



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

“ACTIVOS ALTERNATIVOS: ESCENARIOS Y PREDICCIÓN DE FLUJOS DE CAJA”

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

DIEGO ALONSO VILLAVICENCIO QUEZADA

PROFESOR GUIA:
CARLOS PULGAR ARATA.

PROFESOR CO-GUIA:
ANTONIO BARRIOS NAMUNCURA.

COMISIÓN:
LUIS LLANOS COLLADO.

SANTIAGO DE CHILE
2024

Resumen Ejecutivo

Este informe presenta los principales resultados del proyecto desarrollado en colaboración con una empresa especialista en activos alternativos. El enfoque fue el modelamiento de flujos de ingreso y egreso en fondos de inversión de este tipo de activos. La memoria se centró en modelar, de manera gráfica y sencilla, trayectorias o escenarios para estos flujos, considerando el vintage o año de inicio de inversión. También se exploró el uso de modelos computacionales para predecir dichos flujos en base a historia.

El objetivo principal se dividió en dos etapas: primero, desarrollar un programa en *Python* para generar escenarios basados en datos históricos de curvas que modelan el comportamiento de activos alternativos; segundo, crear un programa en *Python* que use redes neuronales para proyectar programas de inversión con movimientos ya realizados. El estudiante jugó un rol clave en el análisis, preprocesamiento de datos y creación de modelos predictivos, además de investigar nuevas metodologías.

La metodología incluyó estadística aplicada, técnicas de análisis y preprocesamiento de datos, modelado financiero y el uso de *deep learning* con redes neuronales *LSTM* y *GRU*, empleando *Python* y *SQL*.

En la primera etapa, se levantaron datos de un universo de fondos aprobado por el equipo, generando curvas basadas en percentiles que ilustran posibles escenarios pesimistas y optimistas. En la segunda etapa, se construyó una red neuronal que integra variables macroeconómicas con los flujos y valores del fondo, logrando un error cuadrático medio (*MSE*) de validación de 0,0146. Comparado con el modelo de *Yale* utilizado por la empresa, la red neuronal mostró un desempeño predictivo similar, destacando su potencial, pero logrando un tiempo de ejecución considerablemente menor.

El proyecto demostró la viabilidad de utilizar datos históricos para generar escenarios que apoyen la gestión de riesgos en activos alternativos, especialmente en relación con el año de inicio de inversión. También se validó la utilidad de las redes neuronales para predecir flujos de caja y valor de fondos a mediano plazo. Se propone continuar investigando la predicción de la vida completa de un fondo mediante *deep learning* e incorporar métricas de riesgo como el VaR, junto con gráficos que faciliten la comprensión de los riesgos asociados a estos activos.

Dedicatoria

Este trabajo está dedicado a mis padres, quienes me han enseñado mediante el ejemplo varios valores y lecciones tales como la excelencia, autoexigencia y la constancia. También, lo útil e importante que es la educación, siendo esta una inversión a futuro y una herramienta para la vida. A mi hermano, que siempre fue claro sobre los pasos que el siguió para llegar donde esta y mantuvo buena disposición para apoyarme. A mis amigos, que creyeron en mí y siempre me motivaron para desafiarme, no rendirme y tener confianza en mí mismo, sobre todo en los momentos en que quizá me sentía cansado o incapaz. A mi novia, que siempre me apoyo y escucho durante estos largos años, brindando otras perspectivas y puntos de vista que me permitieron afrontar de mejor manera los desafíos que me puso la vida. También, dedicar este trabajo al resto de mi familia, que se mantuvo atenta a mis hitos y avances, siempre buscando mi bienestar. De igual manera, dedicar esta memoria a aquellos que ya no están presentes, pero cuyas presencias siempre vivirán conmigo, sirviendo como cable a tierra y ejemplos a seguir.

Por último, dedicar la carrera al Diego del pasado y del futuro. Ya que, a lo largo de cinco años y medio, logre constantemente demostrarme que pese a lo grande que muchas veces se veía el desafío, mediante consistencia, trabajo y exigencia, todo se puede. También, que esto sirva como ejemplo y recuerdo de que siempre se puede soñar e incluso cumplir dichos sueños, para eso es clave una correcta metodología y confianza en uno mismo.

Pasito a pasito.

Agradecimientos

Quiero partir agradeciendo a mis padres por su dedicación en la crianza de sus hijos, así como por siempre buscar enseñarme y guiarme. A mi padre Jorge, que fue clave en mis primeros años en la facultad, ayudándome a comprender la materia de plan común y siempre enseñando con el ejemplo. A mi madre Cecilia, que siempre se mantuvo preocupada de mi bienestar, buscando ayudarme independiente de lo que eso implicara. A mi hermano mayor Ignacio, que siempre busco aportarme con su perspectiva y experiencia.

También, agradecer a mi compañera y *partner* Martina que, junto con escucharme, apoyarme y ser siempre la fanática número uno de mi trabajo, me enseñó lo importante que es equilibrar la vida, no dejar que esta se torne en solo estudio, si no también disfrutar, compartir y querer.

Extender los agradecimientos a los profesores Carlos y Antonio, que mediante su buena disposición y experiencia me permitieron desarrollar un trabajo del cual me siento sumamente orgulloso.

Por último, agradecer a la empresa patrocinadora de este trabajo, que me brindó el equipo, espacio, tiempo y apoyo para adentrar en el mundo de las finanzas e inversiones que es lo que me apasiona, así como desenvolverme e investigar formas de innovar y desarrollar herramientas útiles.

