



Facultad de Ciencias Económicas y Empresariales

APLICACIÓN DEL MACHINE LEARNING AL FACTOR INVESTING EN RENTA FIJA CORPORATIVA

Autor: 201701546

Director: María Coronado Vaca

MADRID | Junio, 2022

Índice

I. INTRODUCCIÓN

- 1.1. Proposición general y contexto
- 1.2. Objetivos de la investigación
- 1.3. Metodología
- 1.4. Estructura del Trabajo

II. ORIGEN

III. CONCEPTO Y CARACTERÍSTICAS

- 3.1. Definición y características
- 3.2. *Factor investing* en renta fija
- 3.3. Los factores en renta fija corporativa: literatura previa

IV. APLICACIÓN DEL MACHINE LEARNING EN *FACTOR INVESTING*

- 4.1. Introducción
- 4.2. *Penalized Regressions*
- 4.3. *Neural Networks*
- 4.4. *Support Vector Machines*
- 4.5. *Random Forests*

V. EJEMPLO DE APLICACIÓN DE RANDOM FOREST

VI. CONCLUSIONES

I. INTRODUCCIÓN

1.1. PROPÓSITO GENERAL Y CONTEXTO DEL TRABAJO

En este trabajo se estudia la posibilidad de aplicación de técnicas de machine learning a la gestión de carteras, y más concretamente al factor investing en renta fija corporativa.

La gestión de carteras por factores, más conocida por su nomenclatura inglesa “*Factor Investing*”, es un método de inversión que ha experimentado una importante expansión en los últimos años, aumentando tanto en su popularidad como el interés académico dedicado al tema. Se trata de una estrategia de inversión basada en invertir en función de determinadas fuentes de rentabilidad, o factores (uno o varios), para obtener una rentabilidad ajustada a riesgo por encima de la del mercado.

En el contexto de la renta variable, existen numerosos estudios que demuestran como históricamente factores como *value* (compañías baratas desde una perspectiva fundamental), *momentum* (compañías que logran mejores retornos que el mercado en el corto plazo), o *size* (compañías con capitalización bursátil más baja) han conseguido una rentabilidad ajustada a riesgo mayor que sus mercados de referencia. Además, en el ámbito de la renta variable, es un método ampliamente extendido y empleado. No obstante, al contrario de lo que sucede en los mercados de *equity*, el *factor investing* en renta fija está mucho menos difundido, y ha recibido una reducida atención académica. En efecto, los estudios en el marco de los mercados de deuda (bien corporativa, bien pública o soberana) son muy limitados. Ello se debe, entre otras razones, a la mayor complejidad de los mercados de renta fija.

Por otro lado, el *machine learning* y la inteligencia artificial se han consagrado como los *drivers* más importantes de la innovación en prácticamente todos los ámbitos, y también lo están siendo en el mundo financiero y la gestión de carteras. Así pues, a través de aprendizaje automático podemos descubrir patrones ocultos más allá de los referentes al sentido económico; podemos adaptarnos rápidamente a los entornos cambiantes de los mercados financieros; y ponemos analizar y procesar volúmenes de datos fuera del alcance de la mente humana.

1.2. OBJETIVO DE LA INVESTIGACIÓN

Por todo ello, se va a explorar en este trabajo la posibilidad de implementar técnicas de factor investing en renta fija corporativa a través de machine learning para demostrar la mayor rentabilidad ajustada a riesgo de los factores.

1.3. METODOLOGÍA

En primer lugar, para alcanzar el mencionado objetivo, se han analizado y estudiado diversos trabajos publicados por los autores de referencia en el ámbito del factor investing. Tomando como referencia estos trabajos, se ha profundizado en los diversos factores aplicados a renta fija corporativa, mediante la lectura de los artículos disponibles en este ámbito.

A continuación, se ha explorado la posibilidad de aplicación de distintos algoritmos a la inversión por factores, tomando como principal punto de apoyo el libro publicado por Guillaume Coqueret y Tony Guida “*Machine Learning for Factor Investing*”.

Finalmente, se ha tratado de poner en práctica los anteriormente mencionado a través de la implementación propia de técnicas de *machine learning* para demostrar los mayores rendimientos de los factores en el conjunto de datos seleccionado.

1.4. ESTRUCTURA DEL TRABAJO

El presente escrito está dividido en tres partes: en primer lugar, un análisis de los orígenes y otras nociones básicas del factor investing y la gestión de carteras, que faciliten la comprensión del conjunto del trabajo, En segundo lugar, se analizan detalladamente los distintos factores y un análisis más pormenorizado del *factor investing*, especificando su encaje en el ámbito de la renta fija. Finalmente, se estudia la aplicabilidad de distintos algoritmos a la inversión por factores, y se conduce una suerte de experimento propio para el demostrar la mayor rentabilidad de los factores.

II. ORIGEN

Antes de comenzar este estudio sobre el factor investing, su aplicación actual y aplicabilidad al mercado de renta fija corporativa, debemos remontarnos al origen del mismo y el contexto histórico y teórico en el que surge.

De forma previa a la aparición del factor investing y otras formas de gestión de carteras como los fondos cotizados, las carteras de inversión se gestionaban bien mediante técnicas de gestión activa, bien mediante métodos gestión pasiva. La gestión de carteras activa es la técnica mediante la que se trata de superar la rentabilidad de un índice de referencia a través de la selección minuciosa de los activos que presentan mayores rentabilidades¹. La gestión activa conlleva un elevado coste de gestión, y en renta fija consiste en seguir una combinación de: estrategias de expectativas de tasas de interés, estrategias de curva de rendimientos, estrategias de diferenciales de rendimiento...²

Por otro lado, la gestión pasiva trata de conseguir la misma rentabilidad que un índice de referencia o mercado manteniendo en la cartera una muestra proporcional de los valores que componen dicho índice y sus pesos. Esta clase de gestión acarrea un menor coste y se realiza

¹ AFI. (2020, abril). Factor Investing & ETF: Gestión por Factores para una inversión más eficiente.

² de Elejabeitia, P. G. (2018). GESTIÓN DE CARTERAS. GESTIÓN ACTIVA VS. GESTIÓN PASIVA. UNIVERSIDAD PONTIFICIA COMILLAS .

en renta fija bien a través de técnicas de indexación de bonos, o con estrategias basadas en pasivos.²

Como punto intermedio entre gestión pasiva y activa aparece la inversión basada en factores. Su origen se remonta a las décadas de 1960 y 1970, cuando surgen la Teoría Moderna de Carteras de Markowitz y el Capital Asset Pricing Model (CAPM), fundados en la eficiencia de los mercados.

Markowitz definió una cartera eficiente como aquella que ofrece el mínimo riesgo dado un nivel de rentabilidad esperada, y maximiza la rentabilidad esperada dado un determinado nivel de riesgo. Así, existe una frontera de carteras eficientes conformada por las carteras que maximizan la rentabilidad en cada nivel de riesgo³. En estos términos, son eficientes las carteras que maximizan el Ratio de Sharpe:

$$\text{Ratio Sharpe} = \frac{(R_c - R_f)}{\sigma(R_c)}$$

R_c = rentabilidad de la cartera

R_f = rentabilidad del activo libre de riesgo

$\sigma(R_c)$ = desviación típica de la cartera

Éste calcula el exceso de rentabilidad de la cartera sobre el activo libre de riesgo, relativo al riesgo asumido medido por la volatilidad de la cartera.

Por otro lado, según el CAPM existe una correlación entre la rentabilidad de un activo o cartera, y su riesgo medido por la beta:

$$\beta = \frac{\sigma_{cm}}{\sigma^2_m}$$

σ_{cm} = covarianza del activo y el mercado

σ^2_m = varianza del mercado

³ AFI. (2020, abril). Factor Investing & ETF: Gestión por Factores para una inversión más eficiente.

Ésta es una medida de correlación de un valor o una cartera de inversión con respecto a los movimientos del mercado en general, y se calcula como la covarianza entre las rentabilidades del mercado y las de un activo, dividida entre la varianza del mercado³.

Así, la rentabilidad de un activo dado se mide por el CAPM como:

$$R_c = R_f + \beta * (R_m - R_f)$$

R_c = rentabilidad de la cartera

R_f = rentabilidad del activo libre de riesgo

R_m = rentabilidad esperada del mercado

En este contexto, y partiendo de la asunción de que los mercados son eficientes y los inversores son racionales, la beta es el único factor de riesgo que es recompensado por el mercado y hay una relación lineal positiva entre volatilidad y riesgo. Esto significa que una mayor rentabilidad conlleva necesariamente un mayor riesgo (mayor volatilidad de la cartera).

En el CAPM encontramos el primer modelo de un solo factor, que utiliza la covarianza entre el mercado y los rendimientos de las acciones individuales (beta) como único factor de rentabilidad. El razonamiento es sencillo: una acción se mueve en función de su sensibilidad frente al mercado, donde el mercado actúa como un conjunto de múltiples factores. Esto es comúnmente denominado como riesgo sistemático. Este es el único riesgo por el que un inversor debe ser compensado, porque todos los demás riesgos idiosincrásicos pueden ser eliminados mediante la diversificación de las inversiones.⁴

Tanto las hipótesis de la Teoría Moderna de Carteras de Markowitz, como las del CAPM se expandieron rápidamente y siguen presentes hoy en día. No obstante, cuando los datos disponibles comenzaron a ser cada vez mayores y más precisos, surgieron académicos que comenzaron a investigar y cuestionar las conclusiones del CAPM. Este mayor volumen de datos, junto con la mayor capacidad de procesamiento informático, dieron pie a una mayor actividad de investigación empírica, de donde surgieron los primeros estudios, como los artículos de Basu (1997), o Banz & Renganum (1981), que demostraron que empresas con

⁴ van de Wetering, G. J. (2017). Using Factor Investing as Investment Strategy for Corporate Bonds in Europe and the United States. Tilburg School of Economics and Management .

menor PER, y las de menor tamaño en términos de capitalización bursátil, obtienen rentabilidades ajustadas a riesgo por encima del mercado. Así pues, esta tendencia investigadora continuó, y se publicaron multitud de artículos que demuestran la existencia de factores distintos a la beta, especialmente en el ámbito de la renta variable (como el Black-Scholes Model, de Black, Jensen y Scholes).

Sin embargo, la publicación más relevante para el tema que nos concierne, y que da origen a lo que hoy en día conocemos como factor investing, fue el estudio de Fama & French (1993), en el que aplicaron un modelo de tres factores (mercado, tamaño y value) que mejoraba las predicciones de rentabilidad del CAPM, con una menor volatilidad y riesgo.

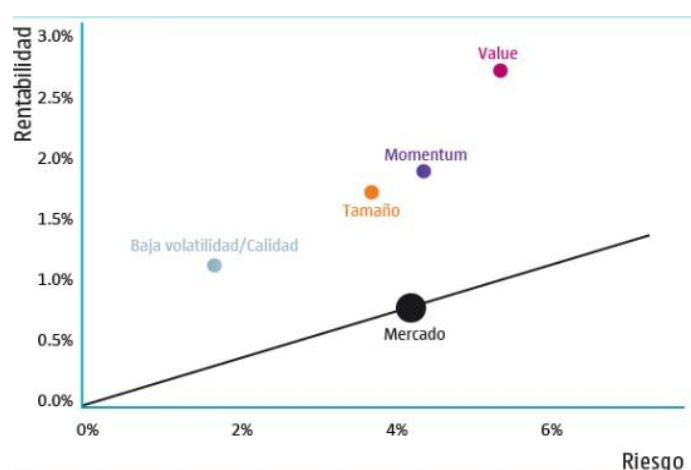
A partir de la publicación de este artículo se expandió rápidamente el estudio e investigación del *factor investing*, pero su auge real y efectiva implementación en carteras de inversión a escala global se dio tras la crisis económica de 2008. A ello contribuyeron mayoritariamente dos fenómenos: dado que se demostró que la gestión activa no proporcionaba protección suficiente contra caídas generalizadas, los inversores comenzaron a diversificar sus carteras en función de los factores de riesgo sistémicos subyacentes (es decir, en base a los factores que generan la rentabilidad de los activos); además, los inversores comenzaron a buscar vehículos de inversión alternativos de menor coste a través de los que lograr rentabilidades similares a las obtenidas por los gestores activos.⁵ Por estas razones, al enmarcarse el *factor investing* como una técnica de gestión de carteras intermedia con menor coste que la gestión activa, y otorgar capacidad a los inversores de exponerse a factores específicos, tuvo gran éxito. Además, otro motivo por el que se ha expandido rápidamente el *factor investing* como técnica de inversión ha sido el crecimiento y popularidad de los fondos cotizados (*ETFs*) y la búsqueda de formas de obtener rentabilidades superiores sin condicionar la volatilidad de las carteras.

