



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

“DESARROLLO DE UN MODELO DE ASSET ALLOCATION PARA FONDOS MULTIACTIVOS”

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

JOSÉ MIGUEL QUIROZ FERRER

PROFESOR GUÍA

CARLOS PULGAR ARATA

MIEMBROS DE LA COMISIÓN:

ANTONIO BARRIOS NAMUNCURA

JOSÉ MIGUEL CRUZ GONZÁLEZ

SANTIAGO DE CHILE

2024

Resumen Ejecutivo

En el 2020, Bloomberg publicó un artículo donde se consultaba a 16 altos ejecutivos de empresas de inversión, como JP Morgan, BlackRock y AQR, sobre sus perspectivas para la próxima década en este ámbito. Todos coincidieron en que el análisis cuantitativo se convertiría en una herramienta esencial para la creación de valor, tanto a nivel global como en el ámbito de las inversiones.

En la actualidad, se genera una vasta cantidad de información a cada segundo, disponible de forma prácticamente instantánea. Esta situación propicia que los mercados sean cada vez más complejos y competitivos. Ya que la información está al alcance de todos. Por lo tanto, la capacidad de extraer *insight's* de esos datos se convierte en una ventaja competitiva fundamental en la generación de valor.

Es en este contexto que la empresa donde el alumno desarrollo su proyecto de título buscaba desarrollar un modelo cuantitativo que apoyara la toma de decisión de los *portfolio managers* en el *asset allocation* de los fondos multiactivos que administran. Por esta razón se formuló el siguiente objetivo general; “Desarrollar un modelo de *asset allocation* global que genere un vector de recomendación suma cero para fondos multiactivos a partir de data macroeconómica y de mercado, maximizando el *information ratio*”.

El modelo se basa en la metodología CRISP-DM, un proceso iterativo de seis etapas propio de la ciencia de datos. El plan consistió en construir una amplia base de datos con información macroeconómica y de mercado, la cual serviría para generar una asignación de activos global (renta variable (RV), renta fija (RF) y Caja) a partir de la construcción de un puntaje o "*scoring*" que reflejara el nivel de convicción del modelo sobre si sobreponderar, subponderar o mantener la posición de un activo. Para posicionar los activos dentro del fondo desde el enfoque global, se siguió una estrategia de tipo "*top-down*".

Los resultados del modelo fueron positivos. Mediante el uso de componentes principales, se logró capturar de manera adecuada la variabilidad de la base de datos original, reduciendo su dimensión de 263 variables a 14 componentes principales. Esto permitió generar un alfa positivo y, en consecuencia, un *information ratio* de 0,97, superando las expectativas iniciales de la empresa, que se ubicaban entre 0,4 y 0,6.

Si bien el modelo buscaba ser explicativo, la utilización de componentes principales dificultó esta tarea, ya que estos reducen la dimensionalidad de la base de datos captando relaciones algebraicas complejas. Para abordar esta limitación, se probaron sin éxito dos estrategias: correlacionar componentes con clases de variables y calcular componentes principales por clases de variables. Siendo las clases de variables macrocategorías como: cíclicas, de condiciones de mercado, sentimiento, técnicas, entre otras.

Tabla de contenido

1. Antecedentes Generales	1
1.1 Industria de las AGF's en Chile	2
1.2 Contextualización de la Empresa	3
1.3 Portfolio Management.....	5
2. Proyecto y justificación.....	7
2.1 Justificación.....	8
2.2 Proceso Inversión Cuantitativa.....	9
3. Objetivo General	10
4. Objetivos Específicos.....	11
5. Alcances	11
6. Marco Conceptual	12
6.1 Series de Tiempo.....	12
6.2 Correlación de Pearson.....	13
6.3 Correlación de Spearman	13
6.4 Test de Causalidad de Granger.....	14
6.5 PCA (<i>Principal Component Analysis</i>)	15
6.6 Teoría Moderna de Portafolio (Markowitz)	15
6.7 Alfa.....	18
6.8 Tracking Error (TE)	18
6.9 Information Ratio (IR)	18
6.10 Look Ahead Bias.....	19
6.11 Metodología CRISP-DM.....	19
7. Metodología	20
8. Resultados	29
9. Discusiones	37
10. Conclusiones	41
Bibliografía	43
Anexos.....	45

1. Antecedentes Generales

El mundo de las inversiones ofrece una amplia gama de instrumentos financieros que permiten a las personas invertir su capital y generar rentabilidades en el corto, mediano o largo plazo, de acuerdo con sus objetivos y perfil de riesgo. Entre los instrumentos más comunes encontramos acciones, bonos, futuros, fondos mutuos y *Real State*, por mencionar algunos.

Los fondos mutuos son una alternativa atractiva como forma de inversión, ya que permiten a los inversionistas invertir en un portafolio diversificado con distintos tipos de instrumentos, en otras palabras, en este tipo de fondos un tercero escoge una serie de instrumentos financieros con los que construye un portafolio, a partir del cual un aportante (persona jurídica o natural) compra un derecho de participación en dicho fondo, una cuota del fondo. Esto permite que los aportantes inviertan de manera directa en distintos instrumentos, permitiéndoles diversificar el riesgo.

Los fondos mutuos son administrados (al menos en Chile) por sociedades anónimas, cuyo objeto exclusivo es la administración de recursos de terceros. Los participantes son los dueños del fondo y asumen los riesgos y beneficios de la inversión. Estas sociedades anónimas deben tener la denominación legal de AGF (Administradora General de Fondos) para administrar fondos de Inversión. La denominación es otorgada por la Comisión para el Mercado Financiero (CMF), institución que se encarga de supervisar que las AGF's cumplan la ley N°20.712 y Decreto Supremo N°129 de Hacienda [1].

Según lo que declara la CMF, en base a la ley N°20.712 [2], existen distintos tipos de fondos mutuos, los que a nivel general sólo pueden invertir en instrumentos de deuda de corto, mediano y largo plazo e instrumentos de capitalización. Los diferentes tipos de fondos mutuos se clasifican dependiendo del tipo de estrategia de inversión que se utilice en su construcción, estas son [3]:

FONDO TIPO 1	Deuda de corto plazo con duración menor o igual a 90 días.
FONDO TIPO 2	Deuda de corto plazo con duración menor o igual a 365 días.
FONDO TIPO 3	Deuda de mediano y largo plazo.
FONDO TIPO 4	Mixto.
FONDO TIPO 5	De Capitalización.
FONDO TIPO 6	De libre inversión.
FONDO TIPO 7	Estructurado
FONDO TIPO 8	Dirigido a inversionistas calificados.

Tabla 1: Tipos de Fondos Mutuos según la CMF.

1.1 Industria de las AGF's en Chile

La industria de los fondos mutuos está bastante desarrollada en el mercado chileno, debido a que existen múltiples actores (AGF's) que compiten entre sí a diario transformando la industria en un mundo de altamente competitivo. El patrimonio administrado por la industria de fondos mutuos [4] alcanzó a los \$57.317.031 MM CLP, lo que significa un crecimiento de un 20 % durante el 2023. Además, si se observa la tabla 2, se pueden apreciar algunas estadísticas generales de la industria elaboradas por la Asociación de Fondos Mutuos, destacándose que a la fecha de diciembre de 2023 estos fondos representan un 19% del PIB. De la misma manera se registró un incremento de un 6,79% de los participantes (individuos que invierten fondos mutuos) en comparación al 2022, mostrándose, por consiguiente, un desafío de generar valor tanto para los participantes como para las AGF's.

Año	AUM (MM\$ CLP)	% AUM	Participes	% Participes	Series	% Series
2018	37.637.882		2.669.827		2.550	
2019	45.052.997	19,70	2.536.645	-4,99	2.614	2,51
2020	51.307.921	13,88	2.650.142	4,47	2.660	1,76
2021	50.987.024	-0,63	2.926.542	10,43	2.666	0,23
2022	47.810.342	-6,23	2.973.508	1,60	2.733	2,51
2023	57.317.031	19,88	3.175.297	6,79	2.860	4,65

Tabla 2: Estadísticas Generales de la Industria de Fondos Mutuos.

Esta industria cuenta con la participación de 23 administradoras a fecha de diciembre de 2023 [4], a continuación, en la Ilustración 1-1 se puede apreciar la distribución que estas tienen tanto a nivel de patrimonio como a nivel de participantes. Por un lado, a nivel de patrimonio destacan Banchile (21,64%), Santander A.M (18,6%) y BCI (12,3%). Por otro lado, en cantidad de clientes destacan Banco Estado (33,1%), Banchile (11,8%) y Santander A.M (11,5%). A partir de lo anterior se deduce que los actores más grandes en la industria son Banchile y Santander A.M.

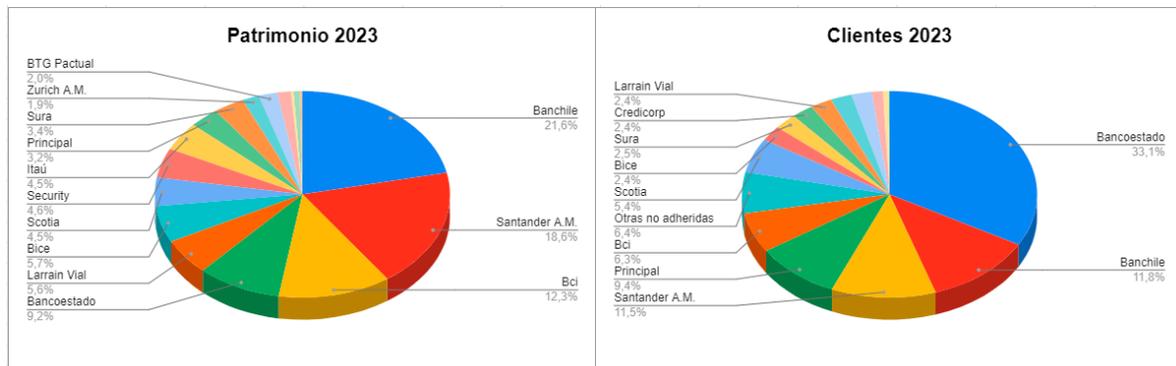


Ilustración 1-1: Repartición del mercado de Fondos Mutuos 2023.

El modelo de negocio de los fondos mutuos se basa en la captación de recursos de clientes, la gestión profesional de las inversiones, la distribución de utilidades y el cobro de comisiones. La rentabilidad del fondo depende del desempeño y la gestión de los activos invertidos.

Existen distintos tipos de incentivos para los *portfolio managers* (PM), de los cuales destacan dos principalmente:

- **Comisión de gestión actual (*Current Management Fee*):** Es un porcentaje que el PM gana típicamente en función del dinero que administra, o mejor dicho, de la cantidad de AUM (*Assets Under Management*) que gestiona.
- **Incentivo por rendimiento del fondo:** El PM obtiene ganancias en función del desempeño que tenga el fondo.

1.2 Contextualización de la Empresa

La empresa en la cual el estudiante realizó la práctica profesional corresponde a una AGF. Sin embargo, la empresa posee otras líneas de negocio que van más allá de ser solo una administradora de fondos, por ejemplo, tiene la posibilidad de invertir en acciones, ETF's, inversión extranjera, seguros de vida, APV, entre otros. Para efectos de esta memoria, el presente trabajo se enfocará en los fondos de multiactivos, los cuales son un tipo de fondo mutuo, que se categoriza entre los tipos 4 a 7 dependiendo de su composición interna.

Los Fondos Multiactivos como se mencionó anteriormente corresponden a un tipo de fondo mutuo que sigue una estrategia de multiactivos [5]. Esta estrategia consiste en combinar diferentes tipos de activos con el fin de crear un portafolio más ágil y diversificado que posee expectativas de retorno a largo plazo. En estos fondos los PM's van decidiendo a través del tiempo las ponderaciones de los activos (*asset allocation*) con el fin de generar lo que se conoce como "alfa" (exceso de retorno). El alfa es una medida ampliamente utilizada en inversiones como una medida de exceso de rendimiento en comparación a un índice o *benchmark* de referencia.

La empresa cuenta con 28 de estos fondos los cuales se distribuyen en cinco categorías Renta Fija Internacional, Balanceados Conservadores, Balanceados Equilibrados, Balanceado Agresivos y Renta Variable Internacional (ver Ilustración 1.2-1). Los activos dentro de los fondos están estructurados de tal manera que sus ponderaciones en total sumen cero.

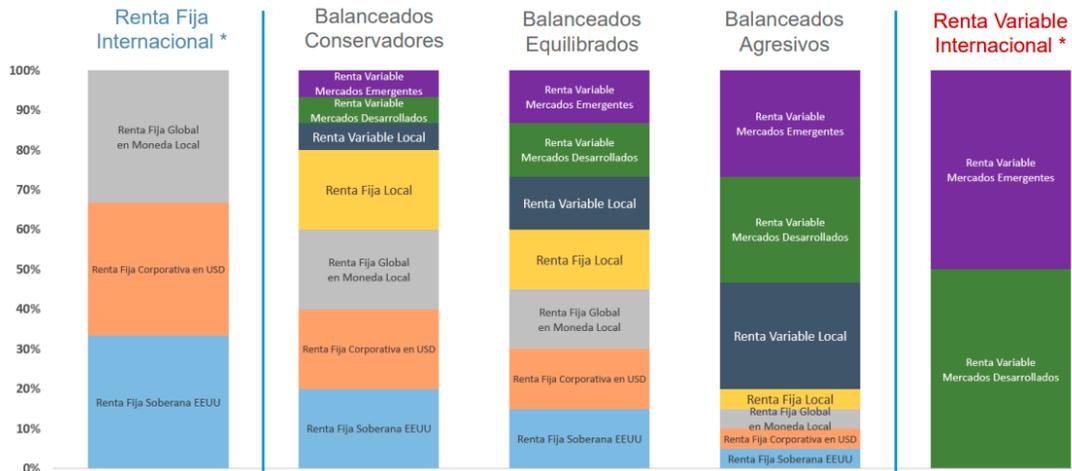


Ilustración 1.2-1: Tipos de fondos multiactivos que la empresa gestiona.

La propuesta de valor de la empresa para estos fondos radica en la posibilidad de invertir en fondos de múltiples jurisdicciones. Lo anterior supone ventajas para los inversionistas de poder acceder a una gama más amplia de activos y de mercados, así como también beneficiarse de las diferentes condiciones económicas y de crecimiento de distintos países, reduciéndose de la misma manera los riesgos asociados a la concentración en un solo mercado o región.

Si bien existen riesgos asociados a la inversión en estos fondos, como el riesgo de mercado, de crédito, de liquidez, de comisiones y de tasa de cambio, estos riesgos pueden mitigarse mediante la correcta gestión y diversificación de los activos por parte del PM. El objetivo de la gestión y diversificación es que los fondos sean capaces de resistir eventos inesperados y proteger el capital de los inversionistas.

El modelo de negocio de los fondos multiactivos de la empresa se basa en los principios descritos anteriormente, con un enfoque en la "*Current Management Fee*", siendo esta la lógica que define los incentivos para los gestores de cartera. El modelo de negocio hoy en día está apuntando a clientes de alto patrimonio, con un monto mínimo de inversión de \$200.000 dólares.

En el proceso de inversión de los fondos multiactivos, los PM's se apoyan mucho en las recomendaciones que provienen de tres fuentes principalmente (ver Ilustración 1.2-2).

- **Equipo de estrategia interno:** la empresa posee un equipo especializado que a partir de sus propios análisis y modelos elabora un vector de recomendación de inversión, que luego en reuniones periódicas se discuten con los PM's con la idea de intercambiar visiones sobre el panorama global.
- **Estrategia Sell Side Externos:** se basa en establecer acuerdos de colaboración con brokers. Estos acuerdos generalmente implican la compra de *securities* a estos, a cambio de recibir vectores de recomendación para los activos de los fondos multiactivos, así como investigaciones de inversión realizadas por el broker.

- **Estrategia Buy Side Externos:** se centra en la adquisición de servicios financieros de terceros, como por ejemplo análisis de inversión independientes, vectores de recomendación para los fondos multiactivos, asesorías financieras, entre otros.

A través de estas fuentes, el PM recopila información y a partir de sus propias convicciones la consolida en un vector que refleja la ponderación de los activos para cada fondo. El vector, que contiene las recomendaciones para los activos para el próximo período, se deriva de la maximización del *information ratio* (IR) producido por la asunción de riesgo previamente establecido en relación con el *benchmark* (tracking error).



Ilustración 1.2-2: Proceso de inversión en los fondos multiactivos de la empresa.

1.3 Portfolio Management

Para comprender mejor el rol del PM y la esencia detrás del *portfolio management*, es necesario comprender que esta disciplina consiste en la selección y supervisión de un conjunto de activos financieros con el fin de alcanzar los objetivos financieros a largo plazo y la tolerancia al riesgo de un cliente o grupo de clientes determinado[14]. Para lograrlo es fundamental que los *portfolios managers* (PM's) comprendan el perfil que poseen sus clientes juntos con sus metas, por ejemplo, una persona natural le interesa tener una buena jubilación por lo que el PM debiera de interesarse por comprender que estilo de vida aspira a tener el futuro, junto con el nivel de riesgo que está dispuesto a correr.

