámetros de un modelo, es decir, indicándole un modelo y estableciendo el posible valor parámetros a probar, puede evaluar el rendimiento del modelo en función de los distintos valores de los parámetros validación cruzada (*Cross Validation*). La validación cruzada consiste en dividir la muestra de entrenamiento en grupos de igual tamaño. Los grupos seleccionados, menos uno, se utilizan para calibrar el modelo, mientras que el grupo restante se utiliza para calibrarlo. Este proceso se sintetiza en el **Cuadro 4**. Este cuadro muestra un *5-fold Cross Validation*, en este caso la muestra de entrenamiento se divide en 5 partes de igual tamaño muestral. A su vez, se realizan 5 iteraciones donde en cada una ellas se ejecuta el algoritmo de *random forest*, incluyendo todo el proceso explicado anteriormente, probando todos los posibles valores de los parámetros que se hayan introducido en el *Gridsearch*, tratando siempre de maximizar la precisión del modelo. La potencia de este algoritmo es que en cada una de las iteraciones la parte entrenamiento y de test es distinta, por lo que el resultado no queda sesgado por una sola división de la muestra, como podría ocurrir si solo tuviéramos una iteración,

2

`RandomForestClassifier(n_estimators='100', criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False, class_weight=None)`

sino que se se obtiene al tener en cuenta la división los datos de 5 formas distintas. Por último, se hará una media de los resultados obtenidos en cada una de las iteraciones y se obtendrán los valores óptimos de esas tres variables, es decir, los valores que optimicen el número de aciertos del modelo.

Los valores óptimos de las tres variables obtenidos finalmente han sido los siguientes: en el caso de Fitch, el número de árboles se fija igual a 100, el número de variables consideradas en cada árbol son todas y el método para evaluar la calidad de cada división del árbol es el de entropía; para los datos de Moody's la única variable cuyo valor cambia respecto a los de Fitch es el número de árboles, fijándose en 125 en este caso y por último, en el caso de S&P, el número de árboles se fija igual a 102, el número de variables consideradas en cada árbol son todas y el método para evaluar la calidad de cada división del árbol es el de Gini.

La principal ventaja del del *random forest* y de su optimización mediante *Gridsearch* es la reducción del sesgo, de la varianza y la del impacto de “*outliers*” en nuestra predicción final (debido al uso del promedio de los resultados del conjunto de árboles, los cuáles son individuales e incorrelados, es decir, se calibrar de forma independiente y con una muestra y unas variables distintas). También se evita un posible problema de multicolinealidad al combinar las variables independientes de manera aleatoria en cada árbol, a pesar de que el número de variables en cada árbol es siempre el mismo y se obtiene mediante *grid-search*. A cambio, es un modelo poco interpretable. La única información que se obtiene es la predicción del *rating* para cada observación y la influencia de cada variable en el proceso. Además, el procedimiento del *random forest* es válido únicamente para muestras grandes y puede dar lugar a un sobreajuste del modelo, aunque se hayan empleado técnicas como la validación cruzada con el fin de evitarlo.

4.2 Método de contraste de modelos

Puesto que el objetivo del trabajo es seleccionar la técnica de determinación de los *ratings* bancarios que ofrezca mejor capacidad predictiva, es necesario disponer de algún procedimiento que permita comparar los resultados. Para ello se utiliza la matriz de confusión. Este método de comparación entre modelos de clasificación cobró popularidad entre las distintas técnicas de *machine learning* en 1998 gracias a una publicación de Kohavi and Provost (Kohavi, R. and Provost, F., 1998), desde entonces su uso se ha generalizado y sigue siendo a día de hoy una de las opciones más comunes para comparar el rendimiento entre modelos.

La matriz de confusión es una matriz de $n \times n$, en este caso de 5×5 al haber 5 valores posibles para los *issuer ratings*, donde las filas están representados por los valores observados de los *ratings* y las columnas por los valores predichos. Así pues, la diagonal representa el número de aciertos, mientras que el resto de los valores de la matriz representan, no solo los errores de predicción, sino también muestra qué valor se ha predicho para cada *rating* en vez de su valor real. El **Cuadro 5** muestra una matriz de confusión de 2×2 donde se reflejan estos conceptos de una forma más clara. De esta matriz se pueden sacar varias ratios, en nuestro caso la más relevante es la ratio de precisión, la cual determina proporción de predicciones correctas. Se determina utilizando la ecuación:

$$\text{Precisión} = \frac{A}{A+E} \quad (3)$$

El resultado de esta ratio en ambos modelos de predicción será el método de comparación empleados.

5. RESULTADOS

En este apartado se presentan los resultados obtenidos a través del logit ordenado. Los resultados se presentarán para cada una de las tres agencias de calificación. Adicionalmente, se comprobará qué modelo predice mejor los *issuer ratings*, para lo cual se utilizará la matriz de confusión, tal y como se ha descrito anteriormente.

El **Cuadro 6** muestra los coeficientes obtenidos del modelo logit ordenado correspondientes a la ecuación (1) anterior. En el cuadro se observa que en general los signos de los coeficientes son los esperados de acuerdo a la literatura previa y a la construcción de las ratios. De esta forma se puede observar como un mayor perfil de riesgo de la estructura de activos de un banco implica una mayor probabilidad de obtener la calificación más baja, ya que el coeficiente de las pérdidas por préstamos es negativo y estadísticamente significativo en las tres agencias de calificación. Además, un aumento de la ratio de capital, de su tamaño, de su rentabilidad (ROA) y una mejora del entorno económico (crecimiento del PIB) implica una mayor probabilidad de obtener la calificación más alta, en aquellas agencias donde dichos coeficientes son significativos. Al mismo tiempo, una mayor proporción del negocio centrado en actividades tradicionales aumenta también la probabilidad de obtener el *rating* más elevado. Por el contrario, el tener más gastos que ingresos y una alta proporción de préstamos sobre activos totales, conlleva a un aumento de la probabilidad de que el *rating* emitido por las agencias de calificación sea el más bajo, al menos en lo que respecta a Fitch y Standard and Poor's ya que en Moody's estos coeficientes no son significativos.