Tabla de Contenido

Resumen Ejecutivo	i
Dedicatoria	ii
Agradecimientos.....	iii
Tabla de Contenido	iv
Índice de Figuras.....	v
Índice de Tablas.....	vi
1. Antecedentes Generales	1
1.1 Definición y Clasificación de Activos Alternativos	1
1.2 Inversión en Activos Alternativos.....	2
1.3 Métricas de Rendimiento.....	4
1.4 Actores en la Industria.....	5
1.5 La empresa	5
2. Descripción del Problema / Oportunidad.....	7
2.1 Modelo de Yale	7
2.2 Oportunidades en el Modelo.....	8
3. Justificación del Proyecto.....	10
4. Objetivos	11
4.1 Objetivo General	11
4.2 Objetivos Específicos	12
5. Alcances.....	12
6. Marco Conceptual	13
7. Metodología.....	16
7.1 Escenarios para las Curvas J.....	16
7.1.1 Investigación y Levantamiento de Ideas	16
7.1.2 Análisis Exploratorio y Preprocesamiento de Datos	17
7.1.3 Implementación de Escenarios y Comparación de Curvas	17
7.2 Run-off de Portfolios.....	18
7.2.1 Preprocesamiento de Datos.....	18
7.2.2 Padding de Datos	18
7.2.3 Implementación de las Redes Neuronales.....	19

7.2.4 Generación de Resultados y Gráficos.....	20
8. Desarrollo y Resultados	20
9. Discusiones	32
10. Conclusiones.....	35
11. Bibliografía	38
Anexos	41
Anexo A: Detalle Contribuciones, Distribuciones y NAV para el Primer Modelo	41
Anexo B: Ejercicio de Fondo con 10 Trimestres de Historia para el Primer Modelo.....	42
Anexo C: Detalle Contribuciones, Distribuciones y NAV para el Segundo modelo	44
Anexo D: Detalle Contribuciones, Distribuciones y NAV para el Tercer Modelo	45
Anexo E: Glosario	47

Índice de Figuras

Figura 1: Funcionamiento de un fondo de inversión de Private Equity.....	3
Figura 2: Representación flujos de caja de un fondo de inversión.....	3
Figura 3: Curva J	4
Figura 4: Análisis de escenarios para flujo de caja neto acumulado proyectado con modelo de Yale para buyout.....	22
Figura 5: Análisis de escenarios para NAV proyectado con modelo de Yale para buyout.	22
Figura 6: Análisis de escenarios para TIR proyectada con modelo de Yale para buyout.	23
Figura 7: Resultados primer modelo para cumulative contributions, cumulative distributions y NAV.....	24
Figura 8: Resultados primer modelo para TVPI.	25
Figura 9: Resultados segundo modelo para cumulative contributions, cumulative distributions y NAV.....	26
Figura 10: Resultados segundo modelo para TVPI.....	27
Figura 11: Resultados tercer modelo para cumulative contributions, cumulative distributions y NAV.....	28
Figura 12: Resultados primer modelo para TVPI.	29
Figura 13: Comparación modelo 3 con Takahashi.	31
Figura 14: Contribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 1).	41
Figura 15: Distribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 1).	41
Figura 16: NAV predicho para 8 trimestres a partir de 20 trimestres de historia (modelo 1).	42
Figura 17: Desempeño real de un fondo (18 trimestres).	42
Figura 18: Predicción del modelo a partir de fondo con 10 trimestres de historia y 10 trimestres de padding.....	43

Figura 19: Resultado concatenado de la predicción a partir de los 10 trimestres de historia real.	43
Figura 20: Contribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2).	44
Figura 21: Distribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2).	44
Figura 22: NAV predicho para 8 trimestres a partir de 20 trimestres de historia (modelo 2).	45
Figura 23: Contribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2 y 3).....	45
Figura 24: Distribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2 y 3).....	46
Figura 25: NAV predicho para 8 trimestres a partir de 20 trimestres de historia (modelo 2 y 3).	46

Índice de Tablas

Tabla 1: Comparación de los modelos tipo red neuronal y sus características.....	30
----------------------------------------------------------------------------------	----

1. Antecedentes Generales

1.1 Definición y Clasificación de Activos Alternativos

Se le llama activo alternativo al grupo heterogéneo de inversiones que se distinguen de las inversiones en acciones, bonos y efectivo que se negocian públicamente y a largo plazo, también conocidas como inversiones tradicionales (CFA Institute, s.f). Además, son características propias de los activos alternativos, la falta de liquidez y su amplia variedad, entre otras cosas. Respecto a los tipos o clases para este tipo de activos, en la literatura existen diferentes formas de clasificación. Para efectos de esta memoria se utilizarán las categorías que entrega el libro del CAIA ⁱ nivel I que se describen a continuación. (*Chambers et al., 2023*)

1. *Real Assets*: Inversiones en las que el subyacente involucra propiedad directa de activos no financieros. Se caracterizan por las implicancias operacionales que arrastra su propiedad, pues para generar retornos es importante involucrarse en las decisiones del día a día que tengan impacto en el desempeño del activo. Algunos ejemplos que recaen en esta categoría son los *recursos naturales* ⁱⁱ, *infraestructura* ⁱⁱⁱ y *propiedad intelectual* ^{iv}.

2. *Hedge Funds*: Vehículos de inversión privada con una naturaleza poco regulada que busca generar oportunidades de inversión que difieren de las que ofrecen los vehículos de inversión tradicional, los cuales están restringidos en su uso de derivados y apalancamiento. Dentro de esta categoría se pueden encontrar fondos de futuros, entre otros.