Los PM's desempeñan un papel crucial en la consecución de objetivos financieros. Su función principal consiste en administrar activos de manera estratégica para maximizar el retorno esperado y minimizar el riesgo inherente a las inversiones.

Es importante destacar que minimizar el riesgo no implica eliminarlo por completo. Las crisis financieras, por su naturaleza excepcional, pueden afectar negativamente a todos los mercados financieros, sin importar la estrategia de inversión empleada [15]. En este sentido, el rol de los PM's se centra en reducir la probabilidad de pérdidas dentro de un portafolio, optimizando la selección y ponderación de activos.

En general el proceso de *portfolio management* consta de tres pasos; planificación, ejecución y *feedback* o retroalimentación [14].

- **Planificación:** En esta etapa se comprenden las necesidades del cliente y se redacta lo que se conoce como IPS (*investment policy statement*), que es un documento que sirve como guía para gestionar las inversiones del cliente, más específicamente se definen los objetivos de las inversiones (plan de jubilación, generación de ingresos, acumulación de capital, entre otros), la tolerancia al riesgo, la estrategia de inversión junto a los activos a invertir (bonos, acciones, *real state*, entre otros), y por último las restricciones (temas regulatorios, pago de impuestos, rescates, etc).
- **Ejecución:** Aquí es donde el PM construye el portafolio del cliente basado en lo definido en el IPS. Armar el portafolio consiste en hacer lo que se conoce como “*Strategic Asset Allocation*”. esto incluye escoger en base a análisis de mercado y económico los activos que van a conformar la cartera al igual que sus respectivos pesos dentro de esta, con el objetivo de cumplir con los retorno y riesgo acordados con el cliente. Lógicamente el PM en esta etapa también debe hacer valer el principio de diversificación de los activos, industrias y regiones, ya que es una práctica que ayuda a reducir el riesgo no sistemático de la cartera, en suma, reducir el impacto negativo que pueda tener un activo por un mal rendimiento.

En el ámbito del *asset allocation*, existen dos metodologías de análisis predominantes: el enfoque *top-down* y el *bottom-up*. Cada una de estas estrategias se caracteriza por su particular secuencia de análisis y toma de decisiones.

El enfoque *top-down* comienza con un análisis macroeconómico y de mercado global. A partir de este análisis, se identifican los sectores, países o regiones con mayor potencial de crecimiento y rentabilidad. Posteriormente, se realiza un análisis más detallado de las empresas dentro de estos sectores o regiones seleccionados, con el objetivo de identificar las compañías que mejor se ajustan a los criterios de inversión establecidos.

En contraste, el enfoque *bottom-up* inicia con el análisis individual de empresas. Se evalúan las perspectivas de cada empresa en términos de su desempeño financiero, situación competitiva, equipo de gestión y otros factores relevantes. Una vez identificadas las empresas con mayor potencial, se procede a analizar el sector industrial y regional al que pertenecen, con el objetivo de evaluar el contexto macroeconómico en el que operan las empresas seleccionadas.

- **Feedback:** Una vez construido el portafolio, es necesario realizar lo que se conoce como “*Tactical Asset Allocation*”, es decir, monitorear y rebalancear los pesos de los activos según sea necesario. Esto se debe a las variaciones en el mercado que pueden afectar los rendimientos de ciertos activos o, en algunos casos, a que los activos no cumplan con las expectativas previstas, como podría suceder con las acciones de una empresa. Además, es fundamental evaluar el desempeño del portafolio para determinar si está alineado con las expectativas y objetivos del cliente.

Dentro de lo que se conoce como gestión o *management* de portafolios existen dos estrategias principales que los PM's suelen seguir; la gestión activa o *active management* y la gestión pasiva o *passive management* [14]. La primera es el tipo de gestión que predomina en el mercado y la que más rentabilidad a generado (ver ilustración 6-2) al menos hasta el 2017. Básicamente en la gestión activa los PM's buscan generar excesos de retorno respecto a un *benchmark* previamente definido, como lo puede ser por ejemplo el S&P 500. En caso contrario, la gestión pasiva no busca vencer a un *benchmark* sino que intenta replicar sus movimiento. Siguiendo esta misma línea es importante mencionar que la gestión de los fondos multiactivos en la empresa corresponde a una gestión activa.

Category	Assets (US\$ trillions)	Revenue (US\$ billions)	Market Share by Assets (%)	Market Share by Revenue (%)
Actively Managed	64	258	80%	94%
Alternatives	12	117	15	43
Active Specialties	15	55	19	20
Multi Asset Class	11	27	14	10
Core	26	59	33	21
Passively Managed	16	17	20%	6%
Total	80	275	100%	100%

Note: Some columns may not sum precisely because of rounding.
Source: Boston Consulting Group.

Ilustración 6-2: Distribución de las estrategias en el mercado.

2. Proyecto y justificación

La empresa está implementando una cuarta fuente de recomendaciones de inversión basada en modelos cuantitativos o "*Quants*". Estos modelos utilizan técnicas de análisis de datos, como regresiones lineales, regresiones multivariadas, *clustering*, optimización, *machine learning* e inteligencia artificial, entre otras. El objetivo de la empresa es aprovechar el potencial de estos modelos para mejorar sus procesos de inversión y generar valor adicional tanto para la empresa como para sus clientes.

En este contexto, el presente trabajo se centra en el desarrollo de uno de los dos modelos cuantitativos de *asset allocation* global que la empresa quiere implementar para complementar el proceso de inversión en los fondos multiactivos.

2.1 Justificación

En este punto, es natural preguntarse por qué es necesaria la implementación de análisis de datos junto con la revisión de múltiples fuentes para tomar una decisión. La respuesta tiene que ver con que la predicción de los movimientos del mercado es como intentar predecir el futuro, en suma, una tarea compleja. Si existiera una fórmula mágica para hacerlo todos serían ricos, pero ese no es el caso. Es por esta razón que año tras año, empresas de inversión, gobiernos, personas y otros actores invierten grandes sumas de dinero y tiempo intentando predecir qué sucederá en el futuro con el mercado y así aprovechar oportunidades y generar excesos de retornos, lo que se traduce en ganancias tanto para la empresa como para los inversionistas.

Esta creciente importancia de los datos a nivel mundial se refleja en el auge del mercado global de "big data". Según el informe "*Big Data-Strategic Business Report*" de 2022 [6], el mercado de big data alcanzó un valor de 154.000 millones de dólares y se prevé que para el año 2030 supere los 353.000 millones de dólares, duplicando su tamaño actual. Este crecimiento exponencial se ve impulsado por el aumento constante de la información disponible, como se ilustra en el gráfico (ver ilustración 2-1).

La disposición de grandes cantidades de información hace que el proceso de inversión y el mercado de valores se vuelvan cada vez más complejo. Por ende, el análisis cuantitativo emerge como una herramienta para apoyar y tomar mejores decisiones de inversión.

Lo anterior se ve reflejado en un artículo que publicó Bloomberg en donde consultaban a 16 altos cargos de empresas de inversiones como JPMorgan, BlackRock, AQR, Bernstein, entre otros, sobre como creían que sería la próxima década en las finanzas [7]. Todos de una u otra forma acordaban en que el análisis cuantitativo es algo que está en pleno desarrollo y que creará mucho valor significativo para las empresas.

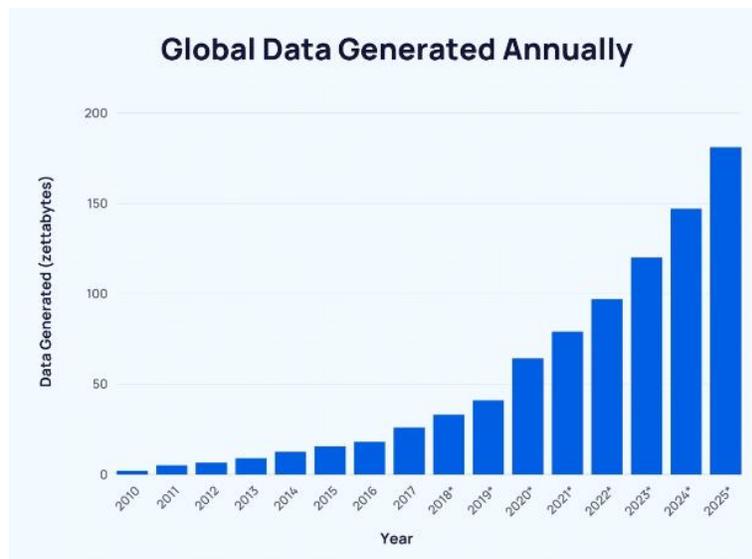


Ilustración 2-1: Cantidad de data creada, consumida y almacenada desde 2010 hasta 2020, con predicción a 2025 [8].
Nota: "zettabyte" (ZB) es una unidad de almacenamiento de información, que equivale a 10^{21} bytes.

En definitiva, la complejidad del mercado financiero actual exige un enfoque basado en datos y análisis cuantitativo para facilitar la toma de decisiones. La combinación de análisis de datos y la revisión de múltiples fuentes permite a los PM's tomar decisiones informadas y mejorar sus perspectivas de éxito. Además, debido al limitado número de analistas y al elevado volumen de información que debe analizarse, la implementación de este modelo contribuirá a optimizar el proceso. Asimismo, desarrollar un modelo que simplifique este proceso permitirá a la empresa ahorrar recursos, como poder computacional, tiempo, necesidad de capacitaciones, mantenimiento, entre otros.

A continuación, se proporcionará una comprensión más profunda sobre en qué consiste un proceso de inversión cuantitativa, junto con algunas de sus ventajas y desventajas.

2.2 Proceso Inversión Cuantitativa

Bajo la mirada del Banco BBVA [9], las finanzas cuantitativas son una disciplina que utiliza técnicas matemáticas, estadísticas, probabilísticas y algoritmos automatizados para analizar mercados financieros, estimar precios, calcular riesgos y predecir tendencias. Este último es lo que la empresa busca hacer con los fondos multiactivos para predecir tendencias y con ellas generar señales de mercado para realizar el *asset allocation* del portafolio del fondo.

Un detalle importante que posee este tipo de modelos es que funcionan sin la intervención humana, es decir, las decisiones que los modelos toman son 100% en base a los datos que se le entregan, por lo que se elimina todo tipo de juicios, experiencias, opiniones y emociones que los PM's puedan tener al momento de tomar decisiones y escoger variables. Esto no sustituye la necesidad de que un ser humano este monitoreando y actualizando el modelo, lo relevante es que en el proceso de toma de decisión el PM no interfiera.

El proceso cuantitativo de inversión se puede reducir en términos muy generales en tres grandes pasos según la CFI [10]:

1. ***Input System (Sistema de Entrada)***: En esta etapa se recolectan en forma de inputs todas variables que se quiere que el modelo utilice en su proceso. Esta puede ser data de mercado (PIB, inflación, sentimiento del consumidor, tasas de interés, entre otros), data de compañías (utilidades, EBITDA, deudas, *divided yield*, entre otros), data de encuestas, data geográfica, demográfica, entre otros. Una vez recolectada la data se realiza lo que se conoce como EDA (Análisis exploratorio de datos), a través del cual se busca comprender y visualizar los datos para identificar anomalías, realizar transformaciones a las variables, identificar patrones y tendencias iniciales, formular hipótesis, entre otras cosas. Esta es la etapa más importante y extensa del proceso, ya que con esto se va a alimentar el modelo. Por último, se debe elegir el o los algoritmos que se van a utilizar para realizar las predicciones. La elección del modelo debe ser de acorde al tipo de dato que se va a utilizar y a lo que se quiere estimar
2. ***Forecasting Engine (Motor de predicción)***: En esta etapa se le entregan al modelo las variables que se escogieron, para que se empiecen a ejecutar los algoritmos de predicción.

3. **Portfolio Construction (Construcción del Portafolio):** Lo que se hace es el *asset allocation* según los resultados de las predicciones, para ello se utilizan optimizadores y/o heurísticas que orientan al modelo en la ponderación o colocación de los activos para el próximo periodo en base a la estrategia que se quiera seguir para alcanzar los objetivos de riesgo y retorno previamente establecidos.

Algunos **pros** de los modelos *quant* son [11]:

- **Insensibles:** Como se mencionó anteriormente este tipo de modelo deja de lado todo tipo de sesgo que los PM's puedan tener a la hora de tomar decisiones, es decir, las decisiones son 100% en base a datos.
- **Económicos:** Las estrategias *quant* no necesitan tener a muchos analistas trabajando y analizando el mercado, basta con tener unos pocos cuanto que actualicen, perfeccionen y desarrollen los modelos.

Respecto a los **contras** [11]:

- **Vulnerabilidad a data manipulada:** Las decisiones que estos modelos toman están 100% basada en la data, por lo que si los inputs son incorrectos o están manipulados los resultados serán negativos o incorrectos.
- **Factores cualitativos:** Detrás de las empresas, gobiernos, instituciones, entre otros. Existen personas, por lo tanto, las decisiones que se toman y que impactan al mercado siempre van a estar influenciadas por factores cualitativos, que son difíciles de medir. Sin embargo, con el desarrollo de algoritmos de procesamiento de lenguaje natural la brecha cada vez será menor.

En base a lo expuesto, se procederá a presentar el objetivo general y los objetivos específicos para este proyecto.

3. Objetivo General

Desarrollar un modelo de *asset allocation* global que genere un vector de recomendación suma cero para fondos multiactivos a partir de data macroeconómica y de mercado, maximizando el *information ratio*.

4. Objetivos Específicos

1. Generar una base de datos suficientemente amplia que permita hacer el *asset allocation* de los diferentes tipos de activos.
2. Encontrar las especificaciones adecuadas que maximicen el *information ratio*, en un rango entre 0,4 y 0,6.
3. Desarrollar una metodología que ayuda a interpretar los resultados por medio de la clasificación de las variables por clases.
4. Documentar las metodologías utilizadas para el desarrollo y cumplimiento de los objetivos.

5. Alcances

Con este modelo cuantitativo la empresa busca tener un vector de recomendación global que se construya únicamente a partir de data macroeconómica y de mercado, con la intención de detectar tendencias y patrones que los PM's no ven, o alternativamente, confirmar las convicciones de posicionamiento que ya tienen. Con esto es relevante destacar que el modelo se utilizará como una herramienta de apoyo para la toma de decisiones, no como un sistema que tome e implemente decisiones por sí mismo.

El modelo por un lado no pretende utilizar técnicas sofisticadas de *Machine learning* o *Deep learning*, sino que intenta simplificar el proceso a través de un algoritmo que no sea una “caja negra”. De esta manera, es posible comprender la lógica detrás de las decisiones y los resultados detrás de las recomendaciones. Por otro lado, desarrollar un modelo que simplifique el proceso permitirá a la empresa ahorrar recursos, como, poder computacional, tiempo, necesidad de hacer capacitaciones, mantenimiento, entre otros.

Debido a limitaciones de tiempo, no será posible medir la eficacia y precisión del modelo en el período de prueba, ya que estos modelos requieren de un período de monitoreo para evaluar si efectivamente genera un alfa mayor a cero y mantiene un *information ratio* entre 0,4 y 0,6. Según lo conversado con la empresa este período de prueba podría durar entre 6 a 12 meses. Por ende, para evaluar la viabilidad de implementar el modelo sin tener que esperar 6-12 meses, se realizará un *backtesting* en la ventana de tiempo de datos que se tiene.

El modelo de *asset allocation* desarrollado en este trabajo se limita a activos de tipo global. Esto significa que el fondo está compuesto por instrumentos financieros emitidos por instituciones extranjeras y denominados en monedas extranjeras, principalmente dólares. Si bien existen fondos multiactivos locales, estos quedan fuera del alcance de este análisis debido a la complejidad adicional que introduce el riesgo moneda, entre otros factores. Cabe precisar que la denominación "local" se refiere a fondos en Colombia, México, Chile y Perú.

6.Marco Conceptual

A continuación, se presentan los principales elementos conceptuales a utilizar en este proyecto.

6.1 Series de Tiempo

Este tipo de serie se entiende como una secuencia de puntos que dependen del tiempo, es decir, los puntos se ordenan de manera sucesiva según el valor que tomaron en un instante de tiempo [12]. Si se grafican los valores de un activo (ver ilustración 6.1-1), se puede ver que cada valor de precio que tomó está asociado a un valor en el tiempo.