En cuanto a las reformas regulatorias, vemos como como sus efectos son muy distintos entre agencias. Así pues, en Standard and Poor's estas reformas implican un efecto diferencia positivo (mejora de los *ratings*) en la categoría de *rating* más elevada, por el contrario, en Moody's el efecto es negativo. En el caso de Fitch la reforma regulatoria de Japón está asociada a un mayor *rating*, a diferencia de la de EEUU, que produjo una reducción de los mismos.

Como se mencionó anteriormente, el signo de los coeficientes es el que cabría esperar, salvo en el caso de la ratio de liquidez, ya que muestra un signo negativo en las agencias. Esto quiere decir que las agencias penalizan la liquidez de un banco. La explicación de este signo puede encontrarse en que las agencias de calificación penalizan un exceso de liquidez como el que en la fase de recuperación han mantenido numerosas entidades financieras.

A continuación, se presentan los resultados obtenidos de la estimación de la ecuación (2), que analiza si las agencias de calificación cambiaron sus criterios de calificación a raíz de la crisis financiera (**Cuadro 7**). El coeficiente de cada variable sin interactuar con la variable ficticia (EC) indica la importancia de cada uno de ellos sobre el *rating* en el periodo precrisis. El coeficiente de la interacción de la variable EC con el resto de las variables muestra la magnitud de este efecto en los años de crisis comparado

con el periodo anterior. Se comprueba que con la crisis hubo un cambio en las políticas de calificación, ya que la importancia de cada factor en la probabilidad de obtener una calificación más alta es distinta. Tan solo se comentarán aquellos factores que hallan experimentado un cambio durante los años de crisis, es decir, aquellos factores cuyo coeficiente sea significativo en los años de la crisis ya que, en el caso de los factores cuyos coeficientes son solo significativos en la época precrisis las agencias de calificación han mantenido sus estándares al evaluar estas variables a pesar de la crisis.

Al observar los resultados del **Cuadro 7**, obtenidos al estimar la ecuación (2), se observa que el cambio en los criterios de calificación no es homogéneo entre agencias, ya que el efecto de la interacción de la variable ficticia (EC) con cada una de las variables explicativas depende de la agencia de calificación analizada. En el caso de Fitch, se observa que en la época de crisis hay un cambio en la forma de considerar los factores de tamaño, liquidez y pérdidas por préstamos. En cuanto al tamaño del banco, un mayor tamaño aumenta la probabilidad de obtener la calificación más alta en el período previo a la crisis. Sin embargo, el coeficiente de la interacción entre el tamaño y la época de crisis (EC) es negativo, es decir, el factor tamaño después de la crisis es positivo, pero más pequeño (la suma de coeficientes con y sin la interacción sigue siendo positiva) que antes de la crisis. En cuanto a los factores de liquidez y de pérdidas por préstamos, en el periodo precrisis no se tiene en cuenta su efecto, ya que los coeficientes de la regresión no son significativos, sin embargo, en la época de crisis estos factores sí son considerados a la hora de emitir un *rating*. Además, ambos factores conducen a un empeoramiento de los *ratings* emitidos, lo cual era de esperar si asumimos que las agencias tratan de penalizar un exceso de liquidez que los bancos vienen realizando desde la crisis y desde la implementación de las medidas de política monetaria ultraexpansiva llevadas a cabo por muchos bancos centrales, especialmente en Europa. Por tanto, la hipótesis que se señalaba en relación a la liquidez en relación a las estimaciones recogidas en el cuadro 1 parecen confirmarse. Es el periodo post crisis cuando las agencias penalizan los excesos de liquidez, y no tanto antes de la crisis.

En el caso de Standard and Poor's, se observa que, durante la crisis, mayor ROA reduce la probabilidad de obtener la calificación más alta (coeficiente negativo), aunque el efecto total de esta variable sigue siendo positivo. Los factores de capital y de liquidez se vuelven negativos y significativos en los años de la crisis. La explicación de estos signos aparentemente contraintuitivos puede señalar que, efectivamente las agencias de calificación pueden haber ajustado sus políticas de calificación para que efectivamente reflejen una evaluación *through the cycle*, obviando, por ejemplo, reducciones en la rentabilidad por ser consideradas transitorias. En el caso del factor ciclo el efecto es el contrario, ya que en los años de la crisis se vuelve positivo y significativo. Por último, en lo que al factor tamaño se refiere, ocurre algo similar al caso de Fitch, ya que un mayor tamaño aumenta la probabilidad de obtener la calificación más alta en el período previo a la crisis, pero el coeficiente de la interacción entre el tamaño y la época de crisis (EC) es negativo, sin embargo, en este caso el factor tamaño después de la crisis es negativo (la suma de coeficientes con y sin la interacción se vuelve negativa).

Finalmente, en el caso de Moody's, los efectos positivos de la ratio liquidez durante el periodo previo a la crisis se revierte, ya que durante el periodo de crisis esos factores tienen un coeficiente negativo y mayor que en el periodo de crisis (la suma de

coeficientes con y sin la interacción se vuelve negativa), constatándose de nuevo el comportamiento *through the cycle* de esta agencia de calificación. Se observa también como en el periodo de crisis, Moody's comienza a tener en cuenta de forma negativa la ratio de crédito, castigando así a los bancos con un mayor perfil de riesgo de la estructura de activos.

A la luz de los resultados, podría concluirse que las agencias de calificación asignan implícitamente ponderaciones heterogéneas a los diferentes factores, y que han cambiado de forma diferente su política como consecuencia de la crisis de financiera.