3. *Private Equity*: El término capital privado es usado por CAIA para referirse tanto al capital como a la deuda que no se tranzan públicamente. Las inversiones de capital privado emergen principalmente de financiar la formación de empresas nuevas, conocidas como *Venture Capital* ^v; del capital de compras apalancadas de empresas existentes, conocido como *Buyout* ^{vi}; de la financiación intermedia o *Mezzanine* ^{vii} de compras apalancadas de empresas; y de deuda en dificultades o también llamada *Distressed Debt* ^{viii}, que resultan del deterioro financiero de empresas anteriormente sanas.

4. *Structured Products*: Corresponden a instrumentos creados para exhibir retornos, riesgos, impuestos u otros atributos particulares. Este tipo de activos generan flujos de efectivo a partir de inversiones tradicionales o de anexar retornos del producto estructurado con uno o más valores de mercado.

1.2 Inversión en Activos Alternativos

A diferencia de los activos tradicionales que tranzan en la bolsa de valores, la inversión de activos alternativos se realiza a través de una sociedad de inversiones, comúnmente llamada *Limited Liability Partnership (LLP)* o *Limited Liability Companies (LLC)*, la cual, a su vez, invierte en diferentes fondos o activos que no son de oferta pública.

En estas organizaciones interactúan dos actores. En primer lugar, está el denominado *General Partner (GP)*, que es el encargado de administrar el fondo y buscar oportunidades de inversión. Por otro lado, se encuentran los *Limited Partners (LP)*, que corresponden a los inversores del fondo, que usualmente suelen ser de tipo institucional, vale decir, fondos de pensiones, *family office*^{ix} o compañías de seguros, entre otros.

Los fondos de inversión tienen un funcionamiento que se diferencia a otros tipos como lo son los fondos mutuos. En primera instancia, los *LP* se comprometen a desembolsar cierta cantidad de dinero, lo que se conoce como *commitment* o *committed capital*. Dicho compromiso puede ser llamado por el *GP* a lo largo de la vida del fondo. En línea con lo anterior, se identifican dos etapas en la vida de un fondo. En primer lugar, se encuentra la fase de inversión, en donde el *manager*^x del fondo busca oportunidades de inversión considerando el *commitment* realizado por los diferentes inversores. A medida que dichas oportunidades se van encontrando, el *GP* realiza *capital calls*, que corresponden al llamado de una fracción del capital comprometido para poder financiar las oportunidades de inversión encontradas. Posteriormente, se llega a la fase de liquidación, en donde dichas inversiones alcanzan la madurez y comienzan a generar retornos, en particular, una vez que el *manager* las liquida, el capital es devuelto a los inversionistas en forma de *distributions*^{xi}.

En la figura 1 se observa la relación entre el *GP* y los *LP* para constituir una *LLP*, y como los actores interactúan entre sí para concretar las inversiones. Por otro lado, en la figura 2, se representa con un diagrama los flujos de caja relacionados con un fondo de inversión, agregando el valor residual, que corresponde al valor de la inversión que todavía no se ha materializado como retornos y que por ende sigue en el fondo.

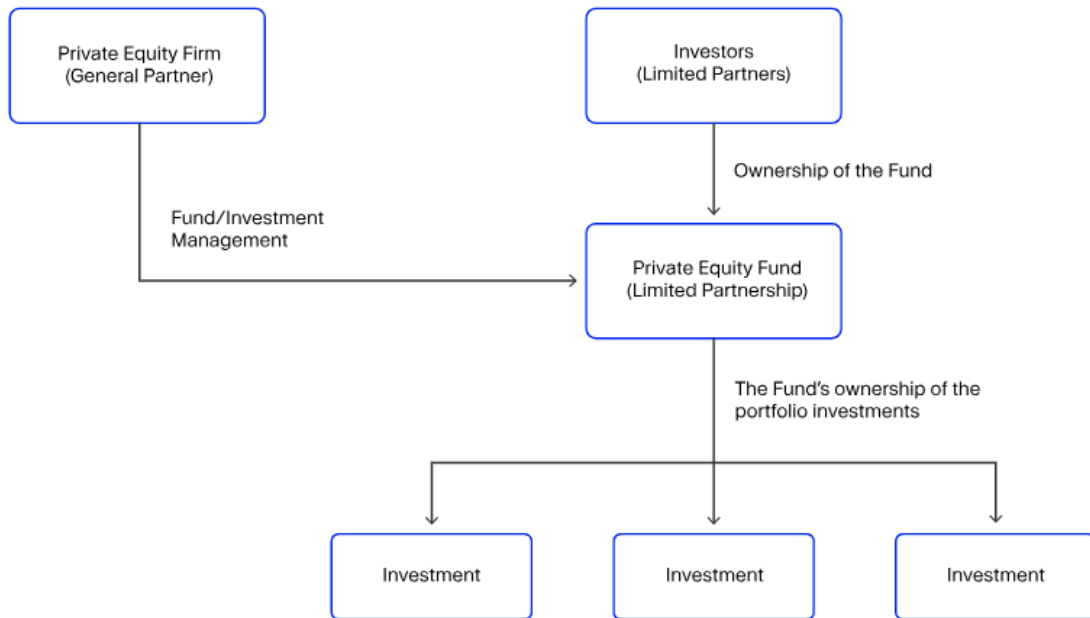


Figura 1: Funcionamiento de un fondo de inversión de Private Equity.