Las series de tiempo se utilizan para analizar cómo es que una variable o variables cambian o se comportan en el tiempo, esto es útil a la hora de analizar la data para hacer predicciones, detectar tendencias, estacionalidades, correlaciones, análisis técnico, detectar patrones, entre otras cosas. En el caso del presente informe se trabajó con índices económicos, ETF's, índices de mercado, entre otros. Por lo que el análisis de las series de tiempo es relevante debido a que, al trabajar con datos temporales, la aproximación, el procesamiento y la comprensión de los datos difieren en comparación con otras estructuras de datos.

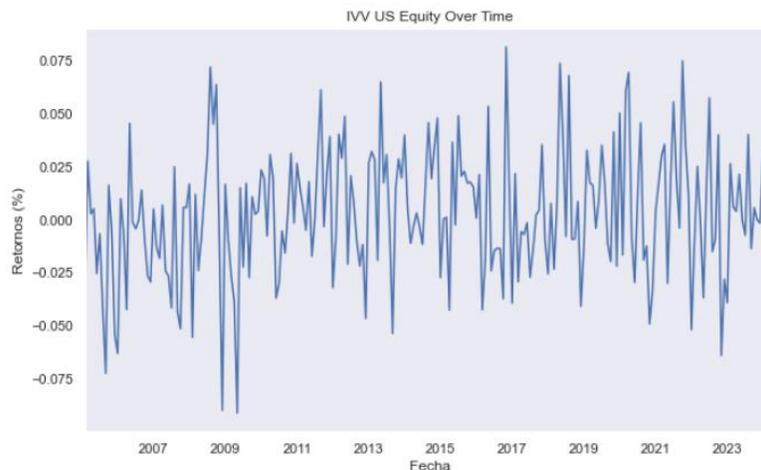


Ilustración 6.1-1: Serie de tiempo de IVV US Equity.

Las ventajas de trabajar con series de tiempo [13] y no con otro tipo de series que usan también en la industria, como los datos transversales, son;

- **Identificación de patrones y tendencias**, resultan útiles a la hora de realizar una predicción de valores futuros, entregando una noción de posibles direcciones de movimientos.

- **Análisis de ciclos**, identificar ciclos en los datos, como ciclos económicos, y estacionalidad en los datos, sirve para predecir, comprender e identificar patrones, en el corto, mediano y largo plazo.
- **Análisis de causalidad**, las series de tiempo pueden resultar útiles a la hora de detectar relaciones causales entre las variables.

Algunas desventajas que las series de tiempo poseen son [13]:

- **Sensibilidad a outliers**, el mal manejo de valores atípicos y del ruido en este tipo de series puede resultar perjudicial en los resultados que entreguen los modelos.
- **Datos faltantes**, tener valores nulos o faltantes podría perjudicar la precisión de los modelos, ya que como se mencionó, una de las características de las series de tiempo es que por un lado dependen del tiempo y el valor siguiente depende del valor anterior, por lo que tener datos faltantes rompe con la cadena de información.

6.2 Correlación de Pearson

Mide la relación lineal entre dos variables, es decir, mide que tan parecido se comportan dos variables en términos lineales, toma valores entre -1 y 1 [20]. Esta es la correlación clásica que se ve en estadística.

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

6.3 Correlación de Spearman

A diferencia de la correlación de Pearson esta es una medida de asociación no paramétrica, es decir, captura la relación no lineal en entre dos variables aleatorias [21]. Al igual que la correlación de Pearson toma valores entre -1 y 1.

$$\rho = \frac{6 \sum D^2}{N(N^2 - 1)}$$

D: es la diferencia de rangos para cada par de observaciones X e Y.

N: número de observaciones.

Tanto la correlación de Pearson como la de Spearman oscilan entre -1 y 1, lo que se interpreta como sigue:

- **Correlaciones cercanas a 1:** Indican una fuerte relación positiva entre la variable y el activo. En este caso, se asigna una ponderación alta (nota mayor)

a la variable, lo que significa que sus valores tienen un mayor impacto en el modelo de predicción.

- **Correlaciones cercanas a 0:** Indican una relación débil o nula entre la variable y el activo. En este sentido, se asigna una ponderación baja (nota nula) a la variable, lo que quiere decir que sus valores tienen un menor impacto en el modelo de predicción.
- **Correlaciones cercanas a -1:** Indican una fuerte relación negativa entre la variable y el activo. Por lo que se asigna una ponderación negativa (nota negativa) a la variable. Esto significa que, si la variable decrece, es probable que el activo suba, por lo que se conserva la ponderación negativa para capturar esta relación inversa.

6.4 Test de Causalidad de Granger

Esta herramienta se utiliza para evaluar la existencia de relaciones causa-efecto entre variables de series de tiempo [22]. Esta técnica no solo identifica si dos variables están correlacionadas, sino que también determina si una variable causa o influye a la otra. El test se realiza en dos pasos:

I. Especificación de los modelos:

$$\text{Modelo 1: } Y_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_n X_{t-n} + \varepsilon_t$$

$$\text{Modelos 2: } Y_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_n X_{t-n} + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \dots + \gamma_t Y_{t-q} + \varepsilon_t$$

Y_t : variable Y en tiempo t.

X_t : variable X en tiempo t.

α : constante.

β_i : coeficiente de regresión para los rezagos de X.

γ_i : coeficiente de regresión para los rezagos de Y.

ε_t : término de error.

- ### II. Prueba de hipótesis, test F. Se realiza el test F para determinar si los coeficientes β_i son conjuntamente distintos de cero, es decir, se busca rechazar la hipótesis nula, ya que al rechazarla a un nivel de significancia aceptable (por ejemplo al 0,05) se puede determinar que estadísticamente X posee un efecto causal sobre Y.

El p-valor que se obtiene a partir del test F indica la probabilidad de rechazar la hipótesis nula, por ende, se utilizará este valor como ponderador para establecer el nivel de relación entre las variables.

6.5 PCA (*Principal Component Analysis*)

El Análisis de Componentes Principales (PCA), también conocido como Principal Component Analysis en inglés, es una técnica de reducción de dimensionalidad ampliamente utilizada en diversos campos, incluyendo el análisis financiero y económico. Su objetivo principal consiste en transformar un conjunto de variables originales en un nuevo conjunto de variables no correlacionadas y ortogonales, denominadas componentes principales, que capturan la mayor variabilidad posible del conjunto original [15].

La esencia del PCA radica en la idea de que la información relevante dentro de un conjunto de variables se encuentra principalmente en su variabilidad. En otras palabras, los cambios en los valores de los indicadores económicos y macroeconómicos son los que aportan información significativa para comprender el comportamiento del sistema. El PCA aprovecha esta premisa para identificar las componentes principales que representan la mayor parte de la variación presente en el conjunto de datos original.

Cada componente principal explica un porcentaje de varianza de los datos, ordenándose de mayor a menor, o sea que el componente principal 1 (PCA 1) posee el mayor porcentaje de variabilidad de conjunto inicial, PCA 2 tiene la segunda mayoría y así sucesivamente hasta que se logre captar el porcentaje de varianza deseado.

Algunos de los beneficios [16] que involucra la utilización de componentes principales es que, por un lado, se maximiza la cantidad de información que estos contienen en el sentido que se captura un gran porcentaje de la variabilidad en unas pocas variables lo que a su vez hace más ligero el modelo en términos de poder computacional, debido a que se elimina información poco relevante. Por otro lado, al generar nuevas variables no correlacionadas se elimina la multicolinealidad.

La principal desventaja de usar componentes principales [16] radica en la interpretabilidad de las variables, ya que al ser combinaciones lineales de todo el conjunto de datos al momento de interpretar y explicar los resultados de manera directa se dificulta, lo que puede traer problemas para entender lo que está ocurriendo. Esto representa una dificultad para el desarrollo del presente modelo, debido a que no solo se requiere generar un *asset allocation*, sino que también comprender cómo es que el modelo toma las decisiones. A pesar de lo anterior se utilizará esta técnica de igual manera porque existen algunas técnicas que se verán más adelante que ayudan a dar cierta interpretación a los componentes principales.

6.6 Teoría Moderna de Portafolio (Markowitz)

Esta teoría, desarrollada por Harry Markowitz en 1952, busca a través de un enfoque matemático crear una cartera diversificada de inversiones que maximice el retorno esperado para un nivel de riesgo dado [6]. Esta resulta ser la teoría base a partir de la cual se desprenden los desarrollos de otros modelos. Algunos de los supuestos más relevantes de la teoría son:

- **Inversores racionales buscan maximizar su utilidad**, esto quiere decir que los inversionistas prefieren activos con mayor retorno, pero con menor riesgo. Matemáticamente esto se puede traducir a una utilidad que se describe como sigue:

$$U_p = E(R_p) - \lambda\sigma_p^2$$

U_p : Utilidad esperada del inversor.

$E(R_p)$: Retorno esperado del portafolio.

λ : Medida de la aversión al riesgo del inversionista.

σ_p^2 : Desviación estándar de los retornos del portafolio (riesgo).

Si se crean distintas combinaciones de retornos esperados y riesgo y se unen los puntos entre sí, se da forma a lo que se conoce como curva de indiferencia (ver ilustración 6-2), que representa de manera visual las distintas combinaciones de riesgo y retorno esperado que le entregan un nivel de utilidad óptimo al inversionista.

- **Los retornos de los activos se distribuyen de manera normal**, por lo que el rendimiento esperado del activo corresponde a la media, la varianza mide la volatilidad y la desviación estándar refleja el riesgo.
- **Los activos no son perfectamente correlacionados.**
- **El mercado es eficiente.**
- **Horizonte de inversión de un solo período.**

Con estos supuestos un PM puede construir un portafolio a partir de los activos o *asset classes* que hayan sido definidos en el IPS, ya que estos activos deben de cumplir tanto los objetivos de inversión como las restricciones. Si a esto se le suman las expectativas de mercado y las correlaciones entre los activos, el retorno esperado del portafolio se calcula como [6]:

$$E(R_p) = \sum_{i=1}^n w_i E(R_i)$$

Donde w_i corresponde al peso que se le asigna a cada activo dentro del portafolio, y el riesgo se calcula como:

$$\sigma_p = \sqrt{\sum_{i=1}^n \sum_{j=1}^n w_{p,i} w_{p,j} Cov(R_i, R_j)}$$

Donde $Cov(R_i, R_j) = \rho_{i,j} \sigma_i \sigma_j$, siendo $\rho_{i,j}$ la correlación entre el activo i y el activo j.

Finalmente, si se traza una línea que una los portafolios factibles en donde se tenga el máximo retorno posible a un cierto nivel de riesgo, se construye lo que se conoce como frontera eficiente. Esta representa las combinaciones óptimas de activos a distintos niveles de riesgo (ver Ilustración 6-1).

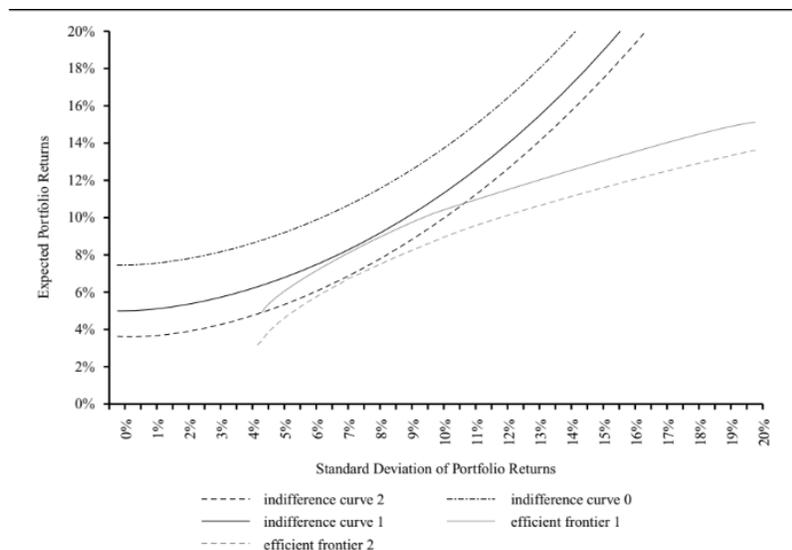


Ilustración 6.6-1: Representación de fronteras eficientes y curvas de indiferencia para dos activos [6].

Al observar las fronteras eficientes junto con las curvas de indiferencia, se puede ver que se interceptan en un punto. Ese punto (punto de tangencia) corresponde al óptimo *asset allocation* para el cliente o inversionista [6]. Además, es relevante considerar que las fronteras eficientes se pueden mover según como sean las expectativas de mercado, dando origen a la frontera eficiente 2 que al mismo tiempo se asocia con un nivel de utilidad esperado menor, lo que conlleva a un nuevo punto de tangencia y por ende un nuevo *asset allocation* (ver ilustración 6.6-2).

El enfoque de Markowitz es útil tanto en la construcción de un portafolio como en la creación de un índice, ya que esta teoría proporciona una perspectiva estratégica de largo plazo al diseñar un portafolio o un índice ajustado por riesgo.

Siguiendo esta línea, dado que el portafolio se gestiona activamente, se compara constantemente con un *benchmark* (por ejemplo un índice), a partir del cual se calculan diversas métricas de éxito, que se detallarán a continuación en esta sección. El ajuste mensual correspondiente a la gestión activa del portafolio se realiza a través de un *tactical asset allocation* (TAA), lo que permite que el modelo cumpla con su rol táctico, maximizando el alfa y, como consecuencia, el *information ratio*.

Algunas de las razones por las cuales la teoría de Markowitz no resulta adecuada, y por las que no se utilizó en este modelo, son las siguientes:

1. Los fondos multiactivos gestionados por la empresa son fondos públicos, lo que significa que cualquier persona, ya sea natural o jurídica, puede invertir en ellos. Esto implica que el *Investment Policy Statement* (IPS) es estandarizado para todos los inversores, y no solo incluye información sobre los retornos esperados y rangos de riesgo, sino también restricciones adicionales, como liquidez, diversificación y requisitos mínimos de ingreso, entre otras. El modelo de Markowitz, al enfocarse exclusivamente en la relación entre riesgo y retorno esperado, no contempla estas restricciones adicionales.

2. Se busca incorporar factores externos en la toma de decisiones, más allá del análisis histórico de los activos. El objetivo era integrar variables macroeconómicas y de mercado que influyen en los rendimientos de los activos, lo que proporciona una visión más amplia y ajustada a las condiciones económicas actuales.

6.7 Alfa

Alfa es un término utilizado en el ámbito del *portfolio management* para medir el rendimiento de una cartera de inversiones respecto a un índice de referencia o *benchmark* [17]. En otras palabras, mide el exceso de retorno que genera una cartera de inversiones en comparación con un *benchmark*. En el caso de los fondos multiactivos, el PM busca generar este exceso de retorno mediante una gestión activa de la cartera, lo que representa el valor agregado que el gestor de portafolios ofrece al cliente por su gestión.

6.8 Tracking Error (TE)

Término utilizado para medir el nivel de divergencia que está teniendo el portafolio respecto al precio del *benchmark*. También puede ser comprendido como el nivel de riesgo que el PM está tomando al momento de gestionar el fondo, y se evalúa como la desviación estándar de los retornos que posee el portafolio respecto al *benchmark* a lo largo del tiempo [18].

$$TE = std(P - B)$$

El valor “P” corresponde a los retornos del portafolio y “B” corresponde a los retornos del benchmark.

6.9 Information Ratio (IR)

El *Information ratio* es la unificación de los conceptos de alfa y *tracking error*, ya que mide como han sido los retornos del portafolio respecto al *benchmark* pero ajustado por el nivel de riesgo o volatilidad de los retornos, es decir, IR estandariza el alfa o exceso de retorno por el riesgo [19]. Esta medida también puede ser comprendida como el nivel de habilidad que posee el PM para generar exceso de retorno. Se calcula como se ve a continuación:

$$IR = \frac{P - B}{TE}$$

“P” representa los retornos del portafolio, “B” representa los retornos del *benchmark* y “TE” el *tracking error* hasta la fecha.

6.10 Look Ahead Bias

Es considerado uno de los sesgos más comunes a la hora de trabajar con modelos de series temporales, tanto en la fase de entrenamiento como en el *backtesting*. Se produce cuando se incorpora información en el análisis que no estaba disponible o no era conocida durante el periodo de estudio [23].

6.11 Metodología CRISP-DM

Dentro de la ciencia de datos, existen diferentes estructuras o metodologías para el desarrollo de proyectos de estas características. En este caso, se optó por el método CRISP-DM, que es la abreviación de *Cross-Industry Standard Process for Data Mining*. Esta metodología fue seleccionada para guiar el desarrollo del proyecto debido a su carácter general y amplia aplicabilidad en el ámbito de la ciencia de datos. Además, proporciona un marco sólido para abordar proyectos de análisis de datos de manera sistemática y eficiente.

CRISP-DM consta de 6 fases (ver ilustración 6-1) las cuales se usarán como referencia para el desarrollo de este proyecto [24], y que además se profundizarán a continuación.