Logit ordenado con efectos fijos vs Random forest

Una vez descrito el cambio en los criterios de calificación de las agencias, el objetivo del trabajo es contrastar si la técnica del *random forest* permite mejorar las predicciones realizadas por la regresión logística para una misma muestra y con las mismas variables explicativas. En esta subsección se muestran los resultados comparados a través de la matriz de confusión de cada modelo (ecuación (1) frente a modelo de *random forest*). Como ya se comentó, uno de los inconvenientes de este método es la pérdida de interpretabilidad debido a que mediante el algoritmo únicamente se pueden obtener las predicciones y la ordenación de las variables explicativas en función de la influencia de cada variable explicativa en sobre el *rating*. Es por ello que no se muestran los resultados del *random forest* en un apartado como tal (como se ha hecho con el logit ordenado), sino que primero se muestra el ranking de variables según la relevancia de cada variable en el *rating* para cada agencia de acuerdo con el procedimiento del *random forest* y, posteriormente, se procede a la comparación de la precisión de ambos modelos mediante la matriz de confusión y la ratio de precisión.

La importancia que cada una de las agencias de calificación le otorga a cada una de las variables independientes de nuestro modelo a la hora de determinar los distintos *ratings* se muestra en el **Cuadro 8**. Como se puede observar, la ordenación de las variables según su importancia en el resultado del *random forest* también apoya la teoría de que el proceso de asignación de rating es heterogéneo entre CRAs. Los resultados muestran que, aunque la última posición, ocupada por Reg_EEUU, la primera posición, ocupada por Tamaño (salvo en S&P que ocupa la segunda posición) y la posición quinta, Crédito (salvo en Moody's que es cuarta), la importancia del resto de variables es bastante heterogénea.

En cuanto a la comparación de las predicciones de ambos modelos se realiza mediante el uso de matrices de confusión, como se ha descrito anteriormente. Estas matrices han sido calculadas con las predicciones del *rating* que ambos modelos realizan con la muestra de contraste. Es decir, estos resultados han sido obtenidos calibrando ambos modelos con un 75% de la muestra y contrastándolo utilizando el 25% de la muestra restante. Es decir, se ha realizado un backtesting con una muestra distinta a la empleada en la calibración para comprobar qué modelo predice mejor. Así se consigue una mejor valoración de la capacidad predictiva de los dos procedimientos, evitando el problema del sobreajuste. El sobreajuste se produciría por la estimación de modelos que ajustasen de forma muy elevada a los datos muestrales (por la selección de un número de explicativas muy elevadas o por formas funcionales muy complejas), con la consiguiente elevada capacidad predictiva dentro

de la muestra, pero que no fuesen capaces de ofrecer buenas predicciones fuera de ella.

En el **Cuadro 9**, donde se encuentran las matrices de confusión de las tres agencias de calificación, puede observarse a simple vista como el *random forest* proporciona un mayor porcentaje de acierto (número de observaciones en la diagonal principal). La ratio de precisión del *random forest* es de un 94,2% en el caso de Fitch, de un 92,3% en Standard and Poor's y de un 66,4% en Moody's, mientras que la del logit ordenado es un 90,4%, 85,8% y 51,1% respectivamente.

La **Cuadro 10** ofrece más información de la comparación de las predicciones de los dos procedimientos, mostrando el porcentaje de acierto de cada modelo para cada rating y agencia. Los tres gráficos muestran un mensaje similar. La ganancia de utilizar el modelo *random forest* se encuentra sobre todo a la hora de rededir los *ratings* para los cuales hay menos información en la muestra (siempre que la muestra sea significativa), como son el 2, 3 y en el caso de Moody's, también el 4. Esto se debe a que, como se ha explicado anteriormente en el trabajo, el modelo *random forest* analiza mucho mejor la relación entre las variables independientes y el *issuer rating*, por lo tanto, a pesar de haber *ratings* para los que la muestra de datos es menor, este modelo es capaz de predecirlos con más precisión que los modelos logit ordenados. Siendo esta otra de las ventajas del *random forest*. Por lo que, las predicciones del *random forest* son más acertadas confirmando la hipótesis postulada en este trabajo, particularmente en las categorías donde el número de observaciones es menor.

6. CONCLUSIONES

En el presente trabajo se analiza el impacto de la crisis de financiera en el comportamiento de las calificaciones de los *issuer ratings* de bancos de Estados Unidos, Europa y Japón, durante el período 2003-2016. Este período incluye años de crecimiento económico de 2003 a 2007 y un subperíodo de crisis de 2008 a 2016. El hecho de que abarque un ciclo completo nos permite contrastar la hipótesis tradicionalmente defendida por las CRAs, por la que las calificaciones se elaboran con una perspectiva a medio y largo plazo (evaluación *through the cycle*), y por lo tanto ignoran los hechos coyunturales (evaluación *point in time*) de los bancos evaluados. Sin embargo, los reguladores han llevado a cabo diferentes reformas en aras de fortalecer su supervisión y evitar conflictos de interés.

La primera parte de este trabajo se centra en corroborar si, efectivamente, ha habido un cambio en la política de fijación de *issuer ratings* por parte de las CRAs utilizando un modelo logit ordenado. Podría concluirse que las agencias de calificación asignan implícitamente ponderaciones heterogéneas a los diferentes factores, como muestra el resultado del *random forest* también, y que han cambiado de forma diferente su política como consecuencia de la crisis de financiera.

Estos cambios en la política de fijación de *ratings* en función de la situación económica global provocan desconfianza sobre las metodologías que emplean las agencias de calificación. Esto es particularmente relevante dado el papel que les otorga el regulador como “sancionador” de la calidad crediticia de las emisiones o

de la solvencia de las entidades. Podría concluirse que, aunque Moody's ha sido la entidad que menos ha cambiado sus criterios de calificación debido a la crisis, las tres entidades han penalizado a los bancos en mayor grado tras la crisis, por lo que su metodología parece ser más *point in time* que *through the cycle*.