En este contexto, pueden surgirnos preguntas como ¿por qué *factor investing*?, ¿cuáles son sus objetivos, y qué ventajas competitivas ofrece frente a otros métodos o técnicas de inversión? Mediante esta técnica de inversión se trata de lograr principalmente un objetivo: mejorar la rentabilidad ajustada a riesgo de una cartera dada. Es decir, se trata de obtener mayor Ratio de Sharpe que el mercado a través de la mayor exposición a determinados valores o títulos que

⁵ Soe Hong Xie CFA, A. (Ed.). (2016). Factor-Based Investing in Fixed Income: A Case Study of the U.S. Investment-Grade Corporate Bond Market.

gozan de una característica determinada, que a su vez es la que motiva o justifica la prima de rentabilidad.

Como se ha mencionado superficialmente, tras la crisis financiera se produjo un cambio drástico, desde estrategias de inversión activa a estrategias de inversión pasiva. No obstante, pese a evitar el riesgo de exposición a decisiones desafortunadas de los gestores activos, la inversión pasiva somete al inversor a un mayor riesgo de arbitraje, y genera unos rendimientos relativamente bajos ajustados a riesgo debido el coste asociado a la gestión. En este contexto, el factor investing trata de lograr mayores rendimientos ajustados al riesgo a la par que mantiene los costes relativamente bajos. Esto se puede observar en el siguiente gráfico, que compara los resultados históricos de factores respecto del mercado de deuda de grado inversión en USD:



Fuente: Robeco, Bloomberg. Deuda investment grade en USD, entre enero de 1994 y diciembre de 2017.

Este gráfico muestra un estudio llevado a cabo por Robeco, en el que se mide la rentabilidad del crédito en relación con la deuda pública de duración equivalente. Así, observamos como la rentabilidad a largo plazo ajustada a riesgo es mayor si los inversores se centran en cada uno (o en una combinación) de estos factores.⁶

El segundo principal motivo para acudir a técnicas de factor investing es la reducción de riesgos a la baja. Esto se ha convertido en una de las mayores preocupaciones de los inversores en los

⁶ Los Essentials del Factor Investing. (s/f). Robeco.

últimos años, y mediante la inversión en factores de baja volatilidad o riesgo, se puede reducir la volatilidad de una cartera sin renunciar a expectativas de rentabilidad.

Además, un inversor puede mejorar la volatilidad de una cartera a través de la diversificación del portafolio mediante factor investing de forma más robusta, que a través del tradicional reparto del mismo entre diferentes categorías de activos y zonas geográficas. Esto ha sido demostrado por diversos estudios, como el realizado en 2012 por Antti Ilmanen y Jared Kizer.⁷

Asimismo, estas estrategias de inversión abren la puerta a que los inversores puedan obtener exposición específica a un factor. Esto es un fenómeno que está teniendo mucho éxito en la actualidad, al proporcionar una oportunidad para incrementar la exposición a objetivos ambientales, sostenibles y de gobierno (*ESG*).

III. CONCEPTO Y CARACTERÍSTICAS

3.1. DEFINICIÓN Y CARACTERÍSTICAS

Una vez detallado el origen histórico y teórico del *factor investing*, procedemos a examinar su definición y características.

Podemos definir los factores como aquellas características de una empresa que impulsan su rentabilidad.⁸ Se trata de los pilares esenciales de rentabilidad de los activos financieros que se pueden emplear para comparar una empresa con sus similares (*peers*). Así pues, son elementos cuantificables y medibles de las empresas. Por tanto, el *factor investing* es una técnica de inversión basada en la asociación de la rentabilidad de activos financieros a estas características medibles, cuantificables y predeterminadas⁹.

El objetivo principal de los modelos de factores es comprender los *drivers* de los precios de los activos, los motivos que impulsan el incremento de su rentabilidad. Así pues, la lógica de la inversión en factores es que los resultados financieros de las empresas dependen de ciertas

⁷ Ilmanen, Antti, Ronen Israel, Tobias J Moskowitz, Ashwin K Thapar, and Franklin Wang. 2019. "Factor Premia and Factor Timing: A Century of Evidence." SSRN Working Paper 3400998

⁸ "Foundations of Factor Investing"; Jennifer Bender, Remy Briand, Dimitris Melas, and Raman Aylur Subramanian; December 30, 2013; MSCI.

⁹ AFI. (2020, abril). Factor Investing & ETF: Gestión por Factores para una inversión más eficiente.

características, ya sean latentes e inobservables o relacionados con características intrínsecas (como las ratios contables, por ejemplo). Tal y como plantea Cochrane (2011), la primera cuestión que ha de plantearse un inversor es ¿cuáles son los componentes que proporcionan realmente información imparcial sobre los rendimientos de un activo? La respuesta a esta pregunta es el primer paso para comprender y poder predecir de forma precisa la rentabilidad de los activos.

En el plano teórico, los modelos lineales de factores pueden verse como excepciones, o también como casos especiales en el marco de la Teoría del Arbitraje (*Arbitrage Pricing Theory*) de Ross (1976), pues asumen que los rendimientos de un activo n se pueden modelar como la combinación lineal de los factores subyacentes f_k :

$$r_{t,n} = \alpha_n + \sum_{k=1}^K \beta_{n,k} f_{t,k} + \epsilon_{t,n},$$

$r_{t,n}$ = retornos en el tiempo t del activo n

K = atributos del activo

Como podemos observar, esta ecuación lineal difiere de la presentada previamente del CAPM, donde el único factor subyacente es el de mercado (que se asume que es plenamente eficiente). Sin embargo, en la teoría de la inversión por factores los rendimientos vienen determinados por cuestiones que pueden ser ajenas o anómalas al mercado. A diferencia del CAPM, el *factor investing* parte de la asunción de que existen diversos factores que también explican y justifican las primas de rentabilidad de activos financieros por encima de la del mercado. En el marco de la Teoría del Arbitraje, se establece que los individuos se aprovechan de las ineficiencias (anomalías) del mercado para generar rentabilidad. Por ello, cuando hablamos de factores también podemos referirnos a ellos como ‘anomalías’.

Entrando en el detalle de los factores y sus características, a nivel general existen dos grandes categorías de factores: los macroeconómicos, que son fuentes de riesgo sistemáticas que afectan a toda la economía, como pueden ser los tipos de interés o la inflación; y factores de estilo, que son características, como el *value* y el *momentum*, que justifican el rendimiento superior de determinados valores.¹⁰

¹⁰ AFI. (2020, abril). Factor Investing & ETF: Gestión por Factores para una inversión más eficiente.

En los mercados de renta variable el *factor investing* es un enfoque ya asentado y consolidado en el que se han empleado desde hace tiempo distintos factores de riesgo para explicar las diferencias entre los rendimientos de distintas carteras. En este ámbito, entre los factores más comúnmente aceptados y extendidos tradicionalmente, encontramos: *value*, *volatility*, *momentum*, *quality* y *size*. Así pues, el *factor investing* defiende que las compañías cuyo perfil encaje dentro de una de estas categorías presentarán en el futuro mejor evolución y rendimiento ajustado a riesgo que el resto del mercado.

Por tanto, los factores más comúnmente empleados y aceptados en los mercados de *equity*, son:

- *Value*: se refiere a las acciones de empresas cuyo precio por acción esté infravalorado en términos fundamentales.
- *Volatility*: que engloba los títulos cuya volatilidad es menor, teniendo simultáneamente una beta estable y próxima a 1 (correlación con el mercado).
- *Momentum*: referente a empresas cuya cotización haya evolucionado favorablemente en el pasado inmediato, presentando una tendencia alcista favorable.
- *Quality*: que concierne los valores con unos mayores márgenes y ROE.
- *Size*: que alude a empresas que tienen una menor capitalización bursátil.

Si bien es ampliamente aceptada la existencia de una prima de rentabilidad sobre el mercado de los factores, sigue habiendo debate en torno a la medida en que ésta es consecuencia de la exposición a un mayor riesgo asumido por los inversores por la mayor exposición a los factores, o si es resultado del aprovechamiento de las ineficiencias del mercado:

En línea con la exposición a un mayor riesgo, la prima tiene una explicación diferenciada para cada uno de los factores. Se plasman a continuación algunos ejemplos:

- En cuanto a las empresas con acciones infravaloradas (*value factor*), existe un mayor riesgo a la baja por su alta sensibilidad a perturbaciones económicas o a épocas de crisis, lo que justifica que los inversores exijan mayores rendimientos en compensación del mayor riesgo asumido.
- Se explica la exigencia de un mayor retorno de los títulos con menor volatilidad y beta (*low volatility*) debido a que los inversores pueden acudir al denominado “*lottery effect*”

(efecto lotería), aprovechando la posibilidad, menor pero existente, de obtener mayores beneficios de títulos con mayor volatilidad.

- Respecto de las compañías con menor tamaño (*size factor*), se observa que tienden a adolecer de una menor liquidez, mayores dificultades crediticias y de financiación, y riesgos de caídas que se ven compensadas por mayores retornos. Además, las empresas de menor tamaño suelen reportar menos habitual y extensamente, dificultando la capacidad de los inversores de acceder a información, lo que también hace que requieran una mayor recompensa para invertir.

Asimismo, la mayor rentabilidad ajustada a riesgo de los factores encuentra también justificación, como hemos mencionado brevemente, en la existencia de ineficiencias de los mercados financieros. En primer lugar, algunas características estructurales del mercado crediticio pueden llevar a ineficiencias, como se observa en los diferentes objetivos de distintos entes del mercado ajenos a la maximización de retornos (como los bancos centrales). La intervención de estos organismos en el mercado causa ineficiencias porque no toman decisiones encaminadas a la optimización de retornos, sino que persiguen metas distintas.¹¹

Además, se observan en el mercado sesgos de comportamiento entre los inversores, que condicionan su comportamiento de inversión por diferentes conductas humanas que les conducen a ignorar o emplear mal información a la hora de invertir, causando deficiencias en el mercado. Ejemplos de ello serían el exceso de confianza o el conocido como '*representative bias*' por el que los inversores tienden a confiar más en las anécdotas que en las estadísticas a la hora de tomar decisiones de inversión.¹¹

Por último, las restricciones de los inversores son otra causa plausible de las mencionadas ineficiencias. Hay muchas carteras que sufren limitaciones importantes, como las de apalancamiento, o las restricciones al *short-selling* (venta en corto) que limitan la capacidad de ciertos inversores para arbitrar las ineficiencias del mercado. Todas ellas surgen bien de normativa, bien de la disponibilidad de activos financieros, pero causan ineficiencias que pueden justificar la exigencia de una prima al estar expuesto a distintos factores.¹¹

¹¹ de Carvalho, T. H. Z. A. I. H. O. L. R. L. (2021). Out-performing corporate bonds indices with factor investing. Forthcoming in Bankers Markets & Investors.