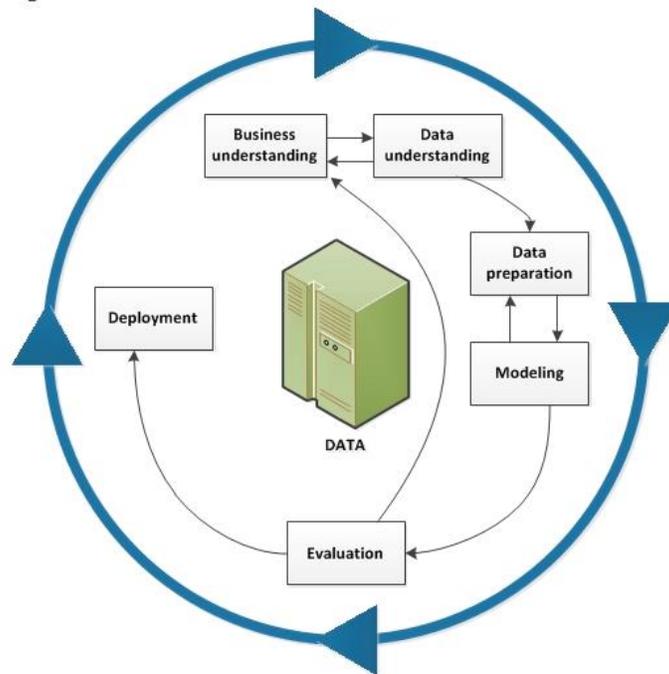


Ilustración 6-1: Diagrama metodología CRISP-DM [24].

Fase 1 (*Business Understanding*): En esta fase, el propósito principal es comprender los objetivos y requisitos del proyecto desde una perspectiva empresarial. Esto se logra mediante la evaluación de la situación, en la cual se recopila información sobre los recursos disponibles, las restricciones y los factores relevantes para el proyecto.

Fase 2 (*Data Understanding*): Aquí se debe lograr tener una comprensión inicial de los datos y su calidad. Esto se logra mediante la recopilación de datos disponibles, descripción de los datos, aplicación de análisis exploratorios y verificación de la calidad de los datos.

Fase 3 (*Data Preparation*): Como el nombre lo indica, en esta fase se preparan los datos de manera que puedan ser utilizados en el modelado. Algunas de las tareas que se realizan son; selección, limpieza, construcción, integración y formateo de los datos.

Fase 4 (*Modeling*): En esta etapa se emplean técnicas de modelado para identificar patrones en los datos y extraer información valiosa que contribuya al logro de los objetivos planteados.

Fase 5 (*Evaluation*): El objetivo de esta etapa es evaluar el rendimiento del modelo y su relevancia para el cumplimiento de los objetivos del negocio inicialmente propuestos.

Fase 6 (*Deployment*): Corresponde a la implementación del modelo en un entorno de producción para su uso en la toma de decisiones. Sin embargo, esta fase no se abordará en este proyecto debido a que excede los plazos previstos. Implementar un modelo de estas características requiere una ventana de evaluación futura considerable para determinar si efectivamente genera alfa en la práctica.

Es importante mencionar que el desarrollo de un modelo de estas características implica volver constantemente a las fases anteriores, ya que se pueden encontrar fallos en las hipótesis iniciales o mejores maneras de completar las fases. Por esta razón, en la ilustración 6-1 se observa que los desarrollos de estos proyectos son cíclicos, debido a que se regresa constantemente a las fases anteriores para realizar ajustes.

7. Metodología

El proyecto se centra en el desarrollo de dos modelos de *asset allocation* para fondos multiactivos. Cada uno de estos modelos se pueden considerar como subproyectos de los cuales uno de ellos es el que se va a desarrollar en el presente proyecto de título, mientras que el que utiliza técnicas de *machine learning* será desarrollado por la contraparte.

La idea detrás del modelo es generar un *scoring* a partir de datos macroeconómicos y de mercado que indique el nivel de convicción sobre la realización de un posicionamiento *overweight* o *underweight* para el siguiente mes. Cabe destacar que la construcción de este *scoring* se fundamenta en un enfoque experimental e iterativo, lo que implica una serie de pasos en cada fase de la metodología CRISP-DM (ver ilustración 7-1), los cuales se detallarán a lo largo de esta sección.

Fase 1 (Business Understanding)	Fase 2 (Data Understanding)	Fase 3 (Data Preparation)	Fase 4 (Modeling)	Fase 5 (Evaluation)
Antecedentes generales del negocio.	Recopilación de los datos y clasificación por clase: 1- Cyclical 2- Financial Conditions 3- Monetary Conditions 4- Technicals 5- Sentiment 6- Valuation	Estandarización variables (z-score).	Especificaciones para el modelo: 1- Tipo de ventana. 2- Normalizar (z-score) 3- Significancia correlación. 5- Correlación Pearson. 6- Correlación de Spearman 7- Test Causalidad de Granger 8- Conjunto Training 9- Conjunto Test	Restricción del scoring obtenido.
Industria.		Rezago datos en un mes.		Asset Allocation a partir del scoring, siguiendo una estrategia top-down.
Contextualización de la empresa.	Reducción del conjunto inicial de variables (muestreo).	Agrupación de los datos por clase utilizando un promedio simple.	Definición de caminos alternativos: 1- Agrupación por clase (promedio simple). 2- Todas las variables. 3- Componentes principales (PCA)	Interpretabilidad resultados.
Justificación proyecto.		Agrupación datos utilizando componentes principales (PCA).		Medición del Performance.
Objetivo general y específicos.	Evaluación representatividad de la muestra.		Entrenamiento del Modelo: 1- División data (Train y Test). 2- Filtro por correlación significativa (Train). 3- Generación de IR de cada activo por cada variable (Train). 4- Filtro por IR's positivos (Train). 5- Weight's por cada IR positivo (Train).	Backtesting -> Resultados
Alcances.	Categorización Activos: 1- Desarrollado 2- Emergente 3- Renta Fija High Beta 4- Renta Fija Low Beta 5- Caja		Generación de scoring, como resultado de recomendación del modelo.	

Ilustración 7-1: Diagrama de los pasos realizados por cada una de las fases de la metodología.

Fase 1 (Business Understanding): En esta fase se comprendieron los problemas, necesidades y objetivos de la empresa en relación con el modelo a desarrollar, identificándose la necesidad de generar por parte de la empresa sus propios vectores de recomendación cuantitativos. El desarrollo de esta fase se da por concluida con la presentación de los antecedentes generales, incluyendo una discusión sobre el negocio de los fondos multiactivos y su propuesta de valor, así como la definición de los objetivos generales y específicos.

Fase 2 (Data Understanding): En la selección y comprensión de la data, se trabajó en conjunto con la empresa para definir qué tipo de variables podrían necesitarse para el desarrollo de un modelo de *asset allocation* para fondos multiactivos. Se establecieron 6 clases (*Cyclical*, *Financial Conditions*, *Monetary Conditions*, *Technicals*, *Sentiment* y *Valuation*), que luego se introdujeron en el buscador de Bloomberg. A medida que las variables coincidían con los campos correspondientes, se iban almacenando por nombre en un Excel. Las clases que se definen como sigue:

- **CYCLICAL (cíclicas):** Estas son variables que en términos muy generales se ven afectadas o influenciadas por los ciclos económicos (ver ilustración 7-2).
- **FINANCIAL CONDITIONS (condiciones financieras):** Corresponden a índices de condiciones financieras de países, diseñadas para capturar de forma general el estado de los mercados financieros. Algunos de los elementos que se considera al momento de construir estos índices son tasas de interés, diferencias de crédito, precios de *equity*, volatilidad del mercado, entre otros (ver ilustración 7-2).

- **MONETARY CONDITIONS (condiciones monetarias):** Hace referencia al estado de la oferta y demanda de la moneda junto con políticas monetarias. Por lo general son índices de inflación a nivel de países, índices de *yields* e índices de créditos (ver ilustración 7-2).
- **TECHNICALS (técnicos):** Cómo su nombre lo indica son variables más de tipo técnicas, o sea variables que surgen a partir de la data como, por ejemplo, índices de volatilidad de algún sector o país, momentum, entre otros (ver ilustración 7-2).
- **SENTIMENT (sentimiento):** Este tipo de variables captan los estados de ánimo del mercado, así como también la actitud en general que están teniendo los inversionistas en cuanto a expectativas en el corto, mediano y largo plazo. Esto se puede medir a través de encuestas las cuales transforman los resultados a un índice (ver ilustración 7-2).
- **VALUATION (valuación):** Son variables que buscan reflejar el valor intrínseco de un activo, típicamente sobre alguna compañía o *security*. Algunos ejemplos de este tipo de variables son ratios como PE (*Price-to-earnings*), PB (*Price-to-book*), *price-to-free cash flow* sobre algún índice, que podría ser el S&P 500 (ver ilustración 7-2).

CLASE	TICKER	NOMBRE
CYCLICAL	NAPMNEMP Index	ISM Services PMI Report on Business Employment SA
	NAPMNNO Index	ISM Services PMI Report on Business New Orders SA
	OEUSKLAR Index	USA OECD Leading Indicators CLI Trend Restored SA
	MPMICNMA Index	Caixin China Manufacturing PMI SA
FINANCIAL CONDITIONS	GSCNFCI Index	GS China Financial Conditions Index
	GSINFCI Index	GS India Financial Conditions Index
	KCFSINDX Index	United States Financial Stress Index
MONETARY CONDITIONS	CNCPIYOY Index	China CPI YoY
	GRCP20YY Index	Germany CPI All Items YoY
	CCOSREVL Index	Federal Reserve Consumer Credit Outstanding Amount Revolving SA
TECHNICALS	RSI_30DIVV US EQUITY	RSI 30 days IVV US Equity
	RSI_14DEZU US EQUITY	RSI 14 days EZU US Equity
	RSI_14DEWJ US EQUITY	RSI 14 days EWJ US Equity
SENTIMENT	SNTEEUGX Index	sentix Economic Indices Euro Aggregate Overall Index on Euro area
	SNTEASPO Index	sentix SentiMent Asia Ex Japan Economy Current Assessment Private Investor
	SNTELAIO Index	sentix SentiMent Latin America Economy Current Assessment Institutional Inves
VALUATION	PE SPX INDEX	PE SPX INDEX
	PB SPX INDEX	PB SPX INDEX
	PX_TO_FREE_CASH_FLOW SPX INDEX	PX_TO_FREE_CASH_FLOW SPX INDEX

Ilustración 7-2: Ejemplos de variables por cada clase.

Cada una de estas clases abarca un amplio espectro de variables, llegando a recopilar más de 1000. Por lo tanto, fue necesario extraer una muestra que fuera lo más representativa posible. El proceso para obtenerla fue el siguiente: en primer lugar, se seleccionaron de manera arbitraria todas las variables que, al momento de ser extraídas de Bloomberg, contaban con una métrica denominada '*relevance value*'. Esta métrica representa la cantidad de alertas que los usuarios de Bloomberg han configurado para un evento económico, en comparación con todas las alertas establecidas para eventos en esa región. En términos simples, se puede interpretar como una medida que indica el nivel de interés que los inversionistas tienen en ese índice en particular. A partir del conjunto inicial de 1000 variables, se extrajo una muestra de más de 250 variables.

El segundo paso fue evaluar si la primera muestra era representativa del conjunto original. Para ello, se examinó la distribución de otros campos asociados a las variables, como por ejemplo, los países. Si se detectaba algún desequilibrio, se ajustaba la muestra manualmente, agregando o eliminando las variables necesarias para lograr un equilibrio en la distribución. Una vez concluido este proceso, se logró reducir el universo inicial de más de 1000 variables a un total de 263.

Los activos que posee el fondo multiactivo, y a partir de los cuales se realizó el *asset allocation* son los siguientes:

Asset Allocation	Selectividad (estrategia)	ETF's
Desarrollados	Estados Unidos	IVV US Equity
	Europa	EZU US Equity
	Japón	EWJ US Equity
Emergentes	Asia Emergente	AAXJ US Equity
	Latinoamérica	ILF US Equity
Renta Fija Bajo Beta	Treasuries Short Duration	SHY US Equity
	TIPS Short Duration	TIPS Corto
	Treasuries Long Duration	IEI US Equity
	TIPS Long Duration	TIPS
Renta Fija Alto Beta	Investment Grade Short Duration	SLQD US Equity
	Investment Grade Long Duration	LQD Us Equity
	High Yield Short Duration	SHYG US Equity
	High Yield Long Duration	HYG US Equity
	EMD LC	EMB US Equity
	EMD HC	EML US Equity
Caja	Caja	SHV US Equity

Ilustración 7-3: Categorización de los activos del fondo a partir de los cuales se genera el *asset allocation*.

Como se puede observar los activos que conforman el fondo corresponden a activos de tipo global que involucran distintos tipos de riesgo. Por ende, lo que define un fondo como más o menos riesgoso es la ponderación que se tengan de los activos.

Fase 3 (*Data preparation*): Para mejorar el entendimiento de las acciones realizadas y establecer una mejor conexión entre las fases 3 y 4, se elaboró el diagrama de la ilustración 7-4. Este diagrama representa de manera más clara y estructurada los pasos y el flujo seguido en la construcción del modelo.

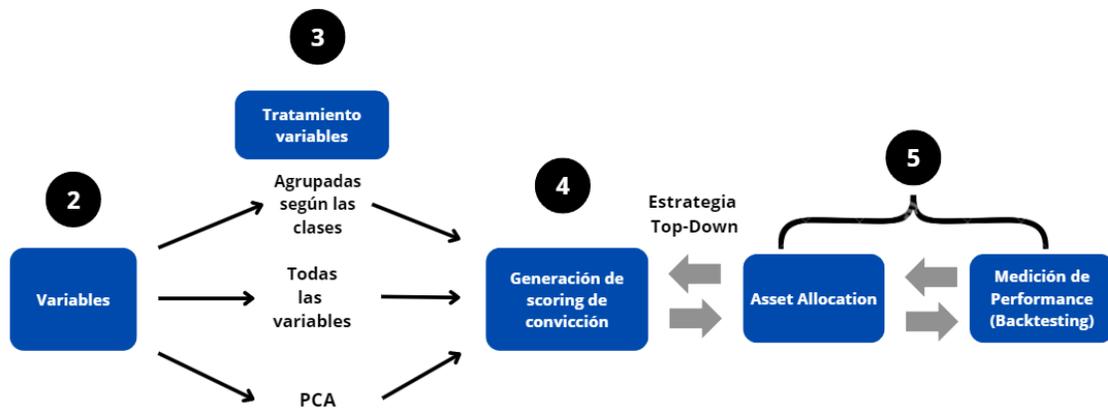


Ilustración 7-4: Diagrama específico de propuesta de construcción del modelo desde la fase 2 a la fase 5.

Como se vio en la ilustración 7-1, la primera parte de la preparación de los datos consistió en estandarizar las variables utilizando z -score. El cálculo de la varianza y la media para el z -score depende de si se utiliza una ventana fija o móvil. Como se observa en la ilustración 7-1, dentro de la fase 4 se formularon especificaciones en el modelo que permiten optar por trabajar con una ventana móvil o fija. La ventana fija se refiere a que la media y la varianza se calculan cada vez que se agregan nuevos datos al modelo y se vuelve a entrenar. Por otro lado, la ventana móvil implica trabajar con un conjunto establecido de datos, por ejemplo, datos de 3 años. Por lo tanto, cada vez que se reentrena el modelo, se consideran datos de 3 años hacia atrás desde la fecha de reentrenamiento.

En la segunda parte, se rezagaron todos los datos en un mes con el objetivo de evitar el sesgo conocido como *look ahead bias*. Esto se debe a que algunos datos, como el de inflación, solo se conocen dentro del mes siguientes al que se está evaluando. Por lo tanto, se optó por utilizar los datos del mes anterior respecto al mes de evaluación para calcular el *scoring*.

La tercera parte de la preparación de los datos consistió en agrupar las variables según cada clase utilizando el promedio simple entre los valores normalizados de todas las variables. Esto con el fin de crear el primer input establecido en el diagrama de la ilustración 7-4. Se aplicaba la misma lógica mencionada en el párrafo anterior cuando se estaba trabajando con una ventana móvil o fija.

La cuarta y última parte consistió en construir el tercer tipo de input establecido en la ilustración 7-4, que corresponde a aplicar componentes principales (PCA) a las variables. Al igual que con el z -score, componentes principales se aplicará dependiendo de si se está evaluando el modelo con ventanas fijas o móviles, por ello cada vez que se reentrenaba el modelo se volvían a calcular los componentes principales de los datos que se tenían hasta la fecha. El número de componentes principales a utilizar se definió de manera experimental, ya que no existe a priori un número establecido de componentes principales a utilizar en el modelo.