Por otro lado, el segundo objetivo de este trabajo era comprobar si los modelos de *machine learning*, en este caso el *random forest*, son más precisos a la hora de estimar los *ratings* con modelos como el logit/probit ordenado. La conclusión de este trabajo es que, a pesar de que el modelo logit ordenado hacía predicciones con una precisión del 90, 5% en Fitch, 85,8% en S&P y un 51,1% en Moody's, el modelo de *random forest* mejora todas las predicciones, consiguiendo hasta un 15% más de precisión en el caso de Moody's, de un 5% y de un 7% más en Fitch y S&P respectivamente. En el caso de Moody's la precisión que proporciona el *random forest* respecto al logit ordenado es tan alta debido a que es la agencia con menos información de los *issuer ratings*, esto es, con menor número de observaciones. Esta falta de información afecta menos (siempre y cuando la muestra sea lo suficientemente grande) a estos modelos de *machine learning*, debido a que aprovechan mejor la información de la muestra. Este resultado es el esperable en la mayoría de los casos, ya que los modelos de *machine learning* se basan en extraer la máxima información de la base de datos de la que se disponga para hacer las predicciones. Además, el modelo de *random forest* también aporta información sobre la relevancia que las variables independientes tienen sobre la variable dependiente, volviéndose a contrastar la heterogeneidad de las variables que fundamentan el *rating* entre las distintas agencias de calificación.

Con todo ello, en futuros estudios sobre la modelización de *issuer ratings* de entidades financieras sería interesante explotar estos modelos, basados en el *machine learning*, para predecir variables financieras, y particularmente los *ratings*, ya que se podrán obtener resultados más precisos que nos ayudaran a entender y a predecir mejor el verdadero proceso de fijación de *ratings* llevado a cabo por Fitch, Moody's y Standard and Poor's, teniendo siempre en cuenta que con estos modelo perdemos interpretabilidad a cambio de obtener una mejor predicción en muchos casos.

REFERENCIAS

- Alp, A. (2013). Structural shifts in credit *rating* standards. *The Journal of Finance*, 68(6), 2435–2470.
- Altman E. and Rijken H. (2004). ‘How *rating* agencies achieve *rating* stability’. *The Journal of Finance*, 28 (11), 2679-2714.
- Altman E. and Rijken H. (2006). ‘A point in time perspective on through-the-cycle *ratings*’. *Financial Analysts*, 54-70.
- Amato, J.D., and Furfine, C.H. (2004). ‘Are credit *ratings* procyclical?’. *Journal of Banking and Finance*, 28 (11), 2641-2677.
- Baghai, Ramin P., Servaes, H., and Temayo, A. (2014) ‘Have *Rating* Agencies Become More Conservative? Implications for Capital Structure and Debt Pricing’. *The Journal of Finance*, 69 (5), 1961-2005.
- Bellotti, T., Matousek, R., and Stewart, C. (2011). ‘Are *rating* agencies’ assignments opaque?’ Evidence from international banks.’ *Expert Systems with Applications*, 38 (4), 4206-4214.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Caporale, G., Matousek, R., and Stewart, C. (2011). ‘EU banks *rating* assignments: Is there heterogeneity between new and old member countries?’. *Review of International Economics*. 19 (1), 189-206.
- D’Apice, V., Ferri, G., & Lacitignola, P. (2014). *Rating* performance and bank business models: Is there a change with the 2007–2009 Crisis? *Italian Economic Journal*, 2(3), 385–420.
- De Moor, L., Luitelab, P., Sercu, P., and Vanpée, R. (2018). ‘Subjectivity in sovereign credit *ratings*’. *Journal of Banking & Finance*, 88, 366-392
- De Young, R., & Torna, G. (2013). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation*, 22(3), 397–421.
- Fitch (2011). *Global Financial Institutions Rating Criteria*. August 2011.
- Gärtner, M., Griesbach, B., and Jung, F. (2011). ‘PIGS or lambs? The European sovereign debt crisis and the role of *rating* agencies’. *International Advances in Economic Research*, 17 (3), 288-299.
- Godlewski, C.J. (2007). Are bank *ratings* coherent with bank default probabilities in emerging market economies? *Emerging Markets Finance and Trade* 43, 5–23.
- Iannotta, G. (2006). ‘Testing for opaqueness in the European banking industry: evidence from bond credit *ratings*’. *Journal of Financial Services Research*, 30 (3), 287-309.
- Iannotta, G., Nocera, G., and Sironi, A. (2008). ‘The impact of government ownership on banks’ *ratings*: Evidence from the European banking industry’ CAREFIN Working paper, Milan: Center for Applied Research in Finance.

- Kim, J.W., Eeistroffer, H.R., Redmond, R.T. (1993). Expert systems for bond *rating*: a comparative analysis of statistical, rule-based and neural network systems. *Expert Systems*, 10 (3), 167–172.
- Kim, S.K. (2005). Predicting bond *ratings* using publicly available information. *Expert Systems with Applications*, 29 (1), 75–81.
- Kohavi, R. and Provost, F. (1998) Glossary of terms. *Machine Learning—Special Issue on Applications of Machine Learning and the Knowledge Discovery Process*. *Machine Learning*, 30, 271-274.
- Lee, Y.C. (2007). Application of support vector machines to corporate credit *rating* prediction. *Expert Systems with Applications*, 33, 67–74.
- Moody's (2007a). Bank financial strength *ratings*: Global methodology. February 2007.
- Morgan, P. (2002). '*Rating banks: risk and uncertainty in an opaque industry*'. *The American Economic Review*, 92, 874-888.
- Öğüt, H., Doğanay, M.M., Ceylan, N.B., and Aktaş, R. (2012). '*Prediction of bank financial strength ratings: The case of Turkey*'. *Economic Modelling*, 29 (3), 632-640.
- Ozturk, H., Namli, E., and Erdal, H. I. (2016). '*Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample*'. *Economic Modelling*, 54:469–478.
- Peresetsky, A. A., Karminsky, A. A., & Golovan, S. V. (2011). Probability of default models of Russian banks. *Economic Change and Restructuring*, 44(4), 297–334.
- Poon, W.P., Firth, M., and Fung, H. G. (1999). '*A multivariate analysis of the determinants of Moody's bank financial strength ratings*'. *Journal of International Financial Markets, Institutions and Money*, 9 (3), 267-283.
- Salvador, C. Pastor, J.M., and Fernández de Guevara, J. (2014). '*Impact of the subprime crisis on bank ratings: The effect of the hardening of rating policies and worsening of solvency*'. *Journal of Financial Stability*, 11, 13-31.
- Salvador, C. Pastor, J.M., and Fernández de Guevara, J. (2018). '*The adjustment of bank ratings in the financial crisis: International evidence*'. *North American Journal of Economics and Finance*, 44, 289–313
- Shen, C.H., Huang, Y.L., and Hasan, I. (2012). '*Asymmetric benchmarking in bank credit rating*'. *Journal of International Financial Markets, Institutions and Money*, 22 (1), 171-193.