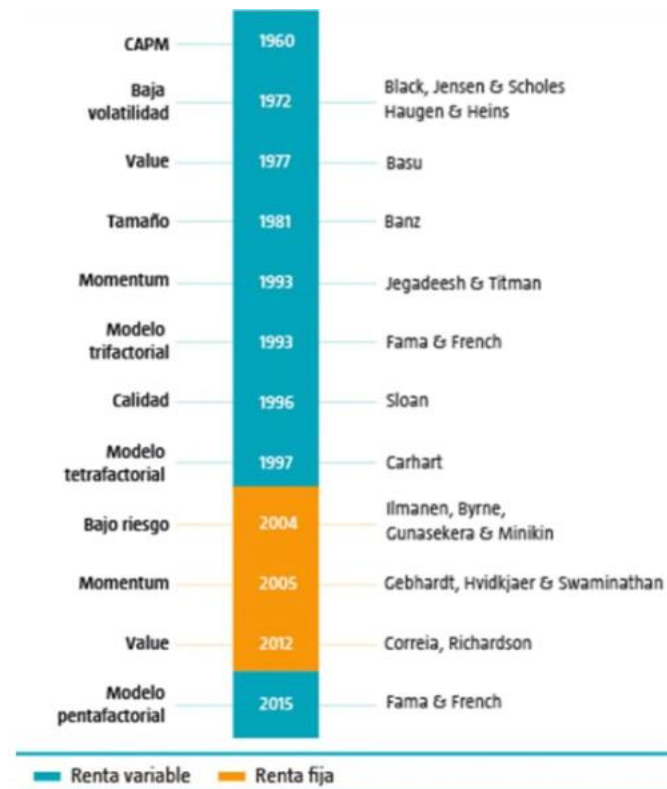
Sin embargo, también se ha de mencionar que son numerosos los retos que encontramos a la hora de obtener mayores rendimientos ajustados a riesgo a través del factor investing. En primer lugar, es destacable que ciertos autores han mencionado en sus publicaciones la existencia de una tendencia a la reducción de las ineficiencias o al ajuste de las mismas a partir de su publicación o anuncio al mercado. Así, hay autores como McLean y Pontiff (2016)¹² o Shanaev y Ghimire (2020)¹³ que demuestran que las mencionadas irregularidades reducen su intensidad una vez son publicadas, puesto que el mercado trata de descompensar el desequilibrio localizado por los inversores. Cuando se descubren los factores y se exponen al mercado, se produce una migración inversora a los valores representativos de los factores, lo que hace que, por el funcionamiento del mercado, la oferta y la demanda, se eliminen o reduzcan las anomalías.

3.2. FACTOR INVESTING EN RENTA FIJA

Si bien la aplicación del factor investing está muy extendida en renta variable, con numerosos estudios que han demostrado los mayores rendimientos de las carteras construidas en base a factores y profundizan en las causas la rentabilidad de cada uno de ellos, no se ha dirigido especial atención, ni se ha estudiado de forma tan extensiva cómo pueden implementarse en gestión de carteras de renta fija factores como *value*, *quality*, *momentum* o *risk*, para obtener rentabilidades mayores a las de los índices de referencia en este ámbito. La investigación académica en la inversión por factores en el ámbito de la renta variable comenzó en la década de 1970 y la se ha extendido de forma notable, siendo un modelo de inversión en renta variable muy extendido hoy en día. Sin embargo, la investigación de los factores de renta fija sigue siendo un campo de estudio en desarrollo. Este desequilibrio en la investigación ha dado lugar probablemente a una menor familiaridad de los inversores con estas técnicas en los mercados de deuda. A modo ilustrativo, véase esta línea temporal con los estudios de investigación académica con mayor trascendencia en el ámbito del *factor investing*:

¹² McLean, R David, and Jeffrey Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance* 71 (1): 5–32.

¹³ Shanaev, Savva, and Binam Ghimire. 2020. "Efficient Scholars: Academic Attention and the Disappearance of Anomalies." *European Journal of Finance* Forthcoming: 1–27.



Fuente: Robeco

En esta línea temporal elaborada por Robeco, se muestra claramente como es mucho más reciente y escaso el *research* académico en el ámbito de mercados de deuda. No obstante, y ante la tendencia de las empresas a lo largo del último ciclo económico de acudir a los mercados de deuda para financiarse, creo que el *factor investing* supone una oportunidad para, con base en la experiencia en bolsa, impulsar las rentabilidades de las carteras de renta fija. Ante tipos de interés en mínimos históricos, y las políticas macroeconómicas expansivas que han provocado una caída en el coste de la deuda, las empresas han tenido la oportunidad de financiarse a un menor coste acudiendo a los mercados de deuda. Si bien es cierto que podemos encontrarnos en un período de transición en términos de política monetaria, me ha extrañado a lo largo de mi estudio que en un mercado creciente que en términos reales alcanzó su máximo histórico a finales de 2019 con la deuda corporativa global en 13.5 billones de dólares según datos de la OCDE, no se haya estudiado más a fondo la aplicación de estas técnicas tan extendidas en los mercados de *equity*.¹⁴

¹⁴ de Carvalho, T. H. Z. A. I. H. O. L. R. L. (2021). Out-performing corporate bonds indices with factor investing. Forthcoming in Bankers Markets & Investors.

No obstante, es bien cierto que las diferencias entre *factor investing* en renta variable y renta fija corporativa son notables. Ésta no es tan intuitiva como aquélla, y hay una serie de diferencias que debemos tener en cuenta.

En primer lugar, debemos considerar que una misma compañía puede (y suele) emitir múltiples bonos con distintos vencimientos y especificaciones, y, por tanto, con distinto nivel de riesgo. Además, el nivel de riesgo de cada bono cambia con el paso del tiempo porque disminuye el tiempo hasta la fecha de vencimiento¹⁴.

Para continuar, y a diferencia de la inversión en bolsa, donde cabe que los inversores puedan esperar obtener rendimientos extraordinarios por la posibilidad de que la capitalización bursátil de las empresas crezca de forma potencialmente ilimitada, los rendimientos obtenidos por los inversores en bonos emitidos por esas mismas empresas no se verán tan afectados ni retribuidos por el aumento de capitalización de las acciones. Es más, una empresa cuyo *market cap* se dispara puede simplemente aumentar su deuda emitiendo más bonos, sin ningún impacto directo palpable en los rendimientos obtenidos por los tenedores de bonos.¹⁴ Así pues, se refleja en el ámbito del *factor investing* la tradicional discrepancia entre los retornos que pueden obtener los inversores en renta variable y los inversores en renta fija.

Asimismo, los mercados de deuda corporativa suelen tener una menor liquidez y muchos títulos son más difíciles de comprar y vender puesto que muchos inversores los mantienen hasta su vencimiento. Ello se ve reforzado por el hecho de que muchos bonos cotizan en mercados libres opacos y fragmentados, que hacen difícil su visibilidad.

Todas estas cuestiones deben ser atajadas a la hora de diseñar estrategias de *factor investing* en renta fija corporativa. Una dificultad adicional surge respecto del *factor investing* en renta variable por el hecho de que en la construcción de modelos de factores en renta fija, se han de vincular los precios de todos los bonos individuales emitidos por una misma empresa a las características fundamentales de ésta de forma robusta. Además, encontramos dificultades extra por el hecho de que un bono puede cotizar bajo el mismo nombre hasta su vencimiento en un mercado de deuda dado, pese a que su empresa emisora cambie de nombre y características (por ejemplo, en caso de fusión con otra empresa), no así en el caso de la renta variable.¹⁴

3.3. LOS FACTORES EN RENTA FIJA CORPORATIVA: LITERATURA PREVIA

Por las razones recogidas anteriormente, la inversión por factores en renta fija difiere notablemente de su homóloga en renta variable. Es más, los factores en renta fija no son los mismos, y que deben definirse de forma distinta. Así pues, procedemos en este punto a explorar y definir los factores en el ámbito de los mercados de deuda, resaltando algunas de las diferencias más notorias respecto de los empleados en renta variable.

Tal y como hemos adelantado previamente, existen motivos por los que los factores generan una prima de rentabilidad por encima del mercado, también así en los mercados de deuda. Por ello, antes de adentrarnos en profundidad en cada uno de los factores, debemos mencionar brevemente que los factores rentabilidad del mercado de renta fija corporativa están basados en principios económicamente demostrados y rigurosos, que históricamente han producido una prima de rentabilidad. Así, al igual que los factores en renta variable, existen motivos que justifican la rentabilidad superior de determinados bonos respecto de sus índices de referencia, entre los que cabe citar: la asunción por parte del inversor de riesgos que son recompensados con una mayor rentabilidad en el largo plazo; los impedimentos estructurales que producen las normas o restricciones del mercado, y que hacen que ciertas áreas del mercado sean de difícil acceso para cierto perfil inversor y crea mayores oportunidades para otro subconjunto del mercado; y finalmente los sesgos de comportamiento de los inversores, que no son perfectamente racionales, ni tienen información completa, lo cual crea oportunidades para explotar tendencias existentes.

Al igual que en el ámbito del *equity*, estos se pueden dividir en dos categorías: los macroeconómicos y los de estilo, y la rentabilidad de cualquier cartera de bonos estará afectada por una combinación de ambos tipos de factores. En este estudio, nos adentraremos en los de factores de estilo, pues son los que afectarán a una cartera específica en función de las características de los bonos que la conforman.

Así pues, múltiples estudios académicos han demostrado la existencia de una serie de factores persistentes en el mercado de renta fija corporativa. Antes de profundizar en cada uno, debemos

llevar a cabo una revisión bibliográfica de la principal literatura académica sobre rendimientos y varianzas de los bonos; para luego entrar al detalle y conclusiones relacionadas con los siguientes factores: *Size, Value, Momentum, Carry, Low Risk, y Liquidity*.

En lo que respecta el rendimiento de un bono, una inversión en deuda corporativa funciona de la misma forma que cualquier otra inversión: las inversiones más arriesgadas o con mayor probabilidad de fracaso requieren un mayor rendimiento (coste de capital más alto), puesto que el inversor tiene menor certeza de recibir su inversión de vuelta. A su vez, dentro de la renta fija, los dos principales componentes del riesgo sistemático son: el riesgo de impago, causado por la posibilidad de que la empresa emisora no reembolse la deuda o financiación que ha tomado en préstamo; y el riesgo de duración, que es el riesgo que se da cuando un bono tiene un vencimiento o término a muy largo plazo, puesto que, a mayor tiempo, mayor será la probabilidad de impago, y de que suban los tipos de interés causando una disminución en el valor del bono. Los bonos soberanos están relativamente más expuestos a estos dos factores, mientras que los bonos corporativos tienen algunos componentes adicionales de riesgo y rendimiento (que suelen ser específicos de cada empresa). En este campo, debemos destacar que ya Fama & French (1993) demostraron que, mientras que los bonos soberanos a corto y largo plazo presentaban coeficientes de determinación (R^2 , medida estadística que mide el *fit* de un modelo con sus datos reales) de 0.79 y 0.87 respectivamente, los R^2 de los bonos corporativos de baja y alta calificación crediticia eran de 0.49 y 0.98 respectivamente.¹⁵ Los coeficientes de determinación indican qué proporción de la varianza de la variable dependiente viene determinada por la varianza de las variables independientes. De esta forma Fama & French (1993) demostraron que, cuanto más seguros son los bonos (mayor calificación crediticia – grado de inversión vs. grado de especulación), más puede explicarse su varianza sólo por la duración y riesgo de crédito.

En otra investigación en este campo, Litterman y Scheinkman (1991)¹⁶ descubrieron que la mayor parte de la varianza de los rendimientos de los bonos estatales puede justificarse atendiendo únicamente a tres factores: el nivel, la inclinación, y la curvatura; o lo que es lo mismo: el tipo de interés, los movimientos en las curvas de rendimientos, o cambios en la

¹⁵ van de Wetering, G. J. (2017). Using Factor Investing as Investment Strategy for Corporate Bonds in Europe and the United States. Tilburg School of Economics and Management .