Fuente: "How does Private Equity Work?", Moonfare, Noviembre 2023

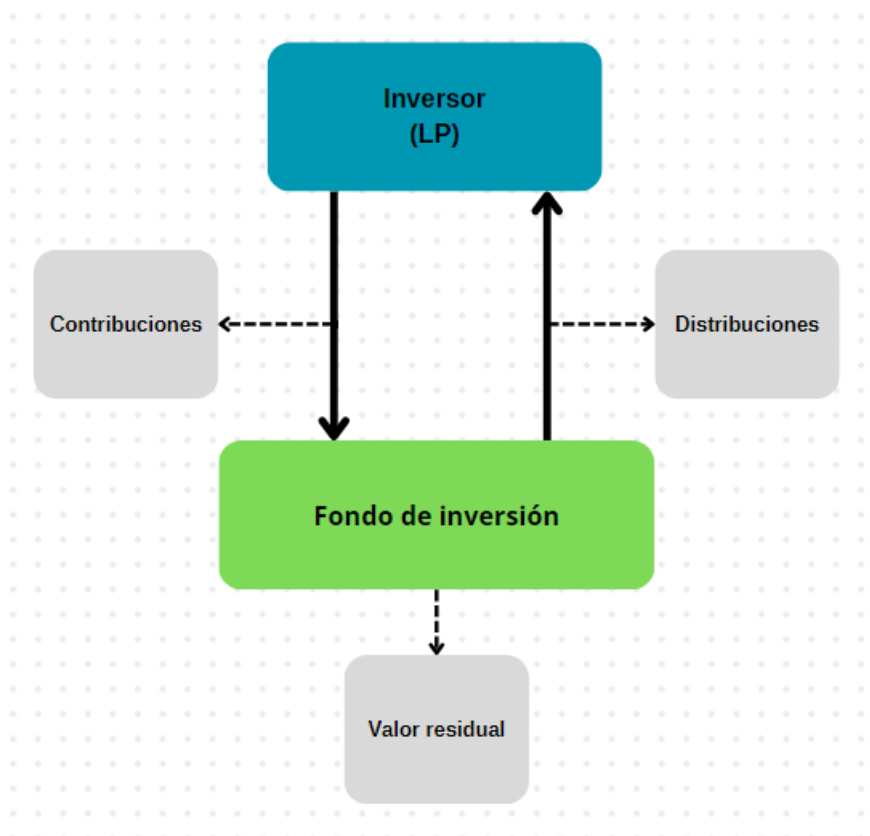


Figura 2: Representación flujos de caja de un fondo de inversión.

A partir de los flujos representados en la figura 2 es que surge la llamada Curva J o *J-Curve*, que consiste en una representación gráfica del comportamiento del flujo de caja acumulado de un fondo de inversión a lo largo del tiempo, considerando las contribuciones como un flujo negativo y las distribuciones como un flujo positivo para el inversor. En la figura 3, se puede observar una representación del fenómeno descrito anteriormente y cómo se conforma dicha curva. En particular, durante la fase de inversión, el flujo neto es negativo ya que, el inversor solamente ha recibido *capital calls*, mientras que a medida que el fondo va alcanzando su madurez y se pasa a la fase de liquidación, el inversor comienza a recibir *distributions*, generando que la curva comience a crecer. El concepto de la *J-Curve* es uno de los pilares centrales de esta memoria.

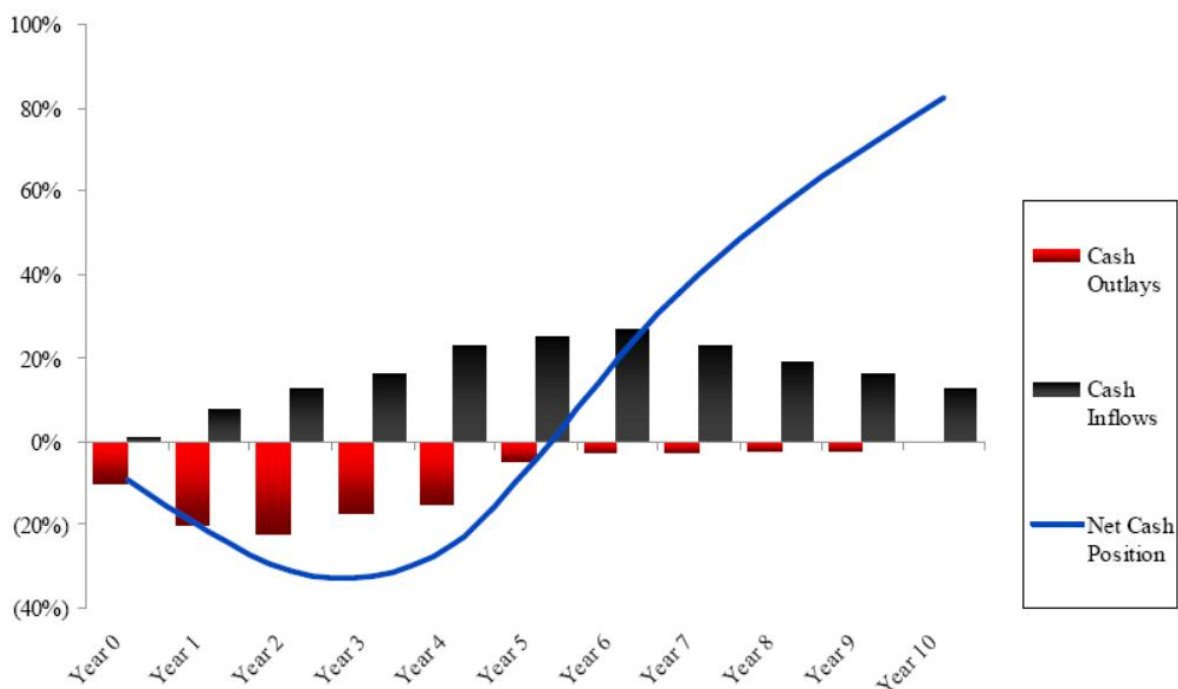


Figura 3: Curva J

Fuente: Curva J, Wikipedia, la enciclopedia libre.