CUADROS

CUADRO 1.

Distribución de la base de datos en función del país y el año. Muestra aquellos países y años para los cuáles alguna de las tres agencias de calificación posee información acerca del *issuer rating*.

COUNTRY	Fecvalor													
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
AT	-	-	-	51	51	54	51	50	52	45	46	35	16	9
BE	3	3	2	3	4	1	3	3	2	3	5	5	6	6
CZ	-	1	3	3	3	3	2	3	3	3	3	3	3	3
DE	7	11	948	1,035	1,048	1,042	1,054	1,369	1,392	1,365	1,357	1,332	1,284	1,221
DK	4	4	4	4	7	9	8	6	8	7	6	5	5	4
ES	34	37	33	30	30	37	33	19	16	10	19	17	21	20
FI	3	3	3	3	2	1	2	2	2	3	3	3	3	2
FR	27	27	27	38	32	44	52	54	67	67	75	75	70	71
GB	3	6	6	6	8	7	9	13	12	13	11	11	6	6
GR	5	6	8	9	8	5	6	6	-	-	3	-	1	3
HU	-	2	2	3	4	4	5	5	4	5	3	2	1	1
IE	-	-	2	-	-	2	1	3	2	2	2	2	1	1
IT	34	37	33	36	39	35	36	36	42	31	26	23	24	25
JP	37	40	56	61	58	53	47	49	43	46	44	46	43	42
LT	3	3	3	3	2	2	2	3	2	2	2	2	1	1
LU	5	4	6	8	6	7	9	8	8	9	10	10	9	10
NL	1		1	1	3	2	3	2	2	2	3	3	4	5
NO	3	3	3	3	3	8	9	10	12	14	15	15	9	9
PL	7	7	6	3	6	5	7	6	6	7	11	11	11	11
PT	3	4	-	-	-	-	-	8	8	6	6	5	2	1
SE	4	1	3	2	3	2	4	4	5	6	6	5	6	4
SI	2	2	4	5	5	3	6	5	5	3	1	3	4	4
US	99	103	109	113	110	109	102	95	95	82	79	72	-	1

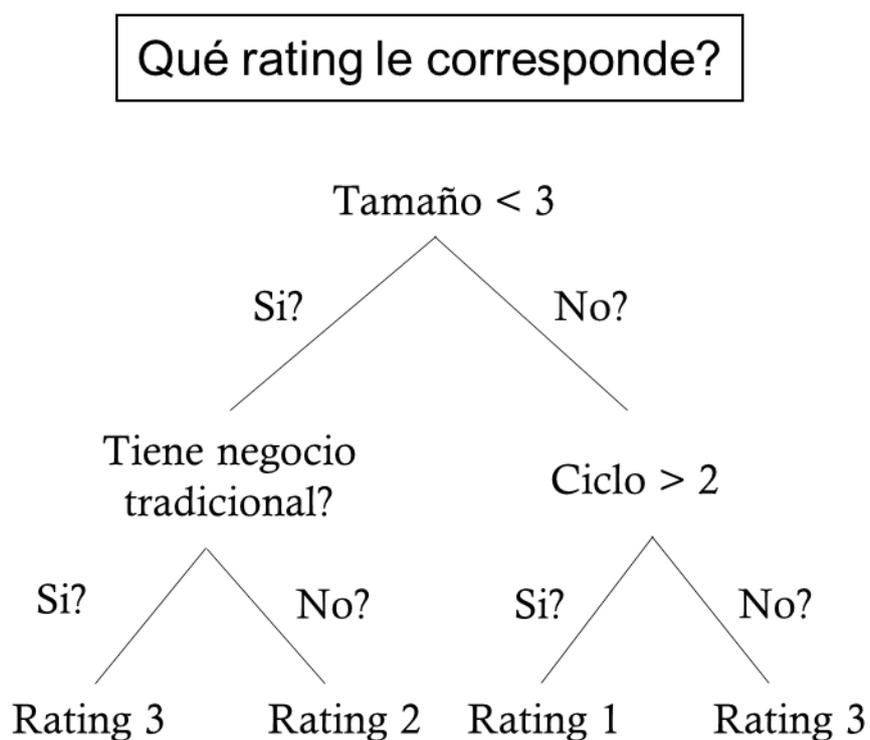
CUADRO 2.

Transformación de la escala del *issuer rating* asignada por Fitch, Standard and Poor's y Moody's a la escala numérica definida en este trabajo, escala 5. A medida que la puntuación disminuye, también lo hace la calidad crediticia y, por lo tanto, aumenta la probabilidad de quiebra de la entidad financiera. Las categorías superior e inferior están agrupadas, dado el pequeño número de observaciones que presenta la muestra.