¹⁶ Litterman, R.B. and Scheinkman, J. (1991) Common Factors Affecting Bond Returns. *Journal of Fixed Income*, 1, 54-61.

curvatura de la curva de rendimientos. En su artículo “*Common Factors Affecting Bond Returns*” los autores demostraban que el 96% de la varianza de los bonos cupón cero (y el 97% de los bonos del Tesoro no exigibles) podía explicarse atendiendo a los tres factores anteriores.

Cuando los rendimientos de los bonos devienen más complejos, como se produce en el caso de los bonos corporativos, el riesgo de liquidez es otro componente del riesgo total que puede emplearse para justificar parte de los rendimientos. Amihud y Mendelson (1986)¹⁷ se encuentran entre los primeros que documentaron el efecto de la liquidez sobre los precios de los activos y sus rendimientos futuros. En su publicación “*Asset pricing and the bid-ask spread*” observaron una relación positiva entre el diferencial de oferta y demanda (la liquidez del activo), y sus rendimientos futuros. A su vez, añadieron que los inversores con horizontes temporales a largo plazo tienden a poseer valores más ilíquidos, lo que provoca un “efecto clientela”. Esta prima de liquidez fue cuantificada más adelante por Warga (1992)¹⁸, y resultó ser una prima de rendimiento anual del 0,55% cuando comparamos una cartera de bonos públicos líquidos y de duración constante, con sus equivalentes ilíquidos.

De esta manera, y ante el amplio diferencial entre la oferta y la demanda de bonos corporativos en comparación con las acciones, la (i)liquidez de los bonos se debe tener en cuenta para la realización de un análisis empírico consistente. Esto fue confirmado por Lesmond, Long y Wei (2007)¹⁹ que, tras la realización de un análisis de 4000 bonos corporativos, demuestran que cerca del 50% de la variación de los diferenciales de rendimiento puede ser explicado acudiendo a la (i)liquidez de los bonos (más que atendiendo a la calificación crediticia de los mismos). A su vez, lo demostrado en este ensayo fue reforzado por el estudio “*Liquidity of Corporate Bonds*”, realizado en 2008 por Bao, Pan & Wang²⁰.

Una vez desglosado y detallado el rendimiento de los bonos, debemos profundizar en los factores concretos previamente adelantados. La mayoría de estos factores fueron demostrados

¹⁷ Yakov Amihud and Haim Mendelson, (1986), *Asset pricing and the bid-ask spread*, *Journal of Financial Economics*, 17, (2), 223-249

¹⁸ Arthur Warga, (1992), *Bond Returns, Liquidity, and Missing Data*, *Journal of Financial and Quantitative Analysis*, 27, (4), 605-617

¹⁹ Long Chen; David A. Lesmond and Jason Wei, (2007), *Corporate Yield Spreads and Bond Liquidity*, *Journal of Finance*, 62, (1), 119-149

²⁰ Bao, Jack and Pan, Jun and Wang, Jiang, *Liquidity of Corporate Bonds* (July 9, 2008).

primero en los mercados de *equity*, y luego se trató de extrapolarlos al ámbito del *factor investing*.

a. Size

La relación entre los rendimientos y el valor de mercado, el factor *size*, fue examinada por primera vez por Banz (1981), quien observó que las acciones de empresas con menor capitalización bursátil obtenían mayores rentabilidades que los demás. Siguiendo sus investigaciones, Fama y French (1993) calcularon la rentabilidad de este factor de tamaño como '*small-minus-big*' (SMB), medido como el rendimiento de una cartera del 10% de los valores más pequeños menos el rendimiento del 10% de las acciones más grandes.

En general, se espera que las pequeñas empresas sean tengan mayor riesgo de inversión que las grandes empresas, porque una recesión económica puede afectar más a las pequeñas empresas que a las grandes. en comparación con las grandes empresas. Además, las grandes empresas pueden beneficiarse de las economías de escala y de la diversificación empresarial, lo que les da más fuerza competitiva, las hace más rentables y disminuye la dependencia del ciclo económico. Además, se puede argumentar que los medios de comunicación, *brokers* y analistas, prestan menor atención a las pequeñas empresas, lo que dificulta la accesibilidad a toda la información que los inversores necesitan para tomar una decisión de inversión adecuada y educada.

Por lo tanto, los inversores exigen una prima de tamaño a la tasa de descuento de las pequeñas empresas para compensar el riesgo adicional en el que incurren por invertir en ellas frente a una de mayor tamaño, lo que reduce la valoración. Aunque el valor de mercado de una acción no influye directamente en su rentabilidad, es una buena aproximación al factor *size*, que puede interpretarse como un conjunto de factores múltiples que conducen a una mejor rentabilidad ajustada al riesgo.

El mismo razonamiento podría aplicarse a los bonos corporativos. Si las acciones de una empresa cotizan con descuento debido a su riesgo, entonces tendría sentido que los bonos de las empresas más pequeñas también se negocien con descuento o, al menos, paguen un tipo de

interés más alto. Bektic, Wenzler, Wegener, Schiereck & Spielmann (2016)²¹ investigan si los cinco factores de Fama y French (2015) conducen a un rendimiento superior en los bonos corporativos también. Utilizan el valor de mercado de los fondos propios del emisor de los bonos como *proxy* del tamaño, y encuentran un rendimiento levemente superior cuando se invierte en una cartera que contiene bonos del 10% de los emisores más pequeños. Dado que la mayoría de las empresas suelen emitir varios bonos diferentes, y en tanto que el tamaño de una empresa puede estar vinculado a un determinado nivel de riesgo, el importe total de la deuda podría ser una mejor medida para el factor *size* (tamaño) en el mercado de deuda corporativa.

Houweling y van Zundert (2016)²² afirman ser los primeros académicos que documentan de forma efectiva y trascendente un efecto del tamaño en el mercado de bonos corporativos, utilizando una métrica diferente a la capitalización del mercado. En su lugar, calculan la suma de las ponderaciones del índice de todos los bonos emitidos por la misma empresa en cada mes, para utilizarla como *proxy* del valor de mercado de la deuda en circulación.

Para la construcción de los portfolios en atención al factor tamaño clasifican la magnitud de los bonos de cada empresa para cada mes, siendo las mejores carteras las formadas por los bonos de las compañías que dentro del 10% de las más pequeñas, y las carteras menos rentables las formadas por bonos emitidos por el 10% de las compañías de mayor tamaño.

Así pues, se emplean la deuda total emitida por una misma compañía, y no los bonos individualizados, argumentando que el factor tamaño clásico del mercado de renta variable está relacionado con la capitalización de la compañía. Otro factor importante a tener en cuenta aquí, es la probabilidad de que exista una prima de liquidez en estos bonos que son emitidos por empresas más pequeñas y que reciben una pequeña ponderación. La prima de tamaño podría, por tanto, captar parte de la prima de liquidez, que es uno de los posibles factores asociados al tamaño.

²¹ Bektic, Demir and Wenzler, Josef-Stefan and Wegener, Michael and Schiereck, Dirk and Spielmann, Timo, Extending Fama-French Factors to Corporate Bond Markets (February 1, 2017). *Journal of Portfolio Management*, 45 (3), 141-158, 2019, <https://jpm.pm-research.com/content/45/3/141>

²² Patrick Houweling, J. van Z. (2017). Patrick Houweling [PDF] from efmaefm.org Factor Investing in the Corporate Bond Market. *Financial Analysts Journal*, 73(2), 1–16.

Sus carteras *long-only*, que contienen el 10% de los bonos más pequeños generan un significativamente más alta de 0,42 (Mercado: 0,12) para los bonos con grado de inversión, y una Sharpe de 0,64 (Mercado: 0,23) para los bonos de alto rendimiento, lo que indica la presencia de una prima de tamaño de tamaño en los bonos corporativos.

b. *Low risk*

En renta variable, Frizzing y Pedersen (2014) llegaron a la conclusión de que betas bajas producen mayores retornos que las betas altas cuando establecieron su factor de apuesta contra beta, el *low risk*. Para los rendimientos de las acciones estadounidenses e internacionales, demuestran que las acciones de beta baja producen un exceso de rentabilidad mensual significativamente mayor, así como alfas y ratios de Sharpe mejores en comparación con las acciones de beta alta.

Mencionan que las acciones de beta alta son más atractivas para los inversores económicamente restringidos que buscan rendimientos altos, lo que lleva a una gran demanda de acciones de alta beta por su parte, lo que, a su vez, aumenta los precios de los valores de beta alto y reduce los rendimientos futuros esperados. La misma idea podría aplicarse a los bonos corporativos, cuando se clasifican en función de la calificación crediticia o de los vencimientos.

En el ámbito de la renta fija corporativa Leote de Carvalho, Dugnonle, Lu y Moulin (2014), fueron pioneros en demostrar empíricamente que los factores de bajo riesgo producen una prima de rentabilidad por encima del mercado. En su publicación, exponen evidencia de que los bonos con menor riesgo no sólo tienen una menor beta relativa al índice de referencia ponderado, sino que además ofrecen mayores rendimientos ajustados a riesgo. Su análisis demostraba este factor independientemente para bonos emitidos en dólares, euros, libras, o yenes, tanto en renta fija corporativa, como en muchas otras áreas de los mercados de deuda.²³

²³ de Carvalho Patrick Dugnonle Xiao Lu and Pierre Moulin, R. L. (2014). Low-Risk Anomalies in Global Fixed Income: Evidence from Major Broad Markets. *The Journal of Fixed Income* Spring 2014, 23(4), 51–70.

Israel et al. (2016) crearon las llamadas carteras defensivas, en las que el riesgo del emisor de un bono se mide utilizando más de una variable, a saber: (1) el apalancamiento de mercado, medido como una deuda neta sobre el valor de la empresa; (2) la duración efectiva; (3) la rentabilidad, medida como el beneficio bruto sobre los activos.

Clasificaron los bonos en 5 quintiles en función de la métrica defensiva. El quintil defensivo genera una rentabilidad un 2,7% mayor que el quintil menos defensivo.

Así pues, probaron que una cartera larga-corta utilizando esta estrategia habría generado un ratio de Sharpe de 1,11, que es (relativamente alto comparado con los resultados de otros estudios, lo cual puede deberse al tipo de gestión activa que requiere esta determinada estrategia).

c. *Value*

Por otro lado, encontramos el factor *value*, que probablemente es el más conocido tanto en mercados de equity, como de deuda. Los defensores más famosos de la inversión en valor son Benjamin Graham y Warren Buffet.

Básicamente, el principio de la inversión en valor consiste en comprar acciones con un valor de mercado que es inferior a su valor fundamental intrínseco. Así pues, se trata de un factor que trata de explotar la infravaloración del mercado de un determinado activo. De esta manera, se parte de la base de que el activo en el futuro alcanzará su valor fundamental, para lo cual debe obtener mejores rendimientos que el mercado.

Fama y French (1993) fueron de los primeros académicos en demostrar la existencia de una prima de valor que justifica la rentabilidad de las acciones de tipo “*value*” a lo largo plazo. La prima de valor la denominan *high-minus-low* (HML) y se mide como la diferencia entre los rendimientos de las acciones con un elevado *book-to-market ratio* (*price to book*, o valor de mercado-frente a valor contable) frente a las que tienen un *book-to-market* bajo.

Fama y French demostraron que un inversor puede aprovecharse comprando el 10% de los valores con mayor BTM, y poniéndose en corto en el 10% de los valores con el BTM más bajo. Esto elimina inmediatamente una gran parte del riesgo de mercado, lo que se traduce en una

mejor rentabilidad ajustada al riesgo. El factor valor de las acciones puede medirse de muchas maneras, pero siempre que se mantenga el fundamento del *value investing*, conducirá a un rendimiento superior independientemente de su forma de medición.