1.3 Métricas de Rendimiento

En el mundo de los activos alternativos existen ciertas métricas que son un estándar para entender el rendimiento de los fondos. Estas se definirán a continuación.

1. *RVPI*: Corresponde al valor que aun reside dentro del fondo, es decir que no ha sido distribuido sobre el capital contribuido hasta la fecha por el inversor (Lehman, 2023).

$$RVPI = \frac{\text{Residual Value}}{\text{Cumulative Contributions}}$$

2. *DPI*: Similar al múltiplo ^{xii} anterior, corresponde al valor del fondo que ha sido distribuido en forma de distribución sobre el capital contribuido hasta la fecha por el inversor (Lehman, 2023).

$$DPI = \frac{\text{Capital Distrivuted}}{\text{Cumulative Contributions}}$$

3. *TVPI*: Se define como el valor total del fondo de inversión sobre el capital contribuido hasta la fecha por el inversor. Considerando como valor total a la suma del valor o capital distribuido y el valor residual (Lehman, 2023). Usualmente, se le refiere como el múltiplo de un fondo.

$$TVPI = \frac{\text{Residual Vuale} + \text{Distributed Capital}}{\text{Cumulative Contributions}} = RVPI + DPI$$

4. *IRR*: Corresponde a la tasa interna de retorno de un fondo o también conocida como TIR. Corresponde a la tasa de descuento que hace que el valor presente de todos los flujos de caja sea igual a 0 (Lehman, 2023).

$$IRR: \sum_t \frac{\text{Cashflow}_t}{(1+IRR)^t} = 0$$

1.4 Actores en la Industria

En Chile, el mercado de fondos de activos alternativos está dominado principalmente por asesores financieros que, a menudo, trabajan junto con distribuidores de fondos internacionales. Dentro de estos, se encuentran los bancos con fondos de su propia marca, administradoras generales de fondos y compañías de seguros, entre otras. Muchos de estos actores terminan entregando recomendaciones sesgadas, ya que los asesores tienden a favorecer fondos de sus empresas para obtener comisiones, creando un conflicto de interés conocido como "*agency problem*" (Chen, 2022). Este conflicto afecta la promoción de fondos de alto rendimiento que no ofrecen comisiones, llevando a una selección subóptima para los inversores. Una solución podría ser la intervención de empresas que ofrezcan los mejores fondos de forma imparcial, cobrando una tarifa plana basada en rendimiento y eliminando comisiones de parte de los distribuidores.

1.5 La empresa

El *sponsor* de este proyecto de título es una empresa especialista en activos alternativos que brinda asesoría financiera y servicios de *OCIO* ^{xiii}. Es decir, brindan asesoría financiera personalizada, centrada en la selección de fondos y sus *managers*, con el fin de maximizar los retornos de cada cliente, siempre cuidando el riesgo asociado. Esta empresa se posiciona dentro de la brecha descrita en el párrafo anterior, pues tiene acceso a algunos de los mejores administradores del mercado y sus fondos, ofreciendo

estrategias pasivas a sus clientes, centradas en el largo plazo y sin mayores cambios en el portafolio, lo que minimiza conflictos originados por comisiones relacionadas con el número de transacciones o comisiones por parte de distribuidoras de fondos. Además, el *sponsor* logra todo esto cuidando la diversificación tanto de la clase de activo como de la geografía. El foco geográfico de las inversiones de la empresa se centra en Estados Unidos y Europa como principales macrozonas. Por otro lado, respecto a las clases de activos en las que invierte, la empresa destaca en *real estate*, *infrastructure*, *private debt* y *private equity*, donde este último representa el 50% de su portafolio, en particular para *buyout*.

La empresa se compone por tres áreas principales. En primer lugar, inversiones, que son los encargados del diseño de portafolio, generación de reportería aparte del análisis de fondos y managers. También, cuenta con un área comercial la cual busca constantemente nuevos clientes y se preocupa de atender a los actuales clientes. Por último, el área de operaciones, que son los encargados de llevar el control de flujos y transacciones de los clientes. Estas tres áreas recaen bajo el alero del gerente general.

2. Descripción del Problema / Oportunidad

2.1 Modelo de Yale

Para comprender la base del proyecto, es esencial entender cómo se modelan los flujos de un fondo de activos alternativos. En este sentido, el *paper* "*Illiquid Alternative Asset Fund Modeling*" (Takahashi y Alexander, 2001) nos entrega un modelo financiero para proyectar y modelar los flujos de caja y el valor de fondos de inversión poco líquidos, como *Venture Capital*, *Real Estate* o *Buyout*. En particular, ofrece fórmulas para calcular contribuciones, distribuciones y el valor neto del fondo por periodos, generalmente trimestrales o anuales, permitiendo a los inversores evaluar el impacto de cambios en los niveles de compromiso o la modificación de otros supuestos. Es por esto, por lo que, la comprensión de los parámetros utilizados como entradas en el modelo es crucial, y estos se definirán a continuación.

1. RC (*Ratio* de contribución): Tasa que indica cómo evolucionan las contribuciones, específicamente cuanto porcentaje del capital comprometido va a ser llamado para cierto periodo.
2. CC (*Capital Committed*) [\$]: Capital que es comprometido por el *LP* para ser invertido en el fondo según vaya siendo requerido por el *GP*.
3. L (*Life of the fund*) [años]: Horizonte de vida del fondo.
4. B (*Bow*): Tasa de ajuste de las distribuciones.
5. G (*Annual growth rate*) [%]: Tasa de crecimiento anual del fondo.
6. Y (*Yield*) [%]: Tasa de rendimiento del fondo
7. RD (*Ratio* de distribución): Tasa que indica cómo evolucionan las distribuciones con el tiempo en base al performance del fondo.
8. *Age*: Corresponde al periodo sobre el cual se está iterando en el modelo y a la edad del fondo modelado en ese instante.