Fitch					Standard and Poor's					Moody's				
Rating	Escala	Escala	Nº de Ratings	% de la Muestra	Rating	Escala	Escala	Nº de Ratings	% de la Muestra	Rating	Escala	Escala	Nº de Ratings	% de la Muestra
AAA	22	5	-	-	AAA	22	5	1	0,01%	Aaa	22	5	17	0,64%
AA+	21	5	10	0,06%	AA+	21	5	20	0,15%	Aa1	21	5	103	3,91%
AA	20	5	52	0,30%	AA	20	5	42	0,32%	Aa2	20	5	167	6,33%
AA-	19	5	1973	11,55%	AA-	19	5	5924	44,81%	Aa3	19	5	335	12,70%
A+	18	4	12668	74,18%	A+	18	4	5237	39,61%	A1	18	4	445	16,88%
A	17	3	908	5,32%	A	17	3	1121	8,48%	A2	17	3	603	22,87%
A-	16	2	393	2,30%	A-	16	2	340	2,57%	A3	16	2	361	13,69%
BBB+	15	1	245	1,43%	BBB+	15	1	215	1,63%	Baa1	15	1	186	7,05%
BBB	14	1	373	2,18%	BBB	14	1	138	1,04%	Baa2	14	1	121	4,59%
BBB-	13	1	198	1,16%	BBB-	13	1	62	0,47%	Baa3	13	1	101	3,83%
BB+	12	1	91	0,53%	BB+	12	1	34	0,26%	Ba1	12	1	52	1,97%
BB	11	1	54	0,32%	BB	11	1	37	0,28%	Ba2	11	1	46	1,74%
BB-	10	1	37	0,22%	BB-	10	1	26	0,20%	Ba3	10	1	43	1,63%
B+	9	1	20	0,12%	B+	9	1	13	0,10%	B1	9	1	20	0,76%
B	8	1	29	0,17%	B	8	1	3	0,02%	B2	8	1	11	0,42%
B-	7	1	19	0,11%	B-	7	1	1	0,01%	B3	7	1	8	0,30%
CCC+	6	1	-	-	CCC+	6	1	3	0,02%	Caa1	6	1	8	0,30%
CCC	5	1	6	0,04%	CCC	5	1	3	0,02%	Caa2	5	1	9	0,34%
CCC-	4	1	-	-	CCC-	4	1	-	-	Caa3	4	1	1	0,04%
CC	3	1	-	-	CC	3	1	1	0,01%	Ca	3	1	-	-
C	2	1	2	0,01%	C	2	1	-	-	C	2	1	-	-
D	1	1	-	-	D	1	1	-	-	D	1	1	-	-
Ratings Totales			17078	100%	Ratings Totales			13221	100%	Ratings Totales			2637	100%

CUADRO 3.

Ejemplo de un árbol de decisión cuyo objetivo es tratar de predecir el *rating* de una entidad financiera. En este caso, esta decisión se tomará en base tres variables independientes. Como resultado se obtienen cuatro nodos terminales, cada uno de ellos con un *rating* determinado, el resultado final dependerá del camino escogido en función del valor de las variables independientes.



CUADRO 4.

Ejemplo del procedimiento *K-fold Cross Validation*. Este ejemplo concreto pertenece a un *5-fold Cross Validation*, ya que la muestra en cada iteración se divide en 5 partes y además hay 5 iteraciones. Las *K-fold Cross Validation* más comunes suelen ser los de 3, 5 y 10.



CUADRO 5.

Ejemplo de una matriz de confusión de 2x2.

		Valor Estimado	
		<i>Rating 1</i>	<i>Rating 2</i>
Valor Observado	<i>Rating 1</i>	A	E
	<i>Rating 2</i>	E	A

Donde

- Acierto (A): Estos son casos en los que el valor observado coincide con la predicción.
- Error (E): Estos, por el contrario, son casos en los que el valor observado no coincide con la predicción.

CUADRO 6.

Resultados de la estimación del *issuer rating* por agencia mediante la ecuación (1) empleando una regresión logística ordenada. En la regresión se han empleado variables *dummy* para aislar los efectos por banco y año. Los errores estándar han sido agrupados por banco.

	Fitch	Moody's	Standard and Poor's
ROA	61.396*** (14.68)	76.747*** (15.93)	114.662*** (19.86)
Capital	9.883** (3.63)	3.162 (4.17)	-2.798 (5.48)
Eficiencia	-0.012** (0.00)	0.004 (0.00)	-0.009** (0.00)
Crédito	-2.702*** (0.70)	-2.368 (1.59)	-3.867*** (0.94)
Negocio Tradicional	0.012*** (0.00)	-0.002 (0.00)	0.015*** (0.00)
Liquidez	-2.314* (0.98)	-0.144 (1.28)	-4.128*** (1.10)
Tamaño	0.175* (0.08)	0.908*** (0.15)	-0.131 (0.10)
Pérdidas activos dudosos	-16.095 (8.77)	-51.964*** (14.45)	-33.165*** (5.31)
Ciclo	0.204*** (0.05)	0.193*** (0.06)	0.652*** (0.09)
Regulación Japón	4.645*** (0.46)	-0.603 (0.36)	2.787*** (0.65)
Regulación Europa	-0.225 (0.17)	-2.346*** (0.24)	1.419*** (0.31)
Regulación EEUU	-6.106*** (0.31)	-0.477** (0.17)	0.516*** (0.11)
cut1	-2.621* (1.22)	10.136*** (2.92)	-6.251 (7.11)
cut2	-1.050 (1.21)	12.110*** (2.93)	-4.371 (7.11)
cut3	2.110 (1.18)	14.829*** (2.94)	1.087 (7.09)
cut4	12.793*** (1.21)	16.842*** (2.94)	7.164 (7.09)

***Significancia al 1%, ** Significancia al 5% y * Significancia al 10%.

CUADRO 7.

Resultados de la estimación del *issuer rating* por agencia mediante la ecuación (2) empleando una regresión logística ordenada. En la regresión se han empleado variables *dummy* para aislar los efectos por banco y año. Los errores estándar han sido agrupados por banco.