Fase 4 (Modeling): Previo a la ejecución del algoritmo, se establecieron los hiperparámetros o especificaciones (ver ilustración 7-5), que coloquialmente podrían describirse como las "perillas" utilizadas en el modelo para ajustarlo y maximizar la generación de alfa y, por consiguiente, del *information ratio*. La idea es probar diferentes combinaciones de estas

especificaciones bajo una misma base para determinar cuál resulta mejor. Para ello, se definen variables booleanas (True o False) que activan o desactivan cada especificación. Así, se le indica al algoritmo si debe utilizar una ventana fija o móvil, normalizar las variables, aplicar un filtro de correlación o causalidad específico, calcular componentes principales, determinar el número de componentes principales a incluir, agregar las variables por clase, y finalmente, definir el tamaño del conjunto de entrenamiento y de prueba.

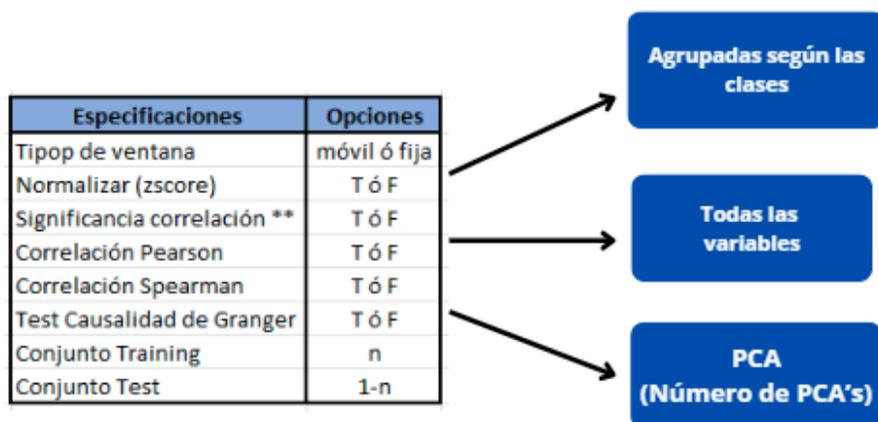


Ilustración 7-5: Diagrama sobre cómo se construyeron las especificaciones para luego combinarlas con uno de los tres caminos propuestos. (**Significancia de correlación filtra las variables según si su correlación con el activo es estadísticamente significativa o no).

Como se observa en las ilustraciones 7-4 y 7-5, el flujo del proceso se divide en tres vías, a las cuales se referirá como caminos. Estos caminos juegan un rol experimental e iterativo dentro del modelo, al igual que las especificaciones, ya que se busca probar tres formas distintas de entregar los datos al algoritmo para que este tome sus decisiones respecto a los posicionamientos de los activos. El primer camino agrupa todas las variables en las seis clases mencionadas anteriormente utilizando un promedio simple. La segunda vía utiliza las 263 variables. Y el tercer camino agrupa las variables utilizando componentes principales.

Luego de definir las especificaciones y caminos del modelo, se establece un proceso iterativo para cada uno de los activos. En este proceso, se toma de forma individual cada activo del *dataset* con el objetivo de generar su *scoring* o puntaje de convicción para cada mes, tal como se ilustra en la Ilustración 7-4 posterior al tratamiento de las variables.

Los periodos de reentrenamiento del modelo por *default* se fijaron en dos meses, aunque más adelante se probaron otros periodos de entrenamiento, siendo el de dos meses el con mejor generación de *information ratio* y por consiguiente alfa.

Se definió el conjunto "X" como el conjunto de variables e "Y" como el conjunto de activos. De cada conjunto, se definió un conjunto de *test* y otro de *train* en base a lo que se mencionó anteriormente, es decir, la cantidad de filas en el conjunto de test se fijó en dos (ver ilustración 7-6).

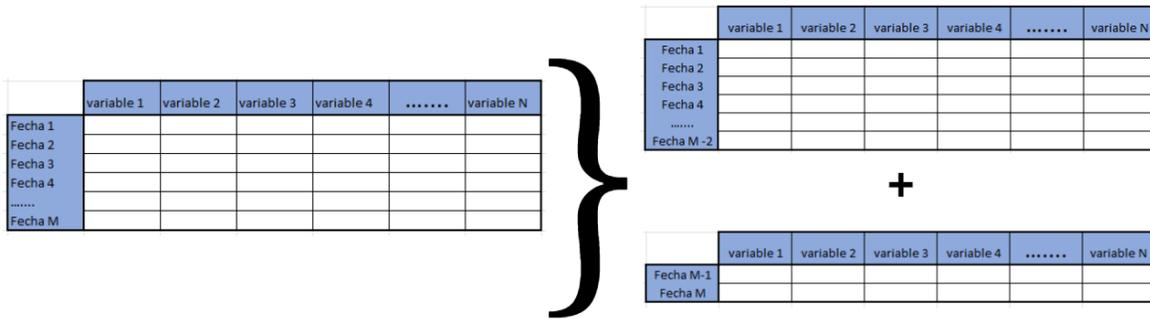


Ilustración 7-6: Diagrama separación en *Train* y *Test*.

El entrenamiento del modelo consiste en los siguientes pasos:

1. Como se mencionó anteriormente se creó un ciclo para evaluar cada activo de manera individual, por lo que se calcularon las correlaciones de cada activo con cada una de las variables, filtrándose por ende las variables que no fueran significativas, dando origen a los primeros parámetros del modelo. El nivel de significancia se midió utilizando el p-valor, que es una medida de la probabilidad de obtener un resultado igual o más extremo que el observado en el estudio, suponiendo que la hipótesis nula es cierta, por lo tanto, si el p-valor era menor a 0,05 se considera que existe la evidencia suficiente para rechazar la hipótesis nula, siendo por consiguiente la correlación significativa.

Existen distintas maneras de calcular la correlación, para este caso se consideraron dos, correlación de Pearson y Spearman, más una tercera vía que es un *test* de causalidad, el test de causalidad de Granger. El objetivo de este filtrado es eliminar aquellas variables que, a primera vista, no parecen tener una influencia estadísticamente significativa en los valores que toman los activos. De esta manera, el enfoque apunta a identificar las variables que realmente aportan información para la predicción de potenciales retornos futuros de los activos.

Cabe destacar que puede darse el caso en que ninguna variable resultara significativa, en esos casos se consideraban todas las variables como si fueran significativas pasando al paso siguiente.

2. Una vez identificadas las variables con correlaciones significativas, se procede a multiplicar cada valor de estas variables dentro de la ventana de tiempo en evaluación por su respectiva correlación. El objetivo de esta multiplicación es ponderar el valor de cada variable en función de la fuerza de la relación que presenta con el activo.

3. Tras identificar las variables con correlaciones significativas, se procedió a calcular el segundo parámetro del modelo, el *information ratio*. La idea de obtener este parámetro es evaluar qué tan relevante ha sido cada variable en la generación de retornos para cada activo.

El IR se calculó como el alfa acumulado generado por la variable durante la ventana de evaluación dividido por la desviación estándar (*tracking error*) del alfa en dicho período. Cuando se habla de “*alfa acumulado generado por la variable*”, se refiere a que se multiplicaron los retornos de los activos por el ponderador obtenido en el paso anterior. Esto da como resultado un tipo de “retorno ponderado”. Por ende, el IR indica el beneficio neto por unidad de riesgo asumido al utilizar una determinada variable para predecir el comportamiento del activo.

4. Se aplicó un segundo filtro eliminando las variables que generaran un IR negativo, ya que se interpreta como que el beneficio neto por unidad de riesgo que generaba las variables en evaluación sobre el activo era negativa, o sea que no generaba alfa.
5. A partir de los IR positivos obtenidos anteriormente, se calcularon los pesos que tiene cada variable sobre el activo en evaluación en comparación al resto de las variables. Esto se hizo de la siguiente manera:

$$w_i = \frac{IR_i}{\sum_{i=1} IR_i}$$

En caso de que todos los IR resultaran negativos, se procedía a darle a todas las variables el mismo peso por igual.

Una vez obtenidos los parámetros y las variables sobrevivientes del modelo, estos parámetros (correlación y w_i) se multiplican por los valores correspondientes del conjunto de test. Los resultados de esta multiplicación se suman para cada fila, generando un valor agregado de convicción para cada activo (ver Anexo A).

Posteriormente, se realiza un ranking utilizando percentiles, asignando dos posibles valores: 0,5 o 1 (ver Anexo B). La asignación de estos valores depende de la comparación entre el valor del mes actual y el valor del mes anterior. Un valor de 0,5 se asigna si el valor del mes anterior es menor que el valor actual, mientras que un valor de 1 se asigna en caso contrario. En otras palabras, considerando que el conjunto de test contiene dos valores para cada activo (el valor del mes actual y el valor del mes anterior), se utiliza la técnica de percentiles para determinar si el valor del mes anterior es superior o inferior al valor del mes actual (ver Anexo B).

El valor del mes de evaluación, que corresponde al *scoring*, se va guardando de manera aislada en un *dataframe* aparte, con la finalidad dejar registro de los *scoring's* que entregaba el modelo para luego hacer el *asset allocation* de los activos.

Fase 5 (Evaluation): Teniendo el *scoring* se procede a generar el *asset allocation*, para ello, se siguió una estrategia *top-down*. Primero se realizó el *asset allocation* de los activos agrupados como renta variable (RV), renta fija (RF) y Caja. Por un lado, se filtró RV y RF para que los valores de los *scoring's* fueran mayores a 0,3 y menores a 0,7. Por otro lado, Caja se filtró entre 0,45 y 0,55. Estas restricciones se implementaron para evitar que el modelo se concentre en un solo activo, manteniendo un nivel de diversificación.

Posteriormente, se ajustaron los *scoring's* para cada fila de manera que la suma de RV, RF y Caja fuera igual a 1. Para lograr esto, se normalizó cada valor dividiéndolo por la suma total de la fila correspondiente. Una vez realizado esto, se restó a cada *scoring* el equivalente a una posición *equally weighted* ($w_i - 1/3$), esto con el fin de capturar las diferencias en los posicionamientos entre el *benchmark* (*equally weighted*) y el modelo. Con estas nuevas magnitudes, se definió cuán corto uno se puede ir en base a la suma de las señales negativas. La magnitud de la señal de venta ahora se puede utilizar para crear gaps y distribuir de manera equitativa lo que se está vendiendo de los activos con valores negativos.

Siguiendo la estrategia *top-down*, como ya está definido qué grupo de activos (RV, RF y Caja) debe mantenerse en posición larga y cuál en posición corta, llega el momento de decidir qué activos dentro de RV y RF comprar o vender y en qué cantidades. Entonces, conociendo el posicionamiento en RV por ejemplo, se define en base a los *scoring's* obtenidos qué activos pertenecientes a RV se les va a aumentar la exposición, a cuáles se les va a disminuir, y a cuáles se les va a mantener.

Un aspecto importante del modelo es que no solo proporcione un posicionamiento, sino que también pueda explicarlo o dar indicios de por qué tomó la decisión que tomó. En el caso de los dos primeros caminos (ver ilustración 7-4), es suficiente con guardar las variables a las que el modelo asignó una mayor ponderación en relación al activo que se está evaluando. Sin embargo, para el Análisis de Componentes Principales (PCA), la generación de insights se vuelve una tarea más compleja, ya que los componentes principales contienen información de toda la base de datos, por lo que no se puede determinar a priori qué variables están afectando la decisión. Para resolver este problema, se adoptó una estrategia que utiliza las correlaciones, que se basa en la siguiente hipótesis:

“Los componentes principales se organizan en función de la cantidad de variabilidad que explican. Si calculamos la correlación entre las clases que etiquetan las variables, encontraremos que las clases con mayor correlación con los componentes son las que más influencia tienen en el comportamiento de los componentes principales.”

Por consiguiente, al examinar cuáles componentes principales tuvieron un mayor impacto en el puntaje de convicción asignado a cada activo, podemos inferir, mediante el uso de la correlación, qué variables están influyendo en la toma de decisiones respecto al activo evaluado. Esto último siempre cuando la correlación posea un valor mayor a moderado (usualmente una correlación entre 0,4 y 0,6 se considera moderada).

Respecto a la medición del rendimiento del modelo mediante un backtesting, se tratará a continuación en la sección de Resultados.

Por último, pero no menos relevante, para el desarrollo del modelo de *asset allocation*, se seleccionó el lenguaje de programación Python como herramienta principal. Esta decisión se basó en las siguientes ventajas que ofrece Python en el contexto de este proyecto:

- Amplia utilización y popularidad.
- Versatilidad y facilidad de uso.
- Herramienta gratuita y de código abierto.
- Experiencia previa y conocimiento interno (la empresa lo utiliza).

8. Resultados

En esta sección se presentarán los resultados del *asset allocation* recomendada por el mejor modelo obtenido, a menos que se indique lo contrario. Por "mejor modelo" se entiende aquel que, mediante sus especificaciones y el camino elegido, proporcionó el mejor *information ratio* y, por ende, alfa (ver figura 8-1). Este modelo corresponde al camino con Análisis de Componentes Principales (PCA), más específicamente 14, la utilización de una ventana móvil de más de 7 años, correlación de Pearson, y un conjunto de entrenamiento y prueba del 98% y 2%, respectivamente.

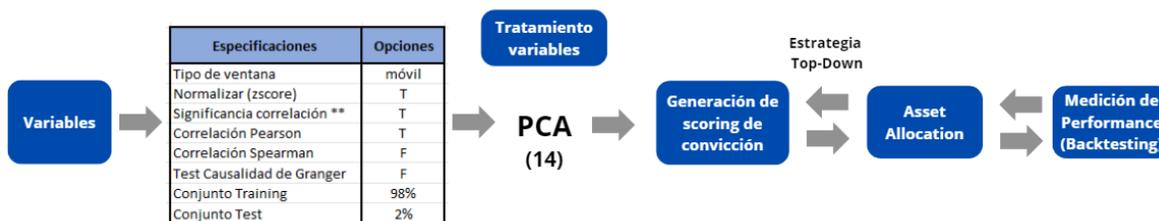


Ilustración 8-1: Diagrama del mejor modelo que se logró de manera experimental, junto las especificaciones y tratamiento de variable utilizado.

La utilización de un conjunto de entrenamiento tan grande en comparación con un conjunto de prueba reducido puede generar inquietudes, ya que en ciencia de datos lo habitual es utilizar un 70-80% de los datos para el entrenamiento y un 20-30% para el *testing* dependiendo de la cantidad de datos disponibles. Esto ayuda a evitar el sobreajuste y permite evaluar qué tan representativa es la muestra. Sin embargo, tras conversaciones con la contraparte, se concluyó que este modelo probablemente capta el momentum del mercado para la toma de decisiones, siendo eficaz para detectar la dirección del mercado en un cortísimo plazo. Esto es valioso para la empresa, ya que el modelo está diseñado para ser táctico y no estratégico.

A pesar de la preocupación sobre el tamaño del conjunto de entrenamiento y su relación con el conjunto de *tesing*, es relevante resaltar que su principal función en este modelo es calibrar parámetros claves, como la correlación y el IR por activo, para así generar un score que determine los posicionamientos recomendados para el mes siguiente. Dado el carácter dinámico de las series de tiempo, utilizar datos que no sean los más recientes podría llevar a la pérdida de información valiosa, comprometiendo la precisión de las recomendaciones y afectando el desempeño táctico del modelo en un plazo muy corto.

A partir de los *scoring's* de cada uno de los activos, se realizó la asignación de activos tal como se indicó en la fase 5 de la metodología. Tras aplicar la restricción de los *scoring's* y definir la ponderación para Renta Variable (RV), Renta Fija (RF) y Caja, se obtuvieron los posicionamientos mostrados en la ilustración 8-2.

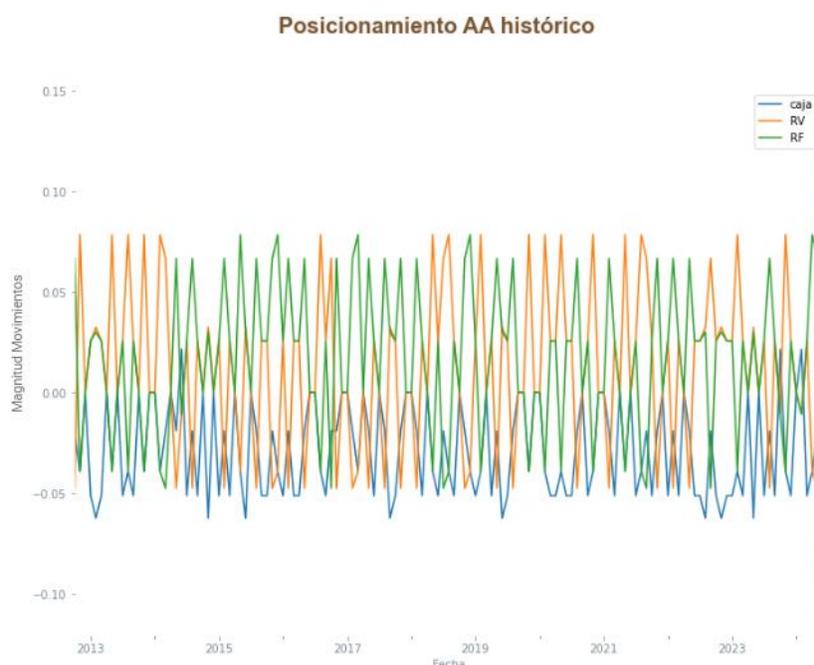


Ilustración 8-2: Gráfico que representa el *asset allocation* de RF, RV y caja a partir de un vector suma cero.