De estos parámetros, tanto el *ratio* de contribución, el *ratio* de distribución y el *bow* son parámetros exclusivos del modelo. El primero corresponde a la tasa con la cual, en cada periodo, se contribuye capital al fondo respecto al *unfunded*^{xiv} o capital comprometido aún no llamado. Por ejemplo, Takahashi propone los siguientes valores para *RC*, 25% para el primer año, 33% para el segundo año y 50% para los siguientes años. Por otro lado, el *ratio* de distribución corresponde a la tasa con la cual el fondo distribuye a los *LP*, esto sobre el valor del fondo en determinado periodo. En particular, el *RD* tiene la siguiente forma.

$$RD = \text{Max} (Y, (Age_{t-1} / L) ^B)$$

Finalmente, el *bow* es un parámetro de control, que explica cómo cambia el *ratio* de distribución a lo largo del tiempo, donde un B alto indica que el fondo distribuirá lentamente durante los primeros años del fondo y rápido para los años finales de este.

Respecto a las variables *output*^{xv} del modelo. En primer lugar, las contribuciones, se plantean proporcionales al capital no llamado (*Uncalled Capital*) del periodo anterior y al *ratio* de contribución, que corresponde a un parámetro propio del modelo.

$$C_t = UC_{t-1} \times RC (Age_{t-1})$$

De forma análoga, las distribuciones, se modelan como un monto proporcional al NAV del periodo anterior de la siguiente manera.

$$D_t = NAV_{t-1} \times (1+G) \times RD (Age_{t-1}, Bow, L)$$

Por último, para el NAV o *net asset value*, el autor postula que esta variable depende de su valor en el periodo anterior, de la tasa de crecimiento del fondo, de las distribuciones y contribuciones del respectivo periodo.

$$NAV_t = NAV_{t-1} \times (1+G) + C_t - D_t$$

2.2 Oportunidades en el Modelo

Dado el funcionamiento de los fondos de inversión, ser capaz de predecir en qué momento y cuánto capital será llamado por el *GP* resulta de suma importancia para los *LP*, ya que les permite a los inversores prepararse y disponer de dicho monto para cuando se requiera. Lo mismo pasa con los montos a ser distribuidos por el *GP*, pues, en un escenario ideal, los flujos que son retornados a los inversores se pueden utilizar para cumplir con los llamados de capital. Es por esto por lo que, muchos actores de la industria acuden a la literatura en búsqueda de soluciones. Actualmente, existen alternativas estocásticas y sumamente complejas, sin embargo, la gran mayoría de modelos para este fin tienen su origen en el ya mencionado *paper* “*Illiquid Alternative Asset Fund Modeling*” o también llamado “*Yale Model*”, debido a su simplicidad e interpretabilidad (Takahashi y Alexander, 2001).

Si bien este modelo cumple la función de lograr estimar las variables para cada periodo de forma sencilla, los parámetros que utiliza muchas veces son determinados por exploración o mediante una optimización. Esto puede derivar en un distanciamiento entre

el valor que toman y el significado práctico o financiero que tienen, lo que se puede traducir en modelos sobre ajustados o variables sin capacidad predictiva.

De igual manera, con situaciones financieras como la ocurrida el 2008, que derivó en la caída de los mercados y años complicados para la industria, donde los fondos de inversión con *vintage*^{xvi} anterior a la crisis o, en otras palabras, fondos que comenzaron a invertir en los años previos a la recesión (por ejemplo 2006 o incluso 2007), esperaban ver retornos para el año 2008 y en consecuencia distribuir a sus inversores, pero se encontraron con un mercado golpeado, en el que no existían compradores dispuestos a arriesgarse. Esto se tradujo en un gran impacto para los inversionistas que no fueron capaces de recuperar sus inversiones. De aquí se desprende la importancia de ser capaz de identificar un mal o buen *vintage* para invertir en este tipo de fondos, a modo de resguardarse ante los riesgos asociados al ciclo económico.

En el contexto del *sponsor*, el modelo de Takahashi es la base teórica sobre la cual se construyen las curvas J. Estas a su vez, son las herramientas que utilizan para elaborar los programas de inversión de los clientes y proyectar sus *portfolios*^{xvii}, tarea que, de ser mal ejecutada, sin tener el cuidado de controlar los riesgos asociados o presentar imprecisiones en los cálculos, termina por impactar directamente a la liquidez del cliente. De hecho, hay que recordar que algunos de los clientes de la empresa son inversores institucionales, tales como, compañías de seguros o fondos previsionales, entre otros. Este tipo de inversionista está altamente regulado, teniendo límites de inversión definidos para los diferentes tipos de activos, dentro de ellos los alternativos. Por ejemplo, para el caso de las AFP^{xviii}, dichas cotas son de entre un 5% y 15% de cada multifondo^{xix}, dependiendo de la rentabilidad y riesgo buscado (Infante, 2024). Por ello, para estas instituciones es sumamente importante ser capaces de predecir con exactitud las distribuciones y contribuciones a realizar a fondos de inversión ya que, si estos resultados son erróneos o imprecisos, podrían derivar en una venta de la posición en el fondo, es decir vender parte de la propiedad en dicho fondo, esto con el fin de no caer en incumplimiento de los límites definidos por la ley, lo que termina por generar pérdidas en la rentabilidad de estos inversores e impactando a las personas que invierten con ellos.