Esta hipótesis de inversión es fácilmente extrapolable al entorno de la renta fija corporativa. El *value investing* en renta fija conlleva invertir en bonos que cotizan por debajo de su valor fundamental, igual que en renta variable respecto de las acciones. Por lo tanto, la dificultad se encuentra en determinar una buena métrica para encontrar dichos bonos corporativos infravalorados. L'Hoir y Boulabiel (2010) señalan que una infravaloración o sobrevaloración puede derivarse de un amplio margen entre el diferencial de crédito real y el diferencial de crédito "justo" que debería reflejar el riesgo del bono. Para predecir el diferencial de crédito justo, realizaron la siguiente regresión:

$$sp_{ijkl,t} = \beta_k ts_{pk,t} + \beta_j ts_{pj,t} + \beta_l ts_{pl,t} + \sum \gamma_m t\chi_{m,t} + \varepsilon_{ijkl,t},$$

$Sp_{k,j,l}$ = diferencial de todos los bonos con una calificación crediticia k, del sector j, con un orden de prelación l

$\beta, \sum, \varepsilon_j$ = conjunto de variables

Esta regresión, mide un conjunto de variables específicas que incluyen una medida de rentabilidad de la empresa (ROA) y la volatilidad de las acciones de la empresa. Los residuos se tratan como señal de valor, donde un residuo positivo indica que los bonos ofrecen una rentabilidad superior a la que predice el modelo, lo que indica que el bono podría estar infravalorado. Lo contrario ocurre con los residuos negativos, que podrían indicar una sobrevaloración.

En esta línea de investigación continúan Correia et al. (2012)²⁴, quienes consiguieron confirmar la existencia de un diferencial entre la probabilidad de impago pronosticada por varios modelos y la probabilidad de impago implícita en los diferenciales de crédito. El principal determinante de los diferenciales de crédito de las empresas es la probabilidad física de impago. La diferencia entre ambas se relaciona positivamente con los rendimientos futuros. Así pues, los autores comparan los modelos existentes basados en la contabilidad y en el mercado para prever el impago. De esta forma, los autores descubren que dichos diferenciales de crédito se ven

²⁴ Correia, M., Richardson, S. & Tuna, İ. Value investing in credit markets. Rev Account Stud 17, 572–609 (2012). <https://doi.org/10.1007/s11142-012-9191>

afectados por tasas de impago previstas con un retraso significativo. Esta conclusión evidencia la existencia del factor valor en los mercados de renta fija corporativa.²⁵

Houweling y van Zundert (2017) van más allá, y añaden la variación a tres meses del diferencial de crédito como variable explicativa, porque se ha comprobado que los diferenciales de crédito se adaptan ya tres meses antes del ajuste real de la calificación crediticia. Sus carteras *long-only*, basadas en el valor generan un ratio de Sharpe de 0,27 (Mercado: 0,12) en el mercado de grado de inversión, y un ratio de Sharpe de 0,49 (mercado: 0,23) en el segmento de grado de especulación, lo que es significativamente superior a la cartera de mercado.²⁶

Finalmente, con casi la misma metodología de estimación que Howeling y van Zundert, pero con un enfoque *long-short* en lugar de una cartera de coste cero, Israel, Palhares y Richardson (2016) obtienen un ratio de Sharpe de 1,75, que es una relación riesgo-rendimiento increíblemente atractiva en comparación con los resultados de otros estudios.

d. Momentum

Asimismo, se han realizado diversas investigaciones y estudios acerca la estrategia de inversión por *momentum*, que consiste en comprar los valores “ganadores” recientemente y vender los “perdedores”. Los primeros en identificar su existencia fueron Jegadeesh y Titman (1993)²⁷. En su pionero estudio, comprueban qué ocurre en términos de rentabilidad al mantener una cartera de los valores más rentables de los últimos 6 meses y reequilibrarla cada mes. Esta estrategia devuelve una rentabilidad superior a la del mercado de más del 12% entre 1965 y 1989.

²⁵ van de Wetering, G. J. (2017). Using Factor Investing as Investment Strategy for Corporate Bonds in Europe and the United States. Tilburg School of Economics and Management

²⁶ Patrick Houweling, J. van Z. (2017). Patrick Houweling [PDF] from efmaefm.org Factor Investing in the Corporate Bond Market. Financial Analysts Journal, 73(2), 1–16.

²⁷ Narasimhan Jegadeesh, S. T. (2017). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>

Tras la publicación de este estudio, muchos académicos han conseguido demostrar la existencia del factor impulso utilizando otras definiciones, mercados y períodos de tiempo.

Cabe destacar que Carhart (1997)²⁸ encontró que los factores Mercado, *value*, *size* y *momentum* explican casi toda la persistencia en los rendimientos de los fondos de inversión en renta variable, especialmente de los que obtienen mejores resultados.

Un argumento contra la estrategia de inversión por impulso es que el rebalanceo cada mes requiere hacer múltiples operaciones, lo que puede dar lugar a unos costes de transacción excesivamente elevados. Pero como los costes de transacción están disminuyendo con el paso del tiempo, este argumento pierde relevancia.

A primera vista, el mercado de bonos corporativos parece menos adecuado para esta estrategia. Con tantos tipos de interés, vencimientos y calificaciones diferentes, un rendimiento superior o inferior se ve impulsado por muchos factores dominantes, como el tipo libre de riesgo y la prima de impago.

Pospisil y Zhang (2010) analizaron el posible efecto de impulso en el mercado de bonos corporativos estadounidense. Los resultados fueron mixtos, ya que la estrategia de impulso resultó ser rentable en los bonos de grado de especulación, pero los resultados fueron negativos e insignificantes para los bonos de grado de inversión. Por tanto, concluyen que el activo libre de riesgo determina una gran parte de los rendimientos, y que parece haber un mayor efecto de inversión en el mercado de bonos corporativos.

Khang y King (2004)²⁹ también encontraron un efecto de inversión en los bonos corporativos. Argumentan que el hecho de que los bonos corporativos se negocien en un mercado repleto de intermediarios es posiblemente la mejor explicación del efecto de inversión en los bonos corporativos. Estos intermediarios agrupan grandes cantidades de bonos, y los venden cuando los precios de los bonos suben lo suficiente, provocando un descenso de los precios, y creando este efecto de inversión de rentabilidad.

²⁸ Carhart, M. M. (2012). On Persistence in Mutual Fund Performance. <https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>

²⁹ King, K.-H. (2004). Return reversals in the bond market: Evidence and causes. *Journal of Banking & Finance*, 28(3), 569–593.

Otros expertos buscan otras definiciones para el factor de impulso. L'Hoir y Boulhabel (2010) examinan el impulso del flujo de caja en lugar de los rendimientos y comparan el impulso del flujo de caja con los rendimientos de las acciones durante el mismo período.

Por su parte, Kan (1996) y Gebhardt, Hvidkjaer y Swaminathan (2005) documentan un impulso de la renta variable en los mercados de bonos corporativos, demostrando que los rendimientos pasados de las acciones predicen con precisión los rendimientos futuros de los bonos corporativos del mismo emisor.

Finalmente, Jostova (2013) muestra que, en una muestra de bonos de high yield y de grado de inversión, los ganadores del pasado tienden a superar a los perdedores perdedores, sobre la base de un período de estimación de 6 meses y un período de tenencia posterior. Una selección del 10% de los bonos con mejores resultados, en términos de rendimiento por encima de un bono del tesoro de duración equivalente, parece ser el método más apropiado para crear carteras de impulso a largo plazo, pero un enfoque de quintiles no es infrecuente y es favorable cuando el tamaño de la muestra es limitado.³⁰

IV. APLICACIÓN DEL MACHINE LEARNING EN *FACTOR INVESTING*

4.1.INTRODUCCIÓN

En un plano general, existen multitud de razones por las que el *machine learning* puede impulsar y mejorar las estrategias de factor investing. En primer lugar, hemos vivido en estos últimos años un aumento exponencial en el volumen de datos que se registran diariamente. Un determinado *dataset* puede contener cientos de miles de observaciones, lo que en el caso de *factor investing* en renta fija corporativa se traduce en una mayor disponibilidad de series históricas de rendimiento, retornos, vencimientos y duraciones de miles de bonos corporativos simultáneamente. Esto ha hecho que el procesamiento humano de dichos volúmenes de datos sea inconcebible. No obstante, estos datos pueden ser aprovechados y procesados por

³⁰ Gergana Jostovac, Doron Avramova, Tarun Chordiab, Alexander Philipovd. (2013). Anomalies and financial distress. *Journal of Financial Economics*, 108(1), 139–157.

algoritmos de *machine learning* e inteligencia artificial que no sólo tienen capacidad de procesamiento para lidiar con ellos, sino que realizan mejores predicciones cuantas más observaciones tienen en cuenta. Además, esto también implica que se puede analizar la contribución a la rentabilidad de una cartera de más factores de forma simultánea y rápida.

Asimismo, a través del *machine learning* podemos lograr encontrar patrones difícilmente extrapolables para el ojo humano, e identificar complejas relaciones no lineales en las series históricas analizadas.

Sin embargo, se ha de mencionar que estas técnicas también encuentran sus limitaciones. Para empezar, y de especial importancia es que estos algoritmos analizan datos pasados de los cuales se pretenden realizar inferencias futuras. Sin embargo, las circunstancias en cada momento son únicas, por lo que las predicciones a futuro pueden no ser satisfactorias en un entorno cambiante.

Además, se ha de ser especialmente cuidadoso a la hora de seleccionar y elaborar los *datasets*, puesto que (especialmente respecto de datos financieros) podemos encontrar mucho ruido que haga que las conclusiones extraídas sean inconsistentes.

Cuando nos referimos a *machine learning*, aludimos al uso de técnicas computacionales que facilitan el aprendizaje automático y la identificación de patrones a partir de una serie histórica de datos. A partir del entrenamiento de modelos predictivos mediante el uso de datos históricos, se pueden inferir predicciones futuras a determinados niveles de confianza. Partiendo de esta base, debemos analizar cómo se pueden aplicar distintos algoritmos al *factor investing*, y qué valor añadido aportan en la gestión por factores.

Una de las características más importantes de los modelos de factores es que deben explicar el comportamiento de los activos que analizan atendiendo a un número limitado de características (factores). Dicho esto, uno de los retos principales a la hora de diseñar un algoritmo es establecer un número de factores suficiente para explicar el comportamiento de las carteras o bonos analizados, pero que a su vez sea lo suficientemente simple como para llegar a soluciones fácilmente comprensibles y razonables, así como computacionalmente manejables. Así pues, el *trade-off* entre simplicidad y poder informativo es un aspecto a tener en cuenta.

Para tomar esta decisión, se ha de realizar una comparativa entre modelos lineales y no lineales y las consecuencias del empleo de cada uno de ellos. En primer lugar, los modelos lineales que se limitan a emplear una serie pequeña de factores corren el riesgo de no explicar de forma efectiva, ni con suficiente nivel de profundidad el comportamiento de un determinado porfolio. Por otro lado, los modelos no lineales no adolecen de este riesgo de excesiva simplicidad, pero también tienen una serie de limitaciones: la primera, que la estructura del modelo tiende a estar fuertemente condicionada por los datos de entrenamiento. Si el número de observaciones crece o se contrae, encontramos que la función que nos proporciona el modelo varía notablemente. Además, en los modelos no lineales la estimación de parámetros no siempre se puede deducir de forma analítica, sino que se realiza mediante métodos iterativos, en los que las optimizaciones pueden no converger satisfactoriamente con los valores que se imputan inicialmente, lo que lleva a soluciones imprecisas.