Luego, siguiendo la estrategia *top-down*, se realizaron los posicionamientos en los activos según la magnitud de los posicionamientos establecidos por el modelo para los activos dentro de Renta Variable (RV) y Renta Fija (RF) (ver Anexos C, D, E, F), ya que Caja es un solo activo.

En la ilustración 8-3 se puede ver cómo queda el portafolio recomendado a partir del *asset allocation* a nivel histórico.

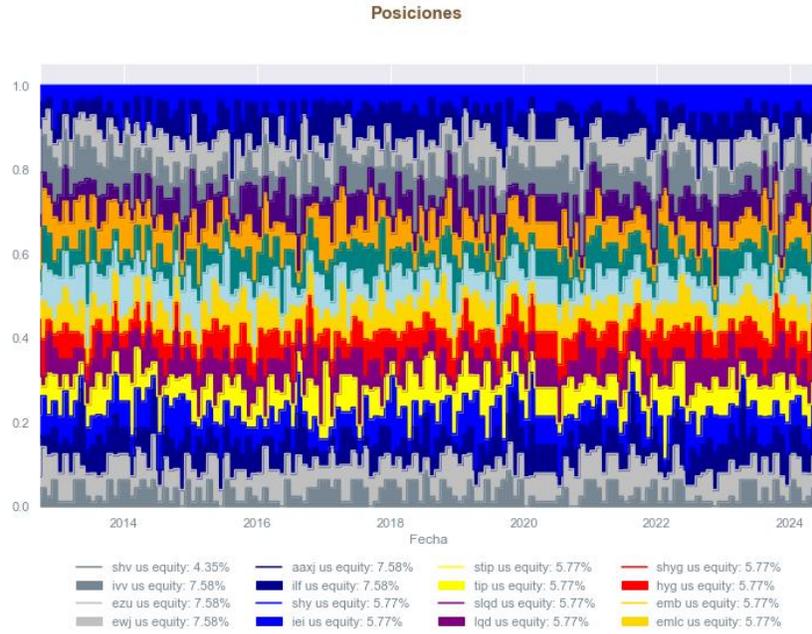


Ilustración 8-3: Ponderaciones históricas de los distintos activos que conforman un portafolio multiactivo.

Al evaluar el rendimiento histórico del posicionamiento mostrado en la ilustración 8-3, se obtiene el gráfico de la ilustración 8-4. Este gráfico refleja el alfa generado a lo largo de la historia a partir de un *backtesting*, así como también el IR histórico. Estos resultados muestran que el modelo está cumpliendo con los objetivos inicialmente propuestos, ya que está generando un IR superior incluso al rango inicialmente propuesto.



Ilustración 8-4: Gráfico de alfa acumulado a lo largo de la histórica junto con el IR.

Para contrastar con los otros caminos, se generó el mismo gráfico de alfa acumulado que se muestra en la ilustración 8-4, aplicándolo a los demás caminos: todas las variables y las agrupadas por clase, utilizando exactamente las mismas especificaciones y componentes principales empleados en el mejor modelo. Los resultados se presentan a continuación en la ilustración 8-5, y en el Anexo G se encuentran otras métricas.

En comparación con el mejor modelo (ver ilustración 8-4), se observa que la utilización de componentes principales ha mostrado generar alfa de manera consistente a lo largo del tiempo, con una acumulación significativa a partir de 2016. Para los modelos mostrados en la ilustración 8-5, la generación de alfa oscila de manera considerable en el camino que agrupa las variables mediante un promedio simple, mientras que es más suave en el camino que utiliza todas las variables. Sin embargo, en ambos casos, a partir de 2020, la generación de alfa se vuelve más relevante y consistente, especialmente en el modelo que incluye todas las variables, esto posiblemente debido a que al utilizar las variables agrupadas por promedio simple se pierde mucha variabilidad y por ende información en los datos.

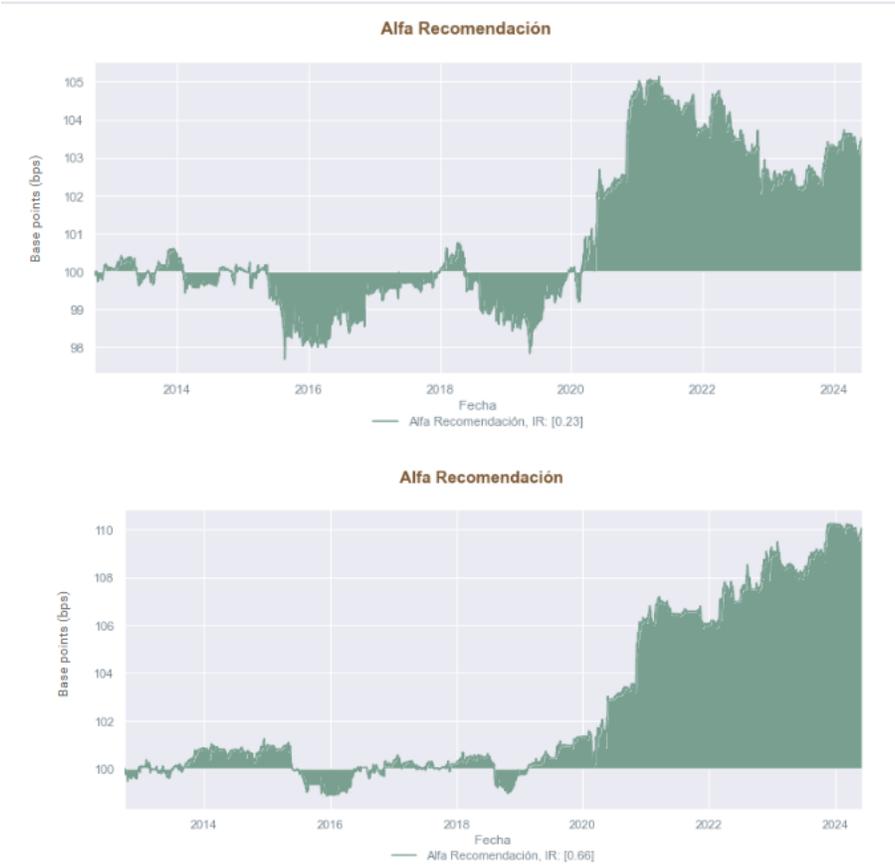


Ilustración 8-5: Gráficos de alfa acumulado e IR. El primero corresponde a la agrupación por clases (IR = 0,23) y el segundo a todas las variables (IR = 0,66).

En cuanto a la elección del número de componentes principales para el modelo, se hizo de manera empírica probando con distintos valores hasta llegar a 14. Este valor cobra sentido también si se observa la varianza acumulada que explican los componentes principales (ver ilustración 8-6).

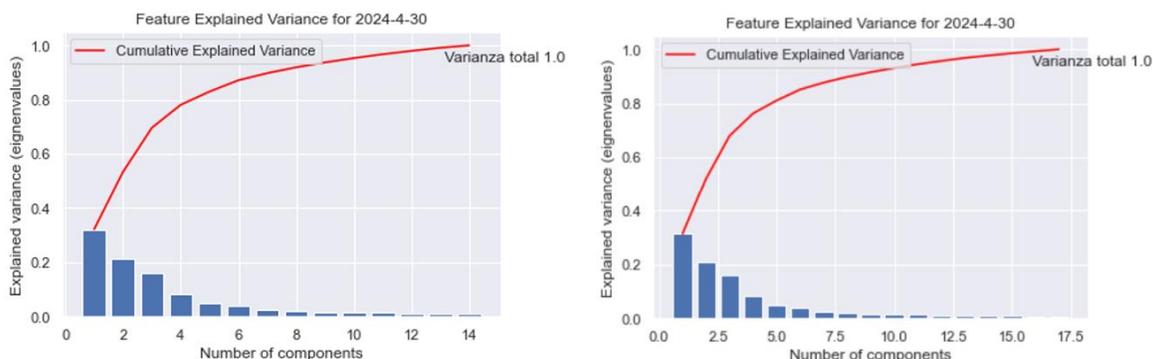


Ilustración 8-6: Varianza explicada según la cantidad de componentes principales.

A partir de la ilustración 8-6 (lado derecho) se observa que prácticamente después del componente número 15 la varianza que explica el siguiente componente es prácticamente nula, incluso si se observa la línea roja que representa la varianza acumulada, al final de la línea dice que el total de varianza explicada es 1, en ambos gráficos, lo que indica que no agrega más información al modelo incluir más componentes principales. Por último, también se hicieron pruebas con 13 y 15 componentes principales, debido a que a partir de la ilustración 8-6 se hace difícil saber cuál entrega un mejor resultado al momento de realizar el *asset allocation*. Se realizaron los posicionamientos con cada variable obteniéndose que el mayor alfa acumulado e IR se alcanzaba con 14 componentes principales (ver Anexos H e I).

Teniendo el mejor modelo identificado, se puede hacer un análisis más detallado sobre la composición del alfa generado en el *backtesting*. Para ello, se llevó a cabo un análisis *performance attribution*, donde se evalúa la contribución al alfa por parte del *asset allocation* y la selectividad de los activos (ver ilustración 8-7).

A partir de la ilustración 8-7 se observa que la mayor contribución al alfa está dada por el *asset allocation*, sugiriendo que la estrategia de asignación de activos ha sido efectiva y consistente en el tiempo. Así mismo, la selectividad de activos también presenta una contribución positiva al alfa, pero en menor magnitud y con una mayor volatilidad. Por un lado, a modo de clarificación, *asset allocation* se refiere a la elección entre RV (RV Desarrollado y RV Emergente), RF (RF *high beta* y RF *low beta*) y caja. Por otro lado, la selectividad hace referencia a la elección de activos específicos dentro de cada categoría de RV y RF (ver Anexo J).

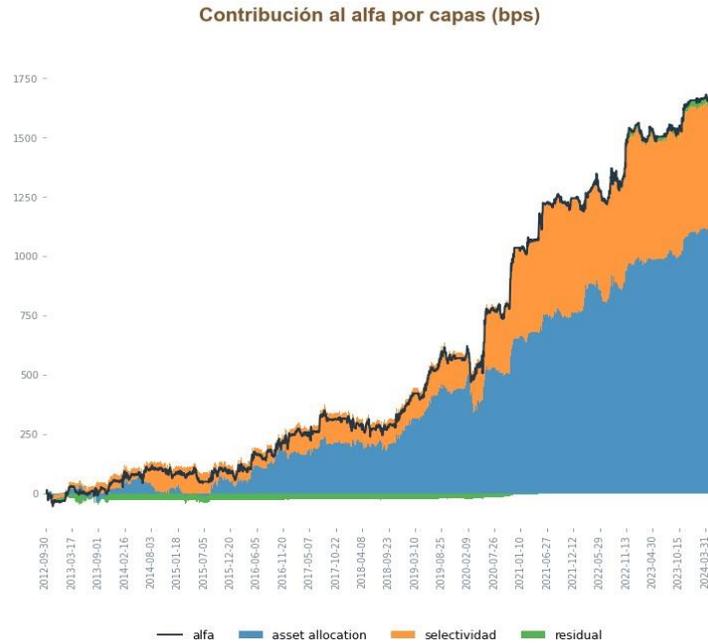


Ilustración 8-7: Contribución al alfa desglosado entre *asset allocation*, selectividad y valores residuales.

Un análisis más profundo dentro del *asset allocation* permite desglosar las contribuciones de RV (desarrollada y emergente), RF (*high beta* y *low beta*) y Caja al alfa generado. La ilustración 8-8 revela que la mayor contribución proviene de la renta variable, con RV desarrollado liderando tanto en contribución como en retornos. Esto sugiere que posiblemente a nivel histórico el modelo ha tomado una posición larga en RV desarrollado. En contraste, RV emergente presenta una contribución significativa, pero con bajos retornos, lo que indica que el modelo lo más probable es que está identificando y tomando de manera correcta posiciones cortas en momentos de bajo rendimiento de RV emergente.

En el caso de la renta fija (RF), la ilustración 8-8 muestra que los retornos históricos, tanto para RF high beta como para RF low beta, son comparativamente más bajos que los de la renta variable desarrollada. Esto era previsible, dado el menor riesgo asociado a la RF. En cuanto a RF high beta, se observa una contribución positiva al alfa, lo que sugiere que el modelo ha sido efectivo al tomar posiciones largas y cortas a lo largo del tiempo. Por otro lado, RF low beta presenta una contribución ligeramente negativa a la generación de alfa, a pesar de que los retornos de RF low beta fueran positivos, lo que podría indicar un desempeño subóptimo del modelo en la selección de posiciones dentro de esta categoría.

Finalmente, para Caja, la ilustración 8-8 revela que, a pesar de sus retornos positivos, su contribución al alfa es negativa. Esto sugiere que el modelo tampoco ha sido efectivo al momento de tomar posiciones en Caja.

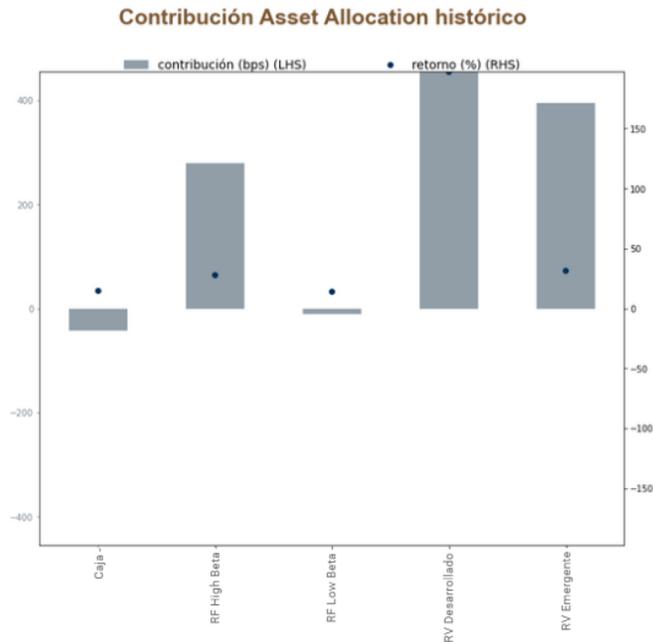


Ilustración 8-8: Contribución *asset allocation* a nivel histórico por parte de RV (Desarrollado y Emergente), RF (*high beta* y *low beta*) y Caja.

Siguiendo la línea de los resultados expuestos, se puede establecer que, por un lado el objetivo general se estaría cumpliendo, ya que el alfa generado en el *backtesting* ha resultado ser positivo al igual que el *information ratio* con un valor de 0,97. Por otro lado, los objetivos específicos 1 y 2 también se estarían cumpliendo, debido a los buenos resultados obtenidos a partir del *Data set* propuesto y a las especificaciones encontrados que han permitido maximizar el *information ratio*.

Tal como se mencionó en la metodología y los objetivos, se busca que el modelo también posea un cierto poder explicativo sobre las decisiones que toma en relación con las variables que considera al asignar el *scoring* para un activo en el mes siguiente. Para los dos primeros caminos, todas las variables y agrupación por clase, esta tarea se logra guardando los valores que resultan para cada variable al momento de multiplicarlas por la correlación y los IR_w , debido a que las variables que resulten con los mayores valores son los que estarán influyendo en mayor medida en la decisión.

Sin embargo, no sucede lo mismo con los componentes principales, ya que como se comentó en el marco conceptual los componentes principales transforman la dimensión completa del *data set* original, y mediante transformaciones algebraicas reducen la dimensionalidad de esta. El problema con aquello es que es difícil saber qué información o variabilidad de las variables originales están almacenadas en los componentes. Por lo tanto, no tiene utilidad almacenar los componentes principales que obtuvieron un mayor peso en la decisión, ya que se desconoce lo que está por detrás. Para solucionar esto se probó utilizar lo que se conoce

como “*correlation circle*”, y calcular componentes principales por clase según las variables que pertenecieran a cada clase.

Correlation circle al igual que una matriz de correlaciones es una forma gráfica de representar cómo es la correlación entre dos variables. Siguiendo con la hipótesis planteada en la fase 5 en la metodología, la ilustración 8-9 muestra que, efectivamente, las correlaciones entre las clases y los componentes principales toman valores distintos de cero (mayores a 0,4). Esto indica que existen componentes principales que poseen una relación considerable a nivel estadístico con una o dos clases, lo que sugiere que dichos componentes principales están más influenciados por esas clases y, por lo tanto, son más sensibles a los cambios en las variables pertenecientes a esas clases. Sin embargo, al observar el lado derecho de la ilustración 8-9, se aprecia que a partir del PC4 las correlaciones se mueven cercanas a cero, lo que implica una pérdida de la relación entre los componentes principales y las clases. Como consecuencia, esta estrategia deja de tener un impacto significativo en la capacidad explicativa sobre el tipo de variables que está predominando en la decisión, y no resulta adecuada para el propósito buscado.