En línea con lo anterior, otro tipo de cliente de la empresa son las familias o los *family office*, que, si bien presentan bastante capital a la hora de invertir, no tienen el nivel de sofisticación o dominio técnico, ni tampoco el capital que tienen los inversores institucionales y profesionales como las AFP o las compañías de seguros. Esto se traduce en que, al invertir en activos poco líquidos como lo son los alternativos, para estos inversores los riesgos asociados se transforman en una de las mayores preocupaciones. Generando así a la empresa el desafío de informar y explicar los riesgos asociados a este tipo de inversión, todo de forma clara y sencilla, sin caer en tecnicismos que dificulten la recepción del mensaje.

Por último, es importante recalcar que la tarea de proyectar programas de inversión utilizando el modelo de Takahashi es una tarea que toma mucho tiempo y recursos computacionales debido al gran número de cálculos y optimizaciones que se deben realizar. Esto termina siendo peor cuando la proyección se realiza sobre *portfolios* de mayor tamaño como los de clientes institucionales o también cuando este proceso se debe realizar para un gran número de clientes. Por lo que si se busca aumentar la escala en que se brinda este servicio, es necesario buscar formas más eficientes de llevar a cabo las proyecciones para programas de inversión ya existentes.

3. Justificación del Proyecto

El proyecto por abordar en esta memoria se divide en dos ejes. En primer lugar, se buscará perfeccionar la metodología ya existente que es utilizada para modelar los rendimientos esperados para las diferentes clases de activos. En particular, se incorporarán escenarios a las curvas construidas con el modelo de Takahashi. Estos escenarios serán generados a partir de datos reales de un universo de fondos, que serán seleccionados especialmente para representar momentos históricos de mal desempeño, o, en otras palabras, se buscará agregar al modelo actual los comportamientos que tuvieron fondos con *vintage* previo a las crisis. Lo anterior debido a que, un fondo que comienza a invertir los años antes de una crisis verá los efectos de esta en su fase de liquidación, es decir, cuando debiese comenzar a distribuir a los inversores. Esta información actualmente no está incorporada al modelo de *Yale*. Para lograr esto, se van a construir curvas con base en la mediana y los percentiles de datos históricos de fondos con determinados *vintage*, tales como podrían ser el 2000, 2006 y 2020, a modo de contrastar la curva J estimada con el modelo de Takahashi, con los flujos empíricos de malos años y así extraer conclusiones de posibles escenarios pesimistas.

En segundo lugar, se buscará explorar una nueva metodología para el proceso de proyección de portafolios que presentan flujos ya realizados. Actualmente, se utiliza el mismo modelo de Takahashi bajo una serie de supuestos y restricciones, en conjunto con los montos a la fecha de los *capital calls* y *distributions* ya efectuados, para así, proyectar los valores futuros de estas variables, junto al NAV de los fondos que constituyen dichos *portfolios*. El proceso actual demanda mucho poder computacional y tiempo por la naturaleza iterativa del modelo de *Yale*, lo que se traduce en un proceso que toma minutos. Es por esto por lo que se busca explorar una alternativa más eficiente en estos aspectos. Específicamente, se buscará generar una herramienta análoga a la planteada en el trabajo investigativo “*Supervised Neural Networks for Illiquid Alternative Asset Cash Flow Forecasting*” (Karatat et al., 2021), pero configurada al contexto de la empresa. En otras palabras, se intentará generar una red neuronal^{xx}, que sea capaz de recibir los flujos de los fondos que ya fueron realizados en conjunto con variables macroeconómicas, para lograr predecir las contribuciones, distribuciones y valor del fondo para los próximos años.

Suponiendo que el proyecto se logra implementar, ambas dimensiones abarcan mejoras significativas para la empresa. Por un lado, se perfeccionaría la herramienta que actualmente se utiliza para modelar los rendimientos esperados de diferentes clases de activos, incorporando nuevas fuentes de información histórica, tales como desempeños de fondos en malos años para invertir. Lo anterior se traduce en valor agregado para los clientes y el equipo, en términos de la visualización y manejo de los riesgos asociados a invertir en un fondo de este tipo, en particular al identificar el riesgo asociado al *vintage* mediante la generación de estos escenarios, todo esto apuntando a una solución sencilla y clara de comprender.

Por otro lado, al implementar y explorar una nueva metodología para la proyección de *portfolios*, se busca generar resultados precisos, eliminar los supuestos iniciales del modelo de Takahashi y hacer el proceso más rápido. Lo que es clave para escalar el negocio y la cartera de clientes que actualmente maneja la empresa. También se busca impactar desde el punto de vista comercial, incorporando conceptos como el autoaprendizaje computacional y el uso de variables macroeconómicas. Además, al desarrollar una nueva herramienta para la tarea de predicción, se abre la oportunidad de complementar el modelo actual y sus resultados con una segunda opinión proveniente de un modelo completamente diferente que incorpora otras fuentes de información. De esta forma, se podría entregar a los clientes las predicciones de ambos modelos, a modo de que estos reciban la mayor cantidad de datos posible, para así aportar la toma de decisiones de estos. En particular, lo anterior sería útil para clientes de tipo institucional, que requieren de un alto nivel de rigurosidad y exactitud debido a sus estrictas cotas de inversión. Por último, en esta misma línea, dados los cambios en materia legislativa, se espera que los límites de inversión antes mencionados incrementen en el mediano plazo, lo que resalta la importancia de contar con modelos y recomendaciones de inversión que se adapten a las necesidades de inversores de este tipo, esto con la intención de poder potenciar este segmento de clientes que podría ser altamente atractivo para el *sponsor* (Infante, 2024).