De igual manera, debemos realizar una distinción entre algoritmos de aprendizaje supervisado y no supervisado. En primer lugar, los métodos de aprendizaje supervisado se caracterizan por el uso de conjuntos de datos de entrenamiento etiquetados. El objetivo principal del aprendizaje supervisado es la generación de reglas e inferencias fiables, aplicables a nuevos conjuntos de datos tanto para problemas de clasificación como de regresión. Así pues, mediante el conjunto de entrenamiento se entrena el modelo y, a partir de las reglas extrapoladas, se predicen o clasifican resultados para datos diferentes.

Por otro lado, los modelos de aprendizaje no supervisado son aquellos en los que, a través de técnicas de *clustering* y reducción de dimensiones, se trata de inferir información de un *dataset* y clasificar las observaciones. El *k-means* es el algoritmo de aprendizaje no supervisado más extendido y común.

Para su aplicación en el ámbito del *factor investing*, resulta conveniente profundizar en los distintos tipos de algoritmos supervisados que se pueden emplear para demostrar la existencia de factores de inversión en el mercado de renta fija corporativa. Como nuestro propósito es demostrar cómo a través del *factor investing* se pueden obtener mayores rentabilidades que los índices de referencia, vamos a usar algoritmos de aprendizaje supervisado, pues debemos realizar predicciones sobre la rentabilidad de los factores. Así pues, vamos a distinguir las diferentes clases de algoritmos de aprendizaje supervisados, su funcionamiento, y las ventajas e inconvenientes que suponen para su aplicación a la cuestión mencionada. A modo de

resumen, vamos a profundizar en: modelos de regresión penalizada lineales, árboles de decisión, redes neuronales, y máquinas de vectores de soporte (SVM).

4.2. PENALIZED REGRESSIONS

En primer lugar, encontramos los modelos de regresión penalizada, cuyas bases teóricas se remontan más de dos décadas atrás. El modelo, matemáticamente, funciona de la siguiente manera: dada una matriz de predictores \mathbf{X} , se busca descomponer el vector de output \mathbf{y} como una función lineal de las columnas de la matriz \mathbf{X} , más un término de error:

$$\epsilon: \mathbf{y} = \mathbf{X}\beta + \epsilon$$

\mathbf{y} = vector de output

\mathbf{X} = matriz de predictores

ϵ = error

La matriz de predictores serían los distintos tipos de factores. Para elección de la beta, se escoge la que minimice el error. Este error mínimo será el que minimice la suma de cuadrados de los residuos:

$$L = \epsilon' \epsilon = \sum_{i=1}^I \epsilon^2$$

L = suma de cuadrados de residuos

ϵ = error

I = observaciones en un tiempo t

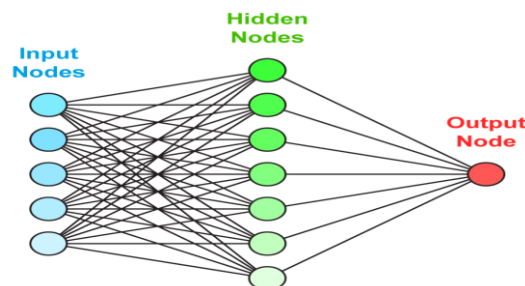
La solución al modelo planteado será la optimización de la β , para lo cual debemos diferenciar suma de residuos al cuadrado respecto de β , puesto que la condición de primer orden requiere que el gradiente sea igual a cero.

Hay varias aplicaciones posibles para estos modelos. La primera es recurrir a las penalizaciones para mejorar la solidez de las regresiones predictivas basadas en factores. El resultado puede utilizarse entonces para alimentar un esquema de asignación de factores. Por ejemplo, Han et al. (2019) y D. Rapach y Zhou (2019) utilizan regresiones penalizadas para mejorar la

predicción de la rentabilidad de las acciones cuando se combinan previsiones que emanan de características individuales de las compañías.

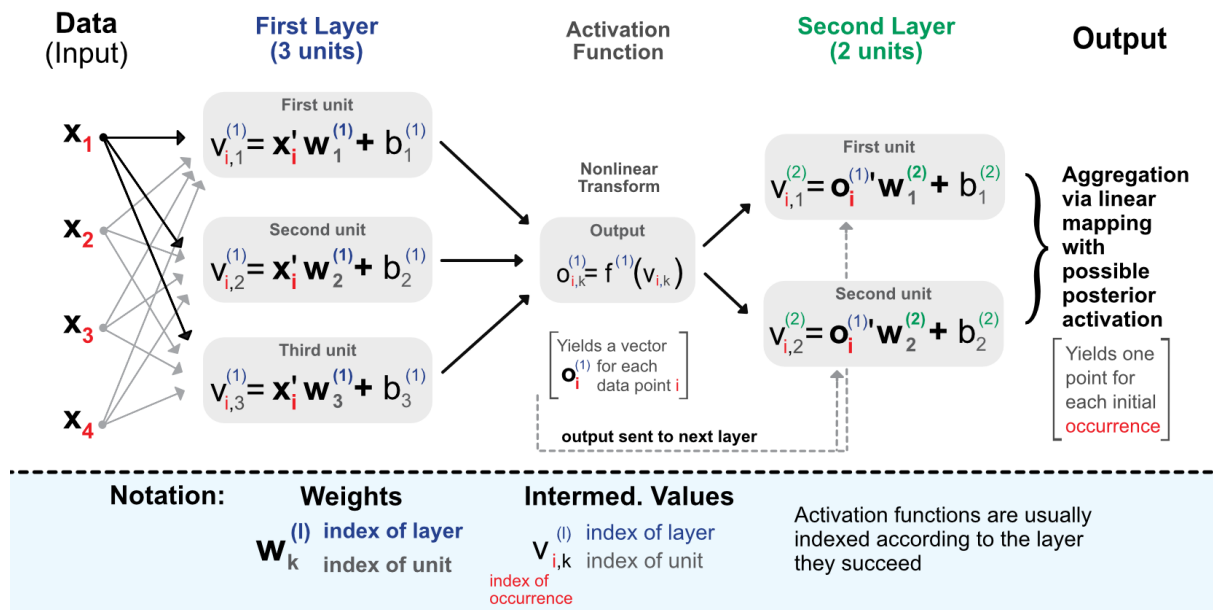
4.3. NEURAL NETWORKS

Otro tipo de modelos de aprendizaje supervisado de aplicación para *factor investing* son las redes neuronales (*neural networks*). François Chollet las define como: "cadenas de funciones geométricas diferenciables y parametrizadas, entrenadas con descenso de gradiente (con gradientes obtenidos mediante la regla de la cadena)". De forma más sencilla, las redes neuronales supuestamente emulan el funcionamiento del cerebro humano, y consisten en unidades de procesamiento (nodos) ordenadas en capas. Estas capas normalmente están formadas por: una capa de entrada, que recoge las observaciones y campos de entradas; una o varias capas ocultas en las que se realizan distintas ponderaciones; y finalmente una capa de salida que devuelve un resultado. De forma visual, esto sería una red neuronal:



Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

Cada uno de estos nodos se puede ver como un modelo lineal al que se le aplica una función en concreto. Éstas reciben el nombre de ‘funciones de activación’, y su propósito primordial es dar paso a elementos no-lineales en el modelo. La idea detrás de las redes neuronales es combinar nodos por bloques (capas). De forma más visual:



Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

El proceso funciona de la siguiente manera: cuando los datos entran en la red, pasan por el mapeo lineal inicial. De forma matemática:

$$v_{i,k}^{(1)} = x_i' w_k^{(1)} + b_k^{(1)}, \text{ for } l = 1, k \in [1, U1]$$

v = vector de resultado del mapeo lineal

X = matriz de predictores para cada observación x

w = peso/ponderación asignada al componente k , de la capa (1)

$$b_k^{(1)} = \text{sesgo}$$

l = capa de la red

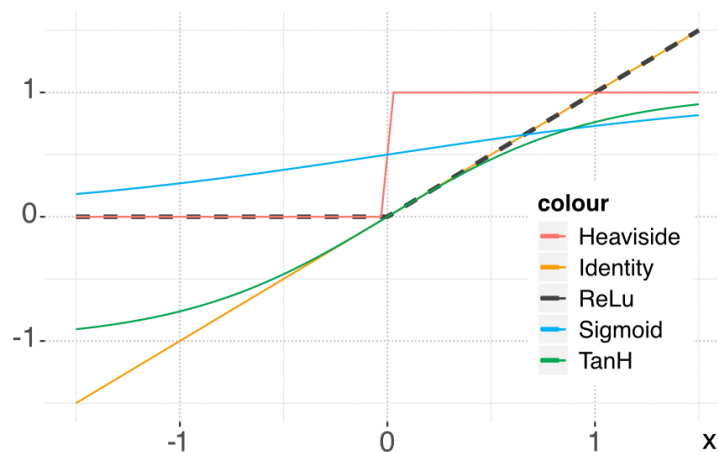
Entonces, los datos son transformados por una función no lineal f^n , y el resultado de esta iteración será el input para las neuronas de la siguiente capa, y así sucesivamente. Las formas lineales se repiten (con diferentes pesos) para cada capa de la red. Matemáticamente, esto viene recogido en la expresión:

$$v_{i,k}^{(1)} = (o_i^{(l-1)}) w_k^{(1)} + b_k^{(1)}, \text{ for } l = 1, k \in [1, U1]$$

Las conexiones entre las capas de la red no son más que los outputs de los nodos anteriores que, a su vez, son puntos u observaciones lineales a las que se les ha aplicado funciones de activación ' f '. El output de la capa l es el input de la capa $l+1$.

Finalmente, el último nodo agrega los resultados de la red en la capa de salida, empleando el mismo mecanismo de ponderación a través de funciones de activación.

Naturalmente, una de las decisiones más importantes que se han de tomar a la hora de implementar este modelo es la elección de la función de activación. Vemos en el siguiente gráfico las más comunes:



Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

Este proceso aplicado a factor investing se traduciría en lo siguiente: los inputs X serían los factores o características de las empresas. A éstas se les asigna un peso y un bias en todos los nodos de la primera capa, en línea con lo que antes hemos denominado “mapeo lineal inicial”. Posteriormente, la combinación lineal resultante es transformada por la función de activación. Así, cada uno de los nodos devuelve un valor que es el input de las neuronas de la siguiente capa, que siguen el mismo proceso. Este proceso itera para toda la red neuronal, con el propósito de que en la capa final devuelva un resultado bien numérico, o bien categórico.