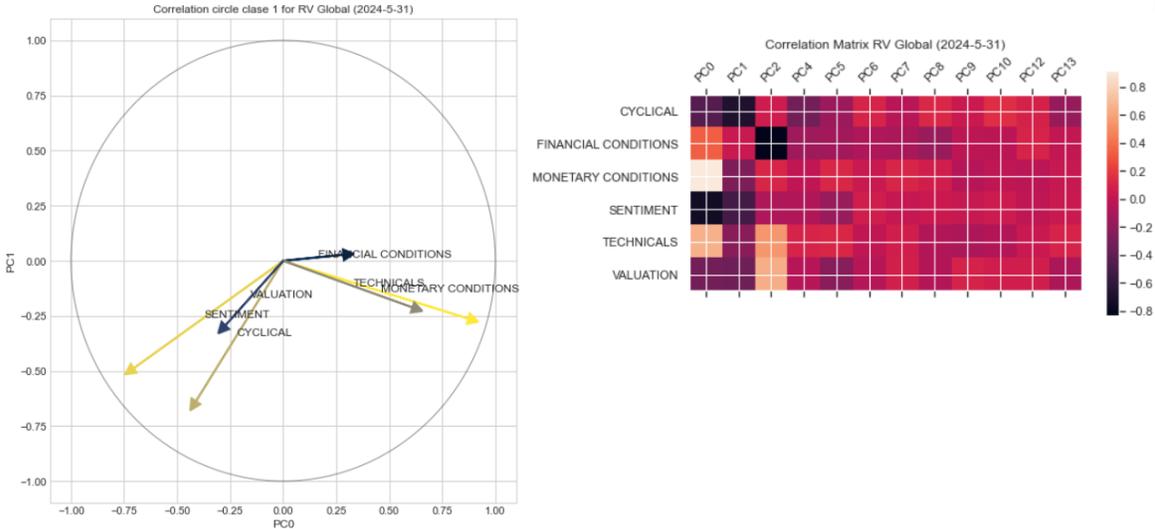


Ilustración 8-9: El gráfico del lado izquierdo corresponde a un *correlation circle* entre PC1 y PC0 para RV, el gráfico del lado derecho corresponde a la matriz de correlación entre los componentes principales y las clases.

En cuanto a la segunda alternativa, que consiste en calcular los componentes principales para cada clase, esta implica agrupar las variables por clase y obtener componentes principales a partir de cada subgrupo, los cuales se etiquetan según su clase correspondiente. Esta alternativa tampoco resultó ser efectiva, ya que generó un *information ratio* (IR) de 0,12 (ver anexos K y L), lo que se considera un valor bajo.

9. Discusiones

Se logró generar una base de datos lo suficientemente amplia para la construcción de un modelo de *asset allocation* global efectivo, lo que se tradujo en un *information ratio* favorable. Sin embargo, la revisión de las variables empleadas podría haberse profundizado aún más. Inicialmente, se consideraron más de 1000 variables, las cuales se redujeron a 263 utilizando principalmente la métrica "*relevance value*". No obstante, la aplicación de esta métrica presenta interrogantes, ya que refleja las variables que los analistas consideran relevantes en el presente, pero su enfoque podría cambiar en el futuro, invalidando el criterio. Por ello, se sugiere actualizar las variables periódicamente, aplicando esta u otras métricas. Sin embargo, esto implicaría una carga adicional para los *portfolio managers*, quienes deberían encargarse de la revisión y actualización constante de las variables, lo cual no es lo ideal. En este sentido, se propone para futuros trabajos evaluar alternativas para automatizar este proceso de selección de variables, ya sea mediante la optimización de la métrica "*relevance value*" o la exploración de otras metodologías que permitan una selección eficiente.

La colaboración con expertos en la selección y clasificación de variables podría plantearse como otra alternativa viable para abordar las limitaciones mencionadas en el párrafo anterior. Sin embargo, la elevada diversificación de los fondos globales, con ETF's que a su vez invierten en una multitud de activos influenciados por diversos factores, torna esta labor costosa en términos de horas-hombre, debido a que elaborar un conjunto de datos especializado que incluya las variables más relevantes para cada ETF, se requeriría la participación de múltiples expertos con conocimientos específicos en cada sector. La creación de un área especializada en este tipo de tareas podría ser ideal, no solo para los fondos multiactivos, sino para una amplia gama de inversiones. No obstante, nuevamente surge el obstáculo del costo, ya que la implementación y mantenimiento de dicha área resultarían poco óptimo para la empresa. A pesar de esto último, existen empresas que sí poseen este tipo de áreas con expertos con enfoque cuantitativo, por lo que habría que hacer un diagnóstico sobre como es su estructura para que sea rentable.

La elección de las seis clases también podría resultar cuestionable, dado que hay clases, como *cyclical* y *sentiment*, que por construcción están muy correlacionadas entre sí. Esto podría traer problemas de multicolinealidad entre las clases y, por ende, problemas en la interpretabilidad de los resultados, como efectivamente ocurrió y se hablará más adelante en esta sección. Este problema puede explicarse porque, al tener dos variables con una correlación muy alta, es posible que los componentes se enfoquen en capturar la colinealidad entre las variables más que la variabilidad significativa de los datos. Como alternativa, se podría haber realizado un análisis más profundo respecto a la relación entre las variables de cada clase, lo que podría haber generado un menor número de clases. Por ejemplo, se podría haber creado una clase que englobara las clases *cyclical* y *sentiment*, o incluso otras clases que no tuvieran una correlación muy alta entre sí.

Otro aspecto cuestionable en el tratamiento de las variables es el rezago aplicado para evitar el *look-ahead bias*. En este caso, se optó por rezagar todas las variables un mes, sin distinguir si realmente había un desfase en la información de cada una. Por ejemplo, los datos de inflación suelen publicarse con retraso y corresponden al mes anterior, mientras que otros datos, como precios, tasas de interés o volatilidad, no presentan rezago, por lo que aplicarles un desfase implicaría perder información. Por lo tanto, se sugiere para futuros trabajos o mejoras, implementar una variable binaria que indique si una variable debe ser rezagada o no. Alternativamente, otra opción sería ver el rezago existente en la información más actualizada, y considerarlo para toda la base histórica. Si bien estos análisis requerirían una inversión de tiempo significativa por parte del PM para revisar cada variable, sería un esfuerzo puntual que luego podría mantenerse en el largo plazo.

Como se observa en la ilustración 8-2, la fluctuación en los posicionamientos del modelo de *asset allocation* para RV, RF y Caja podría generar un impacto en los costos de transacción (compra y venta) asociados a su implementación. Si bien el modelo actual no incorpora explícitamente la restricción de costos de transacción en su toma de decisiones, es importante considerar este factor. Sin embargo, un análisis de la métrica "duración nota" en la última tabla de métricas del Anexo G revela que el modelo mantiene posiciones recomendadas durante un promedio de 62 días. Esta característica indica que el modelo no genera cambios de posicionamiento excesivamente frecuentes, lo que limita el impacto potencial de los costos de transacción. No obstante, la restricción por costos sigue siendo un aspecto relevante en la toma de decisiones de inversión y debería incorporarse en trabajos futuros para una evaluación más completa.

El motivo por el cual no se incorporó una restricción tan relevante como esta puede explicarse por el objetivo principal del modelo, que es entregar recomendaciones de posicionamiento y no implementar las posiciones como tal. Para entender cómo el modelo agrega valor a la empresa, se pueden considerar dos vías. En primer lugar, el modelo puede ser útil para un PM que esté indeciso sobre si tomar una posición larga, corta o neutral respecto a un activo o conjunto de activos, proporcionando directrices que lo ayuden a tomar una decisión informada y justificarla. En segundo lugar, el modelo puede utilizarse para confirmar y justificar las creencias o convicciones del PM respecto a si un activo va a tener un mejor o peor rendimiento en comparación al *benchmark*.

En cuanto al desarrollo de una metodología para interpretar los resultados y comprender qué factores utiliza el modelo para la toma de decisiones sobre un activo, no se logró un avance significativo con el mejor modelo que emplea componentes principales. Las dos alternativas exploradas no arrojaron los resultados esperados.

La primera alternativa, que consistía en utilizar correlaciones entre los componentes principales y las clases, no tuvo éxito. Esto se debe a que la correlación se aproximaba a cero para los componentes principales que explicaban una menor variabilidad, invalidando la hipótesis planteada en la metodología para interpretar dichos componentes. Una posible solución para futuros trabajos sería implementar una función dentro del modelo que otorgue mayor relevancia o valor a los resultados de los primeros componentes principales, ya que estos explican la mayor variabilidad y presentan las correlaciones más fuertes.

La segunda alternativa, que implicaba el cálculo de componentes principales por clase individualmente, tampoco resultó efectiva. Los resultados en la generación de alfa y, por consiguiente, en el *information ratio*, fueron deficientes. Esto podría atribuirse a la pérdida de información valiosa sobre las relaciones o interacciones entre variables de diferentes clases al agruparlas por clase (cíclicas, condiciones financieras, condiciones monetarias, etc), debido a que variables de tipo cíclicas y de sentimiento debiesen de estar altamente correlacionadas. Además, esta división podría generar componentes principales redundantes o irrelevantes, lo que reduce la capacidad del modelo para generar IR. Esta última afirmación se respalda por el hecho de que al aplicar esta técnica se utilizaron más de 50 componentes principales, mientras que el modelo óptimo solo requirió 14.

Respecto a los alcances del proyecto se consideran adecuados y pertinentes a lo que la empresa busca con este modelo, ya que modelos más sofisticados de *Machine learning* o de *Deep learning* se convierten muchas veces en verdaderas cajas negras, por lo que la construcción de un algoritmo con las características ya presentadas simplifica el proceso de toma de decisiones y permite que individuos con pocos conocimientos en *Data Science* puedan entenderlo e incluso modificarlo de ser necesario.

Sin embargo, componentes principales si bien en sí mismo no es un algoritmo de *machine learning* en el sentido tradicional (ya que no realiza predicciones o clasificaciones), sí constituye una técnica que opera como una caja negra. Esto se debe a que no se conoce exactamente qué información se almacena en los componentes principales. Lo que dificulta el seguimiento de las variables relevantes. Por esta razón, se sugiere explorar otras alternativas para aprovechar la información de múltiples variables o consultar con un experto en tratamiento de datos para abordar el problema.

Es importante destacar que, si bien la empresa cuenta con profesionales altamente calificados, su formación no necesariamente incluye un profundo conocimiento en *Data Science*. La relevancia y el potencial de extraer *insights* de los datos han cobrado auge en los últimos años, lo que ha generado una brecha en la formación de profesionales en comparación con las demandas actuales.

La metodología empleada para la construcción de este modelo resultó ser efectiva, ya que se lograron alcanzar buenos resultados, a excepción de la interpretabilidad de los resultados dados por los componentes principales. CRISP-DM es una metodología estándar y ampliamente utilizada en el ámbito de la ciencia de datos, con un enfoque significativo en los objetivos del negocio. Aunque existen otras metodologías como KDD (*Knowledge Discovery in Databases*), SEMMA (*Sample, Explore, Model, Model Assessment, and Deployment*) y P3TQ (*Problem, Plan, Process, Quality, Time*), cada una con sus propias particularidades, la lógica subyacente en la construcción de modelos sigue siendo similar.

CRISP-DM se diferencia por su enfoque integral que combina aspectos técnicos y de negocio, su naturaleza iterativa y su independencia de herramientas específicas. KDD se centra en el descubrimiento de conocimiento a partir de bases de datos, SEMMA se enfoca más en el procesamiento y modelado de datos, mientras que P3TQ enfatiza la gestión de proyectos y la planificación. La elección de CRISP-DM se debió no solo a su éxito previo en

el desarrollo de un modelo similar dentro de la empresa, sino también a su fuerte enfoque en los objetivos del negocio, lo que la hizo más adecuada y preferida sobre las otras metodologías en el contexto actual.

En consonancia con lo expuesto anteriormente y en los resultados, el análisis de la contribución al alfa del mejor modelo reveló un desempeño positivo en el *asset allocation* y en la selectividad para la generación de alfa. Sin embargo, al examinar en detalle la composición del *asset allocation*, se observó que el modelo no era bueno al momento de generar alfa con los posicionamientos de la Caja y renta fija *low beta*, ya que sus contribuciones al alfa fueron negativas a pesar de que sus retornos históricos fueron positivos.

La razón de este comportamiento no está del todo clara, especialmente considerando las dificultades inherentes a la interpretación de los componentes principales. No obstante, cabe mencionar que la tarea de generar alfa es intrínsecamente compleja, y es posible que el modelo haya encontrado dificultades para superar al mercado en estos activos específicos. A pesar de esto, el modelo demostró un desempeño notable en la generación de alfa con los activos de mayor retorno, lo que se alinea con el objetivo central de este tipo de modelos.

En cuanto al *Risk Management*, los principales riesgos que podrían afectar los resultados del modelo son de carácter financiero. Entre estos se encuentra el riesgo de mercado, ya que los activos del portafolio están sujetos a variaciones que provienen del entorno económico, como cambios en las políticas monetarias de los países, fluctuaciones en los precios de *commodities* y acciones, entre otros. También está presente el riesgo de crédito, que surge cuando un emisor tiene compromisos de pago, lo cual es especialmente relevante, dado que el portafolio incluye inversiones en *investment grade, high yield*, y deuda emergente, que tienen diferentes niveles de riesgo crediticio. Por último, se considera el riesgo de liquidez, aunque es menos probable que ocurra, ya que el portafolio está compuesto principalmente por ETF's, que son activos generalmente líquidos. Sin embargo, el riesgo de liquidez podría materializarse si el mercado subyacente de los activos de algunos ETF's en los que invierte el fondo se vuelve ilíquido o si el volumen de transacciones del ETF disminuye significativamente durante periodos de estrés en los mercados.

Siguiendo la línea anterior, el vector de recomendación generado por el modelo se somete a un análisis para determinar si es aplicable bajo las directrices del IPS del fondo. El modelo, por una parte, incorpora restricciones que evitan la toma de posiciones excesivas en ciertos activos. Por otra parte, el vector resultante es ajustado por riesgo para que sea comparable con los demás vectores de recomendación, tanto del equipo de estrategia interno de la empresa como de empresas *buy-side* y *sell-side*. Adicionalmente, se le calculan métricas como Beta ACWI y Vol/TE, entre otras, las cuales se pueden observar en la parte inferior de la ilustración 1.2-3.

Como extensión del proyecto para futuros trabajos, se propone explorar la posibilidad de integrar el *scoring* generado con un modelo de construcción de portafolios. Por ejemplo, una estrategia viable consistiría en ponderar los retornos más recientes de los activos por los *scores* obtenidos, y posteriormente aplicar la teoría moderna de portafolios de Markowitz. De esta manera, se otorgaría mayor relevancia a los activos que el modelo identifique con un

potencial de rendimiento superior para el siguiente mes y también se incorporaría a través del score factores exógenos a las variables.

Otros posibles caminos que se podrían explorar, pero que se desvía un poco de los objetivos específico y alcances, es la utilización de modelos de *Machine Learning* y *Deep Learning*, para generar el *scoring*, ya que estos modelos poseen la capacidad de detectar interacciones y patrones no lineales que modelos más sencillos no serían capaces de detectar. Por lo que sería interesante probar también esa línea de investigación.

10. Conclusiones

El objetivo de este trabajo fue desarrollar un modelo de *asset allocation* para los fondos multiactivos de la empresa, con el fin de generar un puntaje (*scoring*) basado en las relaciones entre los activos del fondo y variables macroeconómicas y de mercado. Este puntaje indicaría el grado de convicción del modelo respecto a si un activo tendría un movimiento positivo o no, para luego convertir dicho puntaje en una posición dentro de un vector suma cero, en relación con el *benchmark*. Si la diferencia con el *benchmark* era positiva, se *sobreponderaba* el activo; si era negativa, se *subponderaba*; y si era neutral, se mantenía la posición.

A partir de los resultados obtenidos, se logró cumplir con el objetivo general ya que el modelo generó un vector de recomendación de suma cero que alcanzó un *information ratio* favorable en el *backtesting*. Este resultado superó incluso las expectativas iniciales planteadas por la empresa, las cuales situaban un rango óptimo entre 0,4 y 0,6, valores que, según la experiencia de la empresa, caracterizan el desempeño de un buen *portfolio manager*. Cabe destacar que, con los resultados presentados en cuanto a la contribución al alfa (ilustración 8-7), la mayor contribución proviene del *asset allocation*. Esto indica que el modelo ha demostrado un buen posicionamiento de los activos en la construcción de las recomendaciones en el *backtesting*. Sin embargo, es necesario evaluar el desempeño del modelo en el período de supervisión antes de su implementación, ya que como es sabido “retornos pasados no garantizan retornos futuros”.