	Fitch	Moody's	Standard and Poor's
ROA	74.362** (25.23)	68.125* (30.50)	121.542*** (21.00)
Capital	14.415* (6.42)	11.152 (5.82)	7.865 (6.26)
Eficiencia	0.324 (0.31)	2320 (1.28)	0.200 (0.72)
crédito	-1.225 (1.21)	1.772 (1.80)	-1966 (1.06)
Negocio Tradicional	-0.023 (0.03)	-0.203 (0.11)	-0.013 (0.06)
Liquidez	2.867 (1.85)	5.376** (1.86)	1526 (1.21)
Tamaño	0.826*** (0.13)	0.997*** (0.18)	0.660*** (0.11)
Pérdidas activos dudosos	-0.098 (4.77)	-5983 (26.69)	-14914 (7.89)
Ciclo	0.151 (0.09)	0.323* (0.13)	0.017 (0.12)
Regulación Japón	4.107*** (0.69)	-2.160*** (0.45)	-3.875** (1.29)
Regulación Europa	-0.333 (0.19)	-2.463*** (0.25)	2.342*** (0.39)
Regulación EEUU	-5.984*** (0.29)	-0.467** (0.18)	0.784*** (0.14)
EC	14.983*** (1.60)	10.705** (4.07)	24.978*** (1.66)
EC_ROA	-33.734 (31.26)	33328 (35.18)	-89.324** (32.88)
EC_Capital	-1.673 (7.29)	-12366 (6.71)	-22.019** (6.90)
EC_Eficiencia	0.281 (0.58)	-2314 (1.28)	-0.192 (0.72)
EC_Crédito	-1.765 (1.19)	-7.471*** (2.03)	-0.990 (0.95)
EC_Negocio Tradicional	-0.550 (0.47)	0.198 (0.11)	0.010 (0.06)
EC_Liquidez	-7.420*** (2.11)	-9.491*** (2.35)	-6.940*** (1.40)
EC_Tamaño	-0.807*** (0.12)	-0.060 (0.17)	-1.075*** (0.08)
EC_Pérdidas activos dudosos	-14.789* (6.41)	-48436 (32.43)	-2436 (9.91)
EC_Ciclo	0.032 (0.10)	-0.176 (0.12)	0.928*** (0.17)
cut1	8.862*** (1.72)	16.478*** (3.78)	7707 (7.16)
cut2	10.584*** (1.72)	18.547*** (3.80)	9760 (7.15)
cut3	13.967*** (1.76)	21.376*** (3.80)	15.410* (7.14)
cut4	24.984*** (1.83)	23.465*** (3.79)	22.177** (7.16)

***Significancia al 1%, ** Significancia al 5% y * Significancia al 10%.

CUADRO 8.

Importancia que el método de *random forest* otorga a las variables independientes, divididas por agencia y excluyendo la importancia de las variables dummies año y país.

Rankig de variables según la importancia en la explicación del <i>rating</i>	Fitch	S&P	Moody's
1 ^a	Tamaño	Reg_Japón	Tamaño
2 ^a	ROA	Tamaño	Capital
3 ^a	Tradinc	Capital	Liquidez
4 ^a	Capital	Tradinc	Crédito
5 ^a	Crédito	Crédito	Pérdidas..
6 ^a	Liquidez	Pérdidas..	ROA
7 ^a	Eficiencia	Liquidez	Tradinc
8 ^a	Pérdidas..	ROA	Eficiencia
9 ^a	Ciclo	Eficiencia	Ciclo
10 ^a	Reg_Euro	Ciclo	Reg_Japón
11 ^a	Reg_Japón	Reg_Euro	Reg_Japón
12 ^a	Reg_EEUU	Reg_EEUU	Reg_EEUU

CUADRO 9.

Resultados de las distintas matrices de confusión. A la izquierda se encuentran las matrices obtenidas por el logit ordenado y a la derecha las obtenidas por el método de *random forest*. Los resultados se muestran para cada una de las agencias de calificación: en la primera fila se encuentran los resultados de Fitch, en la segunda los de Standard and Poor's y en la tercera fila los de Moody's.

RATING Fitch	1	2	3	4	5
1	211	4	7	9	9
2	58	10	10	7	6
3	57	8	127	13	7
4	18	12	6	2981	72
5	18	16	6	27	211

RATING Fitch	1	2	3	4	5
1	211	11	8	6	4
2	36	42	9	2	2
3	26	10	166	5	5
4	10	4	6	3030	39
5	9	4	3	31	231