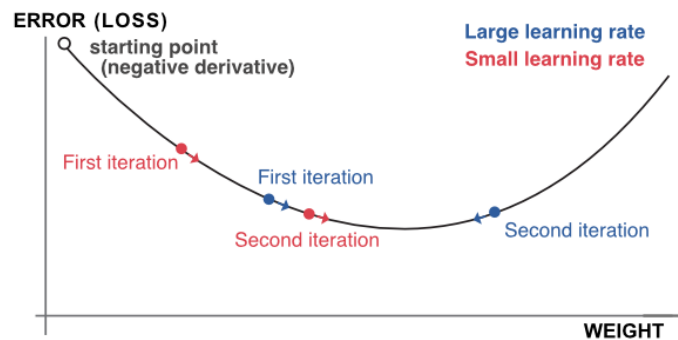
El algoritmo se entrena y “aprende” minimizando una función de pérdida de información o error sujeta a una penalización, el *back-propagation*. Matemáticamente se denota de la siguiente manera:

$$0 = \sum_{i=1}^I \text{pérdida}(y_i, \check{y}_i) + \text{penalización}$$

y_i = parámetros verdaderos que debe devolver el modelo
 \check{y}_i = valores obtenidos por el modelo

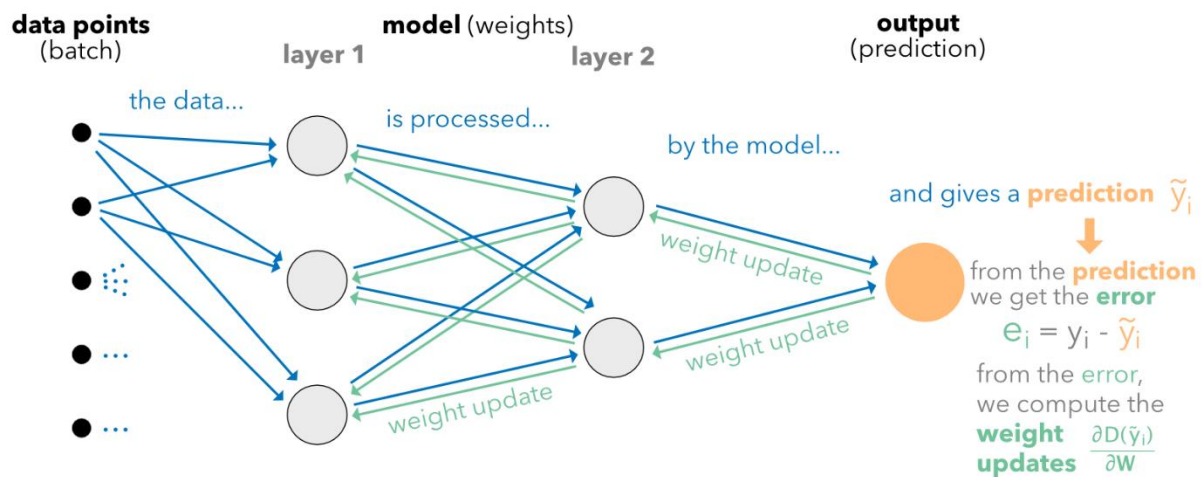
La función de pérdida más comúnmente empleada es la suma cuadrada de errores para regresiones, mientras que suele emplearse entropía cruzada para labores de clasificación.

El objetivo de entrenar las redes neuronales es alterar los pesos y biases para todos los nodos de todas las capas, de manera que la función definida anteriormente sea lo menor posible. Esta alteración de los pesos se produce a través del descenso de gradiente, que determina numéricamente dónde una función genera los valores más bajos, es decir, determina mínimos locales. De forma iterativa, se trata de minimizar la pérdida de información (error) del modelo haciendo que la función se mueva en dirección opuesta a la del gradiente (pendiente de la función). De forma más visual:



Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

Para visualizar el proceso de *back-propagation* véase el siguiente gráfico:



Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

Los datos fluyen a través de los nodos hasta llegar al output o predicción, de la cual se obtiene la función de pérdida de información, de la que, a su vez, se obtienen todas las derivadas (pesos y biases). Éstas se computan desde el nodo de salida hacia atrás, donde cada una de las neuronas recibe aproximadamente la parte proporcional del error que ha registrado del resultado. Este proceso se repite capa por capa y neurona por neurona hasta que toda la red haya recibido una aproximación del error atribuido a cada neurona. Posteriormente, se ajusta el modelo para reducir el error de los puntos donde se produce el error.

Las aplicaciones de los modelos de redes neuronales en el campo financiero y *factor investing* son extensas. En el campo de la renta variable existen multitud de estudios, entre los que podemos destacar los siguientes:

Feng, Polson y Xu (2019), que utilizan redes neuronales para encontrar los factores que mejor explican la sección transversal de los rendimientos de las acciones. En este reciente estudio, los autores construyen un modelo de factores clasificado por características, desarrollando un enfoque *bottom-up* con *machine learning* profundo de última generación. Con el objetivo económico de minimizar los errores de fijación de precios, entrenan una red neuronal utilizando las características de la empresa [*input*], y generan factores [características intermedias], para ajustar los rendimientos de los valores [*outputs*]. La clasificación de los valores en función de las características de la empresa proporciona una activación no lineal para crear ponderaciones para carteras *long-short*. Su modelo ofrece una perspectiva alternativa para la reducción de la

dimensiones de las características de la empresa [*inputs*], en lugar de los factores [características intermedias], y permite tanto la no linealidad como las interacciones en las entradas. Así pues, los autores encuentran pruebas estadísticas y económicas sólidas en las carteras que no forman parte de la muestra y en los rendimientos de las acciones individuales. De esta forma, consiguen ratios de Sharpe significativamente mejores.³¹

Luyang Chen, Pelger y Zhu (2020) estiman el núcleo de precios con una estructura de red neuronal compleja que incluye una red generativa adversarial. Esto, de nuevo, proporciona información crucial sobre la estructura de los rendimientos esperados de las acciones y puede utilizarse para la construcción de carteras (mediante la construcción de una política precisa de máxima relación de Sharpe). Utilizando redes neuronales complejas, producen un modelo de valoración de activos para los rendimientos de acciones individuales, aprovechando la gran cantidad de información disponible, a la par que manteniendo una forma totalmente flexible y teniendo en cuenta la variación temporal.³²

Las innovaciones clave de este estudio residen en que los autores emplean la condición fundamental de no arbitraje como función de criterio, para construir los portfolios de prueba más informativos y abstraer los estados de la economía de las series temporales macroeconómicas. De esta manera, logran construir un modelo de fijación de precios de los activos que supera, fuera de la muestra, todos los enfoques de referencia en términos de ratio de Sharpe, variación explicada y anomalías en la fijación de precios, e identifica los factores clave que impulsan los precios de los activos.³²

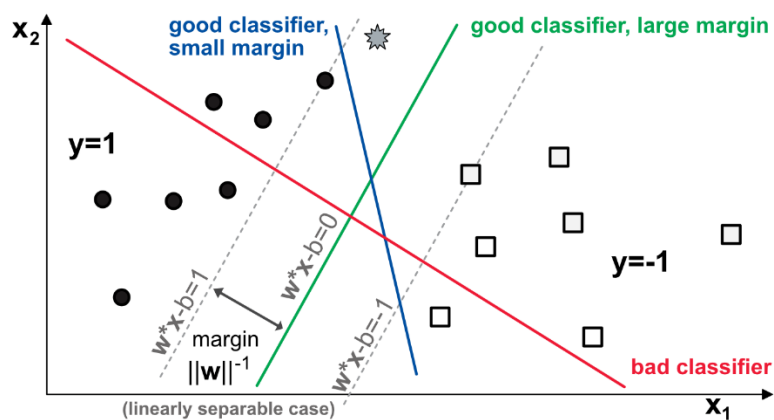
4.3. SUPPORT VECTOR MACHINES

³¹ Guanhao Feng, Nicholas G. Polson, Jianeng Xu. (2019). Deep Learning in Characteristics-Sorted Factor Models. Cornell University.

³² Chen, Luyang and Pelger, Markus and Zhu, Jason, Deep Learning in Asset Pricing (April 4, 2019). Available at SSRN: <https://ssrn.com/abstract=3350138> or <http://dx.doi.org/10.2139/ssrn.3350138>

Para continuar, otro posible algoritmo a emplear son las máquinas de vectores de soporte (*support vector machines*, SVM, en inglés), cuya popularidad y uso ha disminuido en favor de otros algoritmos (como las redes neuronales o los árboles de decisión), pero que se mencionarán brevemente para plasmar otra posibilidad dentro de los algoritmos de aprendizaje supervisado. Su funcionamiento es distinto en función de si se emplean para problemas de clasificación o de regresión. Por motivos de simplicidad y claridad en la explicación, se detallarán los modelos SVM para problemas de clasificación.

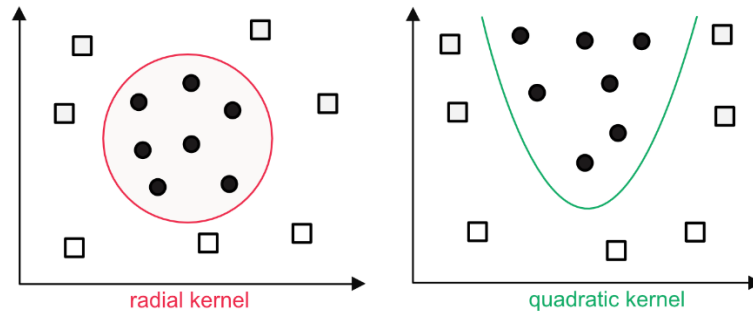
El objetivo de éstos es encontrar un modelo que clasifique correctamente puntos en un plano. Los modelos consisten en pesos que se aplican a las variables y crean una separación lineal natural. De forma más visual, una clasificación binaria:



Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

En esta separación concreta podemos distinguir 3 clasificaciones: la realizada por la línea roja, la de la línea verde y la de la azul. La línea roja no realiza una buena labor, puesto que no logra separar los puntos redondos de los cuadrados. En cambio, las líneas azul y verdes sí logran llevar a cabo esta clasificación. Los vectores de soporte se definen, pues, como las líneas que maximizan la distancia entre el modelo y los puntos más cercanos que estén clasificados correctamente. Así pues, la idea de los SVM es lograr la máxima separación entre los vectores sin cometer errores de clasificación, es decir, tratar de obtener el modelo más robusto entre los modelos que clasifican correctamente.

Esta tarea también se puede realizar para conjuntos no separables linealmente a través de los *kernels*. La idea es realizar, a través de funciones no lineales clasificaciones como las siguientes:

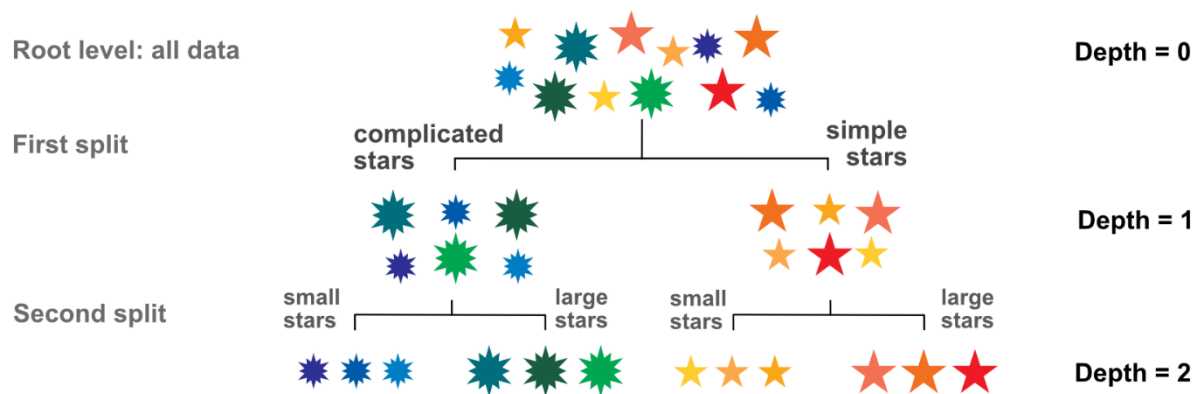


Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

4.4. RANDOM FORESTS

Los árboles de regresión y clasificación, así como sus derivados (random forests...) que se detallan a continuación, son algoritmos simples pero potentes que, además, suelen ser muy eficientes a la hora de realizar predicciones a partir de datos en formato tabla.

El método más sencillo en este campo son los árboles de decisión, cuyo objetivo es dividir los conjuntos de datos analizados en grupos homogéneos. Es decir, dado una variable exógena Y , y características X , un árbol trata de dividir el conjunto en grupos que sean lo más homogéneos en Y como sea posible. Las divisiones se realizan en función de una variable entre el conjunto de sus características. Si la variable es numérica, entonces el árbol será un árbol de regresión, mientras que, si la variable es categórica, hablaremos de árboles de clasificación. Se ilustra a continuación un ejemplo de este proceso de partición:



Fuente: Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

En este ejemplo, la variable Y sería el color, de forma que el objetivo del algoritmo es realizar distintas divisiones para lograr agrupar los datos en función de ciertas variables, para lograr homogeneizar el color dentro de los conjuntos finales. De esta manera, la primera partición se realiza en función de las puntas de las estrellas. Así pues, podríamos establecer como criterio que el número de puntas sea mayor que 5, o menor o igual a 5. En un segundo nivel, se realiza otra partición entre las estrellas grandes y las más pequeñas (con un área total de más de $x \text{ cm}^2$, por ejemplo). La comparación de este sencillo ejemplo con factor investing es simple y muy ilustrativa: la variable Y sería (color) sería el rendimiento del título o activo, siendo, por ejemplo, el azul los que logran mayores rendimientos, y los rojos los que menos. Los factores y características de las empresas relacionadas con los títulos (bien acciones, bien bonos corporativos) serían las que se emplean para realizar las particiones (por ejemplo, capitalización bursátil, LTM EPS...). En resumen, el propósito del modelo será encontrar las características que permiten dividir las empresas entre las que logran rendimientos mayores, y las que no.