Respecto a los objetivos específicos, el conjunto de datos propuesto ha demostrado ser suficiente para proporcionar la información necesaria al modelo para tomar decisiones que generen un alfa positivo y un buen *information ratio*. No obstante, el desempeño del modelo en cuanto al posicionamiento en liquidez (caja) y renta fija (RF) *low beta* ha sido deficiente. Si bien los retornos históricos de estos activos fueron positivos, no contribuyeron a la generación de alfa. Por lo tanto, se estima que, por un lado, podría existir falta de información sobre estos dos activos en el *data set* y, por otro lado, el modelo podría no estar detectando correctamente los patrones de dichos activos.

Tras un proceso exhaustivo de pruebas y ajustes de las especificaciones, se logró identificar una combinación óptima que maximizó el *information ratio*. Las especificaciones clave de este modelo incluyen una ventana móvil, normalización de variables con z-score, filtrado mediante correlación de Pearson, un conjunto de entrenamiento del 98% y un conjunto de *testing* del 2%, junto con la selección de 14 componentes principales.

Sin embargo, la aplicación de componentes principales para explicar los resultados del modelo presentó limitaciones en el cumplimiento del tercer objetivo específico, que buscaba identificar las variables con mayor incidencia en la toma de decisiones. A pesar de probar dos estrategias distintas, no se lograron los resultados esperados. En este contexto, resulta necesario explorar enfoques alternativos que faciliten la comprensión de la relación entre los componentes principales y las variables utilizadas. Esto podría implicar la investigación de técnicas de reducción de dimensionalidad alternativas.

En cuanto al último objetivo planteado, también se considera logrado, ya que el desarrollo de este documento junto con el código comentado más las presentaciones dentro de la empresa, permiten documentar de buena manera la metodología para el desarrollo y cumplimiento de los objetivos expuestos en este informe.

Teniendo en cuenta las críticas constructivas al modelo mencionadas en esta discusión, podría ser necesario replantear lo establecido en los alcances, que inicialmente se limitaba a monitorear el modelo en un entorno práctico, comparándolo con otros modelos o con las recomendaciones del equipo de estrategia interno o de empresas buy-side y sell-side para evaluar su viabilidad. Esto se debe a que han surgido elementos relevantes que requieren una segunda revisión, con el fin de asegurar que el modelo sea sólido, más allá de generar posiciones acertadas en el tiempo o producir alfa. Además, si se agrega el hecho de que el modelo no logra ofrecer interpretabilidad en sus resultados, sería difícil confiar en su implementación, ya que carecería de una justificación racional que explique las decisiones tomadas.

En cuanto a futuras líneas de investigación, se propone explorar la posibilidad de combinar el *scoring* generado por el modelo con metodología de optimización de portafolios, como la teoría moderna de portafolios de Markowitz. De igual manera, se recomienda evaluar la aplicación de modelos de *machine learning* o *deep learning* para el cálculo del *scoring*, con el objeto de mejorar aún más el desempeño del modelo.

Bibliografía

- [1] Comisión para el Mercado Financiero (CMF), «¿Qué son los fondos Mutuos?,» 2023. [En línea]. Available: <https://www.cmfchile.cl/educa/621/w3-propertyvalue-987.html>.
- [2] M. d. Hacienda, *LEY 20712 ADMINISTRACIÓN DE FONDOS DE TERCEROS Y CARTERAS INDIVIDUALES Y DEROGA LOS CUERPOS LEGALES QUE INDICA*, Santiago de Chile: Biblioteca del congreso, 2013.
- [3] Comisión para el Mercado Financiero (CMF), «¿Qué tipos de fondos mutuos existen?,» 2023. [En línea]. Available: <https://www.cmfchile.cl/portal/principal/613/w3-article-641>.
- [4] Asociación Administradora de Fondos Mutuos, «Asociación de Fondos Mutuos Informe anual,» Santiago, 2023.
- [5] Black Rock, «Que son las estrategias de multiactivos,» 2023. [En línea]. Available: <https://www.blackrock.com/co/educacion/fondos-multiactivos#:~:text=Los%20fondos%20de%20multiactivos%20ofrecen,riesgos%20o%20buscar%20mejores%20oportunidades..>
- [6] «Big Data: Global Strategic Business Report,» Global Industry Analysts, Inc, 2024.
- [7] Justine Lee, «Sixteen Leading Quants Imagine the Next Decade in Global Finance,» *Bloomberg*, 29 Enero 2020.
- [8] P. Taylor, «Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025,» Statista, 16 Noviembre 2023. [En línea]. Available: <https://www.statista.com/statistics/871513/worldwide-data-created/>. [Último acceso: 7 Mayo 2024].
- [9] BBVA, «Qué son las finanzas cuantitativas: origen y aplicaciones,» 29 Noviembre 2023. [En línea]. Available: <https://www.bbva.com/es/salud-financiera/que-son-las-finanzas-cuantitativas-origen-y-aplicaciones/>.
- [10] Corporate Finance Institute, «Quant fund,» 2023. [En línea]. Available: <https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/quant-fund/>.
- [11] J. McWhinney, «Quantitative Analysis: A Simple Overview,» Investopedia, 26 Marzo 2024. [En línea]. Available: <https://www.investopedia.com/articles/investing/041114/simple-overview-quantitative-analysis.asp#toc-pros-and-cons-of-quant-trading>. [Último acceso: 7 Mayo 2024].
- [12] CFA Institute, «Time-Series Analysis,» de *Quantive Methods, Economics- level 2-volume 1*, 2023, pp. 111-190.
- [13] «Ventajas y desventajas del análisis de las series de tiempo,» One Marketing, [En línea]. Available: <https://onemarketing.es/ventajas-y-desventajas-del-analisis-de-las-series-de-tiempo/>. [Último acceso: 3 mayo 2024].
- [14] CFA Institute, «Portfolio Management, Ethical and Professional Standards-level 1- volume 6,» CFA Program, 2024, pp. 3-82.

- [15 B. Trabelsi, «Medium,» 1 julio 2021. [En línea]. Available: <https://bechirtr97.medium.com/feature-selection-in-unsupervised-learning-problems-585a56d024bd>. [Último acceso: 3 mayo 2024].
- [16 J. Frost, «Statistics By Jim Making statistics intuitive,» [En línea]. Available: <https://statisticsbyjim.com/basics/principal-component-analysis/>. [Último acceso: 3 mayo 2024].
- [17 CFA Institute, «Measuring Value Added,» de *Portfolio Management, Ethical and Professional Standards-level 2-volume 6*, 2023, pp. 91-92.
- [18 CFA Institute, «The Information Ratio,» de *Portfolio Management, Ethical and Professional Standards-level 2-volume 6*, 2023, p. 98.
- [19 CFA Institute, «The Informatio Ratio,» de *Portfolio Management, Ethical and Professional Standards-level 2-volume 6*, 2023, pp. 98-100.
- [20 P. & G. M. Samuels, «Pearson Correlation,» *Stats Tutos* , pp. 1-4, 18 octubre 2023.
- [21 Springer, «Spearman Rank Correlation Coefficient.,» de *The Concise Encyclopedia of Statistics*, New York, 2008.
- [22 C. W. Granger, Investigating causal relationship by econometrics and cross-spectral models, *Econometrica*, 1069.
- [23 CFA Program, «Good Bad and Ugly of Bias in AI,» CFA Institute, [En línea]. Available: <https://www.cfainstitute.org/en/professional-insights-stories/good-bad-and-ugly-of-bias-in-AI>. [Último acceso: 2024 junio 25].
- [24 IBM, «Conceptos básicos de ayuda de CRISP-DM,» IBM, 17 agosto 2021. [En línea]. Available: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>. [Último acceso: 4 mayo 2024].

Anexos

	variable 1	variable 2	variable 3	variable 4	variable N
Fecha M-1	$w_i * test_{M-1}$	$w_i * test_{M-1}$	$w_i * test_{M-1}$	$w_i * test_{M-1}$	$w_i * test_{M-1}$
Fecha M	$w_i * test_M$	$w_i * test_M$	$w_i * test_M$	$w_i * test_M$	$w_i * test_M$

➔

	Prediction
Fecha M-1	$Sum(w_i * test_{M-1})$
Fecha M	$Sum(w_i * test_M)$

Anexo A: En el lado izquierdo se observa la data de test ponderada por los pesos calculados a partir de los IR's, y en el lado derecho la suma del resultado del lado izquierdo por fila.

	Prediction
Fecha M-1	$Rank(Sum(w_i * test_{M-1}))$
Fecha M	$Rank(Sum(w_i * test_M))$

➔

	Prediction
Fecha M-1	1
Fecha M	0,5

o

	Prediction
Fecha M-1	0,5
Fecha M	1

Ilustración B: Cálculo del ranking según percentil entre los dos meses de la data de testing, en donde el scoring toma el valor 1 si es que M-1 posee un valor mayor que en M, y viceversa

Posicionamiento AA histórico

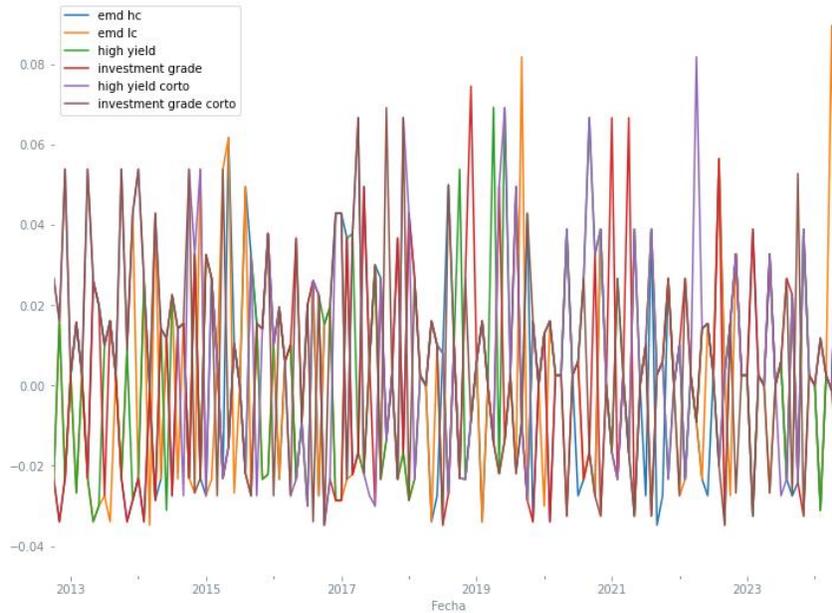
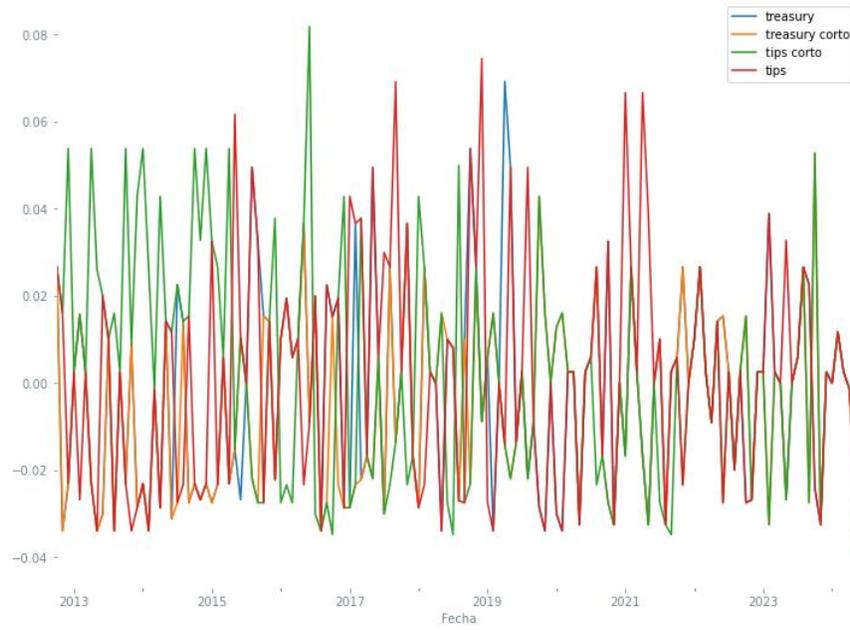


Ilustración C: Asset allocation histórico realizado por el modelo para Renta Fija High Beta.

Posicionamiento AA histórico



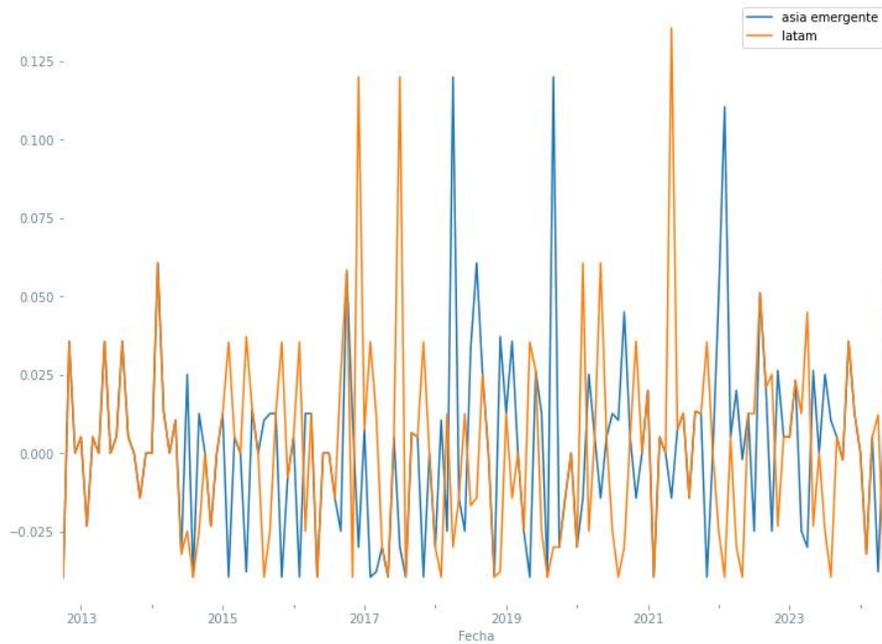
Anexo D: Asset allocation histórico realizado por el modelo para Renta Fija Low Beta.

Posicionamiento AA histórico



Anexo E: Asset allocation histórico realizado por el modelo para Renta Variable Desarrollado.

Posicionamiento AA histórico



Anexo F: Asset allocation histórico realizado por el modelo para Renta Variable Emergente.

Metrica	valor
Fecha inicial	2012-09-30
Fecha Final	2024-05-15
Alfa anualizado	0.38%
Tracking error	1.22%
Turnover	0.17
Duracion nota	65.11
Probabilidad de éxito	51.08%
IR	0.31

Metrica	valor
Fecha inicial	2012-09-30
Fecha Final	2024-05-15
Alfa anualizado	0.84%
Tracking error	1.20%
Turnover	0.16
Duracion nota	62.81
Probabilidad de éxito	50.22%
IR	0.69

Metrica	valor
Fecha inicial	2012-09-30
Fecha Final	2024-06-03
Alfa anualizado	1.14%
Tracking error	1.18%
Turnover	0.18
Duracion nota	62.2
Probabilidad de éxito	50.93%
IR	0.97

Anexo G: Tablas de métricas para los tres tipos de inputs. El primero corresponde a la agrupación por clases, el segundo a todas las variables y el tercero a PCA.



Anexo H: Gráfico de alfa acumulado e IR a lo largo de la historia de la data con 13 componentes principales.



Anexo I: Gráfico de alfa acumulado e IR a lo largo de la historia de la data con 15 componentes principales.

Global	
Asset Allocation	RV
	RF
	Caja
	RVG DM
	RVG EM
	RFG LB
	RFGHB
	Caja
Security Selection	EE.UU.
	Eurozona
	Japón
	Asia ex-japón
	Latam
	T - LD
	T - SD
	TIPS - LD
	TIPS - SD
	IG - LD
	IG - SD
	HY - LD
	HY - SD
	EMD HC
	EMD LC
	caja

Anexo J: Distribución de los activos locales entre *asset allocation* y *security selection*.



Anexo K: Resultados en la generación de alfa utilizando componentes principales por cada clase.

Metrica	valor
Fecha inicial	2012-09-30
Fecha Final	2024-06-03
Alfa anualizado	0.26%
Tracking error	2.19%
Turnover	0.2
Duracion nota	72.42
Probabilidad de éxito	50.82%
IR	0.12

Anexo L: Métricas de resultado del modelo utilizando componentes principales por cada clase.