RATING S&P	1	2	3	4	5
1	85	7	5	7	8
2	27	14	9	17	11
3	16	7	37	81	99
4	8	1	31	1254	29
5	6	6	20	33	1191

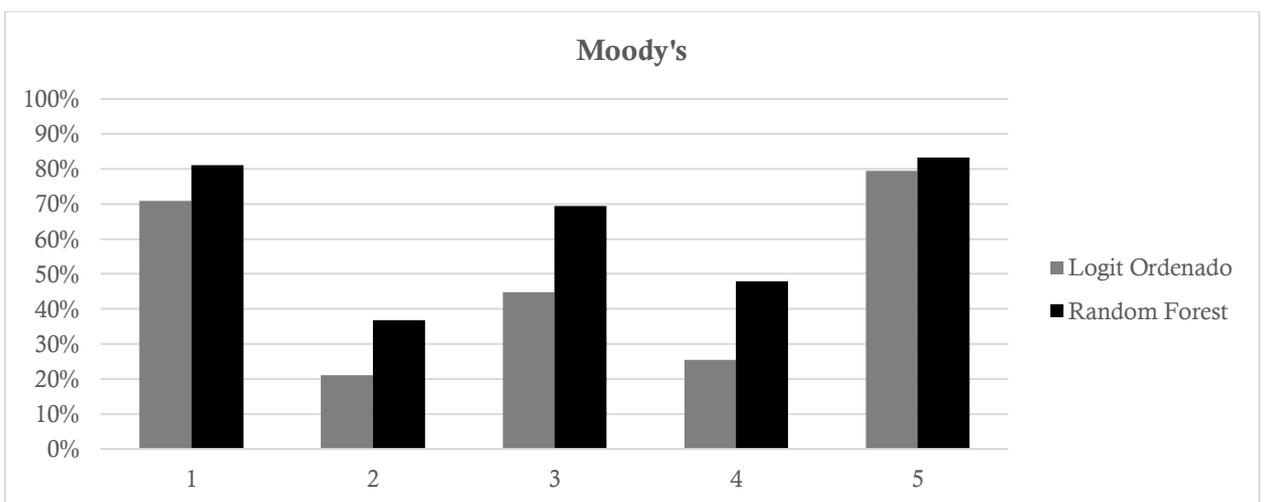
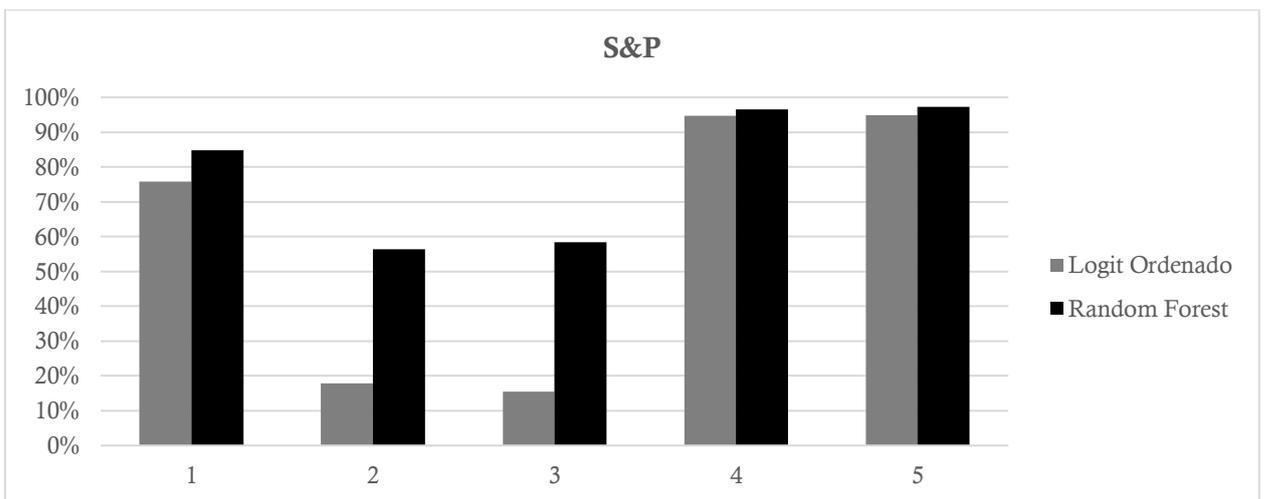
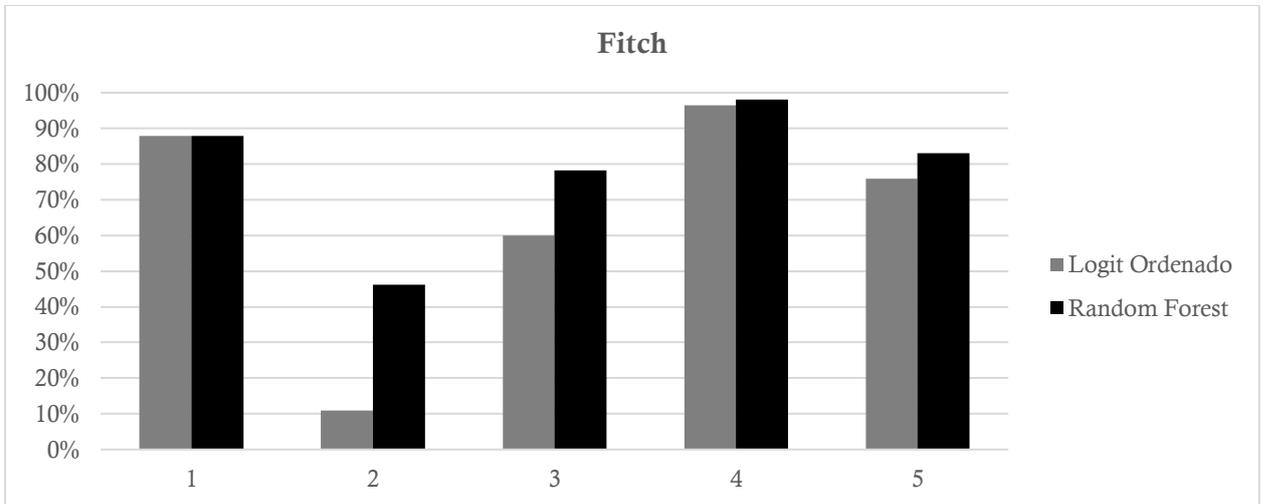
RATING S&P	1	2	3	4	5
1	95	13	3	0	1
2	14	44	17	3	0
3	6	12	140	33	49
4	2	0	23	1277	21
5	4	7	17	7	1221

RATING Moody's	1	2	3	4	5
1	90	10	15	7	5
2	28	20	17	19	11
3	27	16	69	12	30
4	12	4	7	24	47
5	6	3	7	11	104

RATING Moody's	1	2	3	4	5
1	103	6	11	6	1
2	27	35	20	8	5
3	14	8	107	14	11
4	4	6	14	45	25
5	1	2	6	13	109

CUADRO 10.

Los gráficos muestran el porcentaje de aciertos de ambos métodos de predicción, dividido por rating y por agencia de calificación.



ANEXO

Conjunto de ratios empleadas para definir las variables independientes del modelo. También se muestra tanto su abreviatura, como el número de asiento en Bankscope.

$$\text{Rentabilidad}(ROA) = \frac{p2105}{p2025} = \frac{\text{PreTaxProfits}}{\text{Totalassets}}$$

$$\text{Tamaño}(sz) = \log(p2025) = \log \text{of TotalAssets}$$

$$\text{Capital}(cap) = \frac{p2055}{p2025} = \frac{\text{Equitylevel}}{\text{Totalassets}}$$

$$\text{Liquidez}(te) = \frac{p2075}{p2025} = \frac{\text{Liquidassets}}{\text{Totalassets}}$$

$$\text{Eficiencia}(eo) = \frac{p2090}{p2080 + p2085} = \frac{\text{Overheads}}{\text{NetInterestRevenue} + \text{OtherOperatingIncome}}$$

$$\text{Credito}(cred) = \frac{p2000}{p2025} = \frac{\text{Loans}}{\text{Totalassets}}$$

$$\text{Pérdidasporprestamos} = \frac{p2095}{p2000} = \frac{\text{LoanLossProvisions}}{\text{Loans}}$$

$$\text{Tradinc} = \frac{p2080}{p2080 + p2085} = \frac{\text{NetInterestRevenue}}{\text{NetInterestRevenue} + \text{OtherOperatingIncome}}$$

$$\text{Ciclo} = \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}} \times 100$$