Una vez visualizado el proceso, se procede a detallar matemáticamente cómo funcionan los árboles de decisión. Dado un conjunto de datos y_i, x_i de tamaño I , un árbol de regresión tratará de dividir los datos de forma que la variación total de y_i sea mínima. Así, primero encuentra para todos los factores x_i^k , el mejor criterio de partición para que los dos conjuntos sean lo más homogéneos posibles en y_i . Después, encuentra el factor que ha logrado el mayor nivel de homogeneidad.

La homogeneidad en los árboles de decisión está ligada a la varianza de los datos. Como lo que se busca es que la varianza de cada subconjunto sea lo más similar posible, lo que el algoritmo trata es de minimizar la dispersión dentro de cada uno y luego sumar estas dos magnitudes. Para tener en cuenta el tamaño de los subconjuntos la varianza será la total de cada grupo, es decir, multiplicando la varianza por el número de elementos de cada *cluster*. De forma matemática lo anterior se denota con la siguiente expresión:

$$V_I^{(k)}(c^{(k)}) = \sum (y_i - m_I(c^k))^2 + \sum (y_i - m_I(c^k))^2$$

$V_I^{(k)}$ = varianza del cluster k para todas las observaciones I

$c^{(k)}$ = cluster k

m_I = media del cluster k

En su máxima expresión, un árbol de regresión se puede dividir (cortar) tantas veces como se quiera, hasta que cada observación pertenezca a un subconjunto diferenciado, o no quepa dividir más las observaciones de los miembros de uno en función de las características definidas. Esto recibe el nombre de “hacer crecer el árbol completo”. No obstante, y como es obvio, esto llevaría muy seguramente a la producción de problemas de *overfitting*, de forma que sería un modelo muy preciso con los datos de entrenamiento, pero inútil con conjuntos de datos diferentes. Para evitar hacer en este problema, es importante determinar el número de particiones que se hace a los datos. Hemos de decir que las primeras particiones son las más importantes, pues determinan los patrones más relevantes y aglutinan la mayor parte de la información. Las particiones realizadas a niveles más profundos del árbol aportan menos información y explican características específicas de cada subgrupo que de poco sirven al análisis general. De esta manera, es importante limitar el número de particiones, o “talar” el árbol, para que deje de realizar particiones cuando éstas dejen de aportar información.

Existen numerosos métodos para ello: imponer un número mínimo de datos para cada nodo terminal (para cada hoja del árbol); establecer una condición de que los nodos deban tener un número mínimo de observaciones para hacer otra partición; exigir un cierto umbral de mejora en la aportación de información de la partición en cuestión, es decir, si una división no reduce suficientemente el error en la predicción, se considera innecesaria; o limitar la profundidad del árbol, es decir, limitar el número de particiones que se pueden realizar.

Como se ha detallado, los árboles de decisión pueden ser muy útiles para analizar y comprender conjuntos de datos complejos. Sin embargo, pueden mejorarse a partir de la idea de los *ensembles*, con los Random Forests. Éstos, básicamente hacen crecer varios árboles, que forman un “bosque”, y realizan una predicción a través una media ponderada de los resultados devueltos por todos ellos. Matemáticamente esto se expresa de la siguiente manera:

$$f^x = \frac{1}{J} \sum_{j=1}^J h_j(x)$$

f^x = predicción del random forest

J = número total de clusters

$h_j(x)$ = resultado del árbol j para la variable x

Este algoritmo emplea distintos conjuntos de entrenamiento para cada uno de los árboles mediante el *bootstrap*, que permite que los datos del conjunto original se muestreen con o sin reemplazo, formando conjuntos de datos de entrenamiento distintos para cada uno de los árboles del bosque. La agregación de un gran número de árboles entrenados vía *bootstrapping* se conoce como *bagging*. Esto enriquece la predicción, y la hace más robusta, reduciendo el riesgo de *overfitting*. Si se emplea sólo un árbol para realizar predicciones, el algoritmo puede acabar aprendiendo relaciones altamente idiosincrásicas entre los datos y, por tanto, ajustar en exceso el modelo. Sin embargo, usando *bootstrapping* el modelo puede identificar relaciones a nivel más general de las variables y características.

Los *Random Forests* devuelven dos resultados: el primero es simplemente un conjunto de valores condicionales, y el segundo es la importancia de factores (*feature importance*, FI, en inglés). Como su propio nombre implica, muestra la relevancia de cada variable explicativa y su contribución a la predicción de la variable. La *feature importance* calcula el grado en que el poder predictivo del modelo se diluiría si se realiza una partición más.

V. EJEMPLO DE APLICACIÓN DE RANDOM FOREST

Para poner en práctica lo anteriormente recogido, realizo un breve ejemplo de aplicación de *machine learning* para predecir el precio de un determinado bono. De cara a realizar dicho ejercicio, uso un conjunto de datos que describen los registros históricos de transacciones, los cálculos intermedios y los precios históricos puestos a disposición (en Kaggle).

Este *dataset* contiene 61 atributos observados para cada uno de los 762.678 bonos: 3 nominales, 12 ordinales discretos, 1 peso de observación y 45 continuos (ratio). Para cada uno de ellos, se da el siguiente detalle³³:

- **id**: El id de la fila
- **bond_id**: El id único de un bono para ayudar a la reconstrucción de la serie temporal. (Esta columna sólo está presente en los datos de entrenamiento)
- **trade_price**: El precio al que se produjo la operación. (Esta es la columna a predecir en los datos de prueba)
- **peso**: El peso de la fila a efectos de evaluación. Se calcula como la raíz cuadrada del tiempo transcurrido desde la última operación y se escala para que la media sea 1.
- **cupón_actual**: El cupón del bono en el momento de la operación.
- **tiempo_hasta_el_vencimiento**: El número de años hasta el vencimiento del bono en el momento de la operación.
- **is_callable**: Valor binario que indica si el bono es rescatable o no por el emisor.
- **reporting_delay**: El número de segundos después de que se produjera la operación que fue reportada.
- **trade_size**: El importe nocional de la operación.
- **trade_type**: 2=venta del cliente, 3=compra del cliente, 4=comercio entre distribuidores. Es de esperar que los clientes obtengan, por término medio, peores precios que los concesionarios.
- **curve_based_price**: Una estimación del precio justo basada en las curvas de riesgo y financiación implícitas del emisor del bono.

³³ Benchmark Bond Trade Price Challenge. (s/f). Kaggle. <https://www.kaggle.com/competitions/benchmark-bond-trade-price-challenge/data?select=train.mat>

- `received_time_diff_last{1-10}`: La diferencia de tiempo entre la operación y la anterior {1-10}.
- `trade_price_last{1-10}`: El precio de la operación de las últimas {1-10} operaciones.
- `trade_size_last{1-10}`: El importe nominal de las últimas {1-10} operaciones.
- `trade_type_last{1-10}`: El tipo de operación de las últimas {1-10} operaciones.

Asimismo, se emplea este código³⁴, para realizar diversos análisis, cuyos resultados varían en función de las asunciones tomadas.

```
library(randomForest)
library(plyr)

# Seed the random number generator for reproducibility
set.seed(68768768)

print("Reading the data")
train <- read.csv( "train.csv", header = TRUE, na.strings = "NA")
test <- read.csv( "test.csv", header = TRUE, na.strings = "NA")

# Uncomment the line below to use only a small portion of the training data
# train = train[1:20000,]

# Append indicators for "NA" and replace NA values with something else
appendNAs <- function(dataset) {
  append_these = data.frame( is.na(dataset[, grep("received_time_diff_last",
names(dataset))]) )
  names(append_these) = paste(names(append_these), "NA", sep = "_")
  dataset = cbind(dataset, append_these)
  dataset[is.na(dataset)] = -1000
  return(dataset)
}

print("Resolving NAs")
train <- appendNAs(train)
test <- appendNAs(test)

print("Training the Random Forest")
rf <- randomForest(train[,-c(1,2,3)],train$trade_price,
do.trace=TRUE,importance=TRUE, sampsize = 10000, ntree = 200)

print("Making Predictions on the Test Set")
predictions = predict(rf, test)

print("Creating the Submission File")
predictions_df <- data.frame(test$id, predictions)
names(predictions_df) <- c("id", "trade_price")
write.csv(predictions_df, file = "random_forest_sample_submission.csv", row.names =
FALSE)
```

³⁴ Benchmark Bond Trade Price Challenge. (s/f). Kaggle. <https://www.kaggle.com/competitions/benchmark-bond-trade-price-challenge/data?select=train.mat>

Bibliografía

AFI. (2020, abril). *Factor Investing & ETF: Gestión por Factores para una inversión más eficiente*.

Carhart, M. M. (2012). *On Persistence in Mutual Fund Performance*.
<https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>

Chen, Luyang and Pelger, Markus and Zhu, Jason, Deep Learning in Asset Pricing (April 4, 2019). Available at SSRN: <https://ssrn.com/abstract=3350138> or
<http://dx.doi.org/10.2139/ssrn.3350138>

Cfa, T. P. C. J. (Ed.). (2019). *Fixed income factors: Fixed income's worst kept secret*.
Systematic Fixed Income | By The Numbers.

de Carvalho Patrick Dugnonle Xiao Lu and Pierre Moulin, R. L. (2014). Low-Risk Anomalies in Global Fixed Income: Evidence from Major Broad Markets. *The Journal of Fixed Income Spring 2014*, 23(4), 51–70.

de Carvalho, T. H. Z. A. I. H. O. L. R. L. (2021). Out-performing corporate bonds indices with factor investing. *Forthcoming in Bankers Markets & Investors*.

de Elejabeitia, P. G. (2018). *GESTIÓN DE CARTERAS. GESTIÓN ACTIVA VS. GESTIÓN PASIVA*. UNIVERSIDAD PONTIFICIA COMILLAS .

Gergana Jostovac, Doron Avramova, Tarun Chordiab, Alexander Philipovd. (2013).
Anomalies and financial distress. *Journal of Financial Economics*, 108(1), 139–157.

Guillaume Coqueret, T. G. (2020). *Machine Learning for Factor Investing*. Chapman and Hall/CRC.

Guanhao Feng, Nicholas G. Polson, Jianeng Xu. (2019). Deep Learning in Characteristics-Sorted Factor Models. Cornell University.

King, K.-H. (2004). Return reversals in the bond market: Evidence and causes. *Journal of Banking & Finance*, 28(3), 569–593.

Los Essentials del Factor Investing. (s/f). Robeco.

Narasimhan Jegadeesh, S. T. (2017). *Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency*. <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>

Patrick Houweling, J. van Z. (2017). Patrick Houweling [PDF] from efmaefm.org Factor Investing in the Corporate Bond Market. *Financial Analysts Journal*, 73(2), 1–16.

Soe Hong Xie CFA, A. (Ed.). (2016). *Factor-Based Investing in Fixed Income: A Case Study of the U.S. Investment-Grade Corporate Bond Market*.

van de Wetering, G. J. (2017). *Using Factor Investing as Investment Strategy for Corporate Bonds in Europe and the United States*. Tilburg School of Economics and Management .