4. Objetivos

4.1 Objetivo General

El objetivo general de este trabajo de título es incorporar escenarios al modelo actual de la empresa para modelar las clases de activos y desarrollar una alternativa computacional para proyectar carteras con flujos realizados, beneficiando así a la empresa y a sus clientes.

4.2 Objetivos Específicos

- Desarrollar un programa en *Python*^{xxi} para calcular y graficar las curvas tipo escenario para los diferentes *vintage* seleccionados.
- Generar un documento tipo informe, que contenga la forma en que se hicieron los cálculos, modos de uso y el análisis de las curvas generadas con Takahashi respecto a los escenarios.
- Desarrollar un programa en *Python* que contenga el modelo computacional capaz de proyectar las contribuciones, distribuciones y valorizaciones de fondos a partir de flujos previamente realizados.
- Generar un documento tipo informe, en el cual se explique el funcionamiento del modelo, la forma en que se entrenó y una comparación entre los resultados predichos y los valores reales.

5. Alcances

Del presente proyecto se esperan dos cosas. En primer lugar, agregar valor a la metodología que actualmente es utilizada por el *sponsor* para modelar las diferentes clases de activo. En particular, se busca incorporar información histórica del desempeño de los fondos a las curvas generadas bajo la metodología de Takahashi, aportando así al equipo de inversiones una herramienta que les permita identificar de forma visual y sencilla, posibles riesgos o alertas relacionadas con el desempeño de diferentes clases de activos. Por otro lado, se espera crear una nueva herramienta específica para la proyección de portafolios con flujos realizados, que sea capaz de predecir los valores de las variables con un error bajo y en un tiempo menor que las proyecciones generadas con el modelo actual basado en las curvas de Takahashi. De esta manera, se busca desafiar el *statu quo* del modelo utilizado actualmente por la empresa para proyectar flujos mediante una nueva metodología.

Para medir el éxito de ambos ejes del proyecto, se espera que ambos informes elaborados para mostrar los procesos de construcción, modo de uso de las herramientas, resultados de estas, junto al análisis de ambos programas, sean presentados en un comité de desarrollo y posteriormente aprobados por el equipo de inversiones.

En cuanto al proyecto, se decidió abordar las dos líneas mencionadas anteriormente por varias razones. En primer lugar, el tiempo asignado es limitado a un semestre y en modalidad de media jornada, lo que hace poco viable desarrollar e implementar un nuevo modelo que reemplace al modelo de *Yale* para modelar las diferentes clases de activos. En su lugar, se optó por mejorar el modelo actual mediante la creación de escenarios para las curvas generadas. Además, al *sponsor* le interesaba explorar nuevas metodologías para la proyección de portafolios y utilizar algoritmos computacionales avanzados, pero ante la restricción de tiempo y la falta de experiencia

en modelos computacionales complejos por parte del alumno, se optó por seguir la metodología descrita en el artículo “*Supervised Neural Networks for Illiquid Alternative Asset Cash Flow Forecasting*” (Karatas et al., 2021), adaptándola a la realidad de la empresa, en vez de diseñar e implementar un modelo completamente nuevo.

Por último, es relevante entender que aspectos quedan fuera del trabajo realizado. Para partir, queda fuera de los alcances del proyecto de título, los usos y las modificaciones futuras que se le puedan dar a los escenarios desarrollados para las curvas del modelo de Takahashi. También, en esta memoria no se abarcará en profundidad la explicación del código de *Python*, más bien, el trabajo realizado se centrará en los resultados obtenidos a partir de este y en detalles claves de la metodología utilizada. Por último, respecto al uso de redes neuronales para la proyección de *portfolios* con historia, es relevante entender que este trabajo se centra en una exploración de una metodología para predecir flujos en el corto plazo (dos años u ocho trimestres), por lo que usos y modificaciones posteriores que se le puedan dar al modelo para lograr poder predictivo de largo plazo, quedan fuera del alcance del proyecto de título. Sin embargo, esto último es un área de investigación y desarrollo con mucho potencial, por lo que este tema quedará propuesto en las discusiones.

6. Marco Conceptual

En el contexto de proyectos relacionados con el desarrollo de modelos financieros para predecir flujos de caja, se emplean diversas herramientas provenientes de la ingeniería industrial. A continuación, se describen algunas de ellas.

En primer lugar, dentro del área del análisis de datos, es importante conocer el concepto del análisis exploratorio de datos o *EDA (Exploratory Data Analysis)*. Corresponde al conjunto de técnicas y prácticas que se realizan sobre los datos disponibles previo a su utilización, para así, ser capaz de identificar, posibles errores, anomalías, relaciones y la estructura de estos. Dentro de las técnicas contempladas dentro del *EDA* se tienen, los histogramas, matrices de correlación, estadísticas descriptivas como los promedios, medianas, percentiles y medidas de dispersión (IBM, s.f).

También es importante entender el preprocesamiento de datos. Este paso sigue al análisis exploratorio y es crucial antes de implementar un modelo o calcular métricas con los datos. El preprocesamiento implica limpiar, transformar e integrar los datos para prepararlos para una tarea específica. Incluye eliminar datos nulos o no deseados, integrar con otras tablas, crear nuevas variables y normalizar los datos, entre otras prácticas importantes (GeeksforGeeks, s.f.).

Ambos procedimientos son fundamentales en proyectos de este tipo, pues preceden al desarrollo y análisis, optimizando el uso del tiempo y recursos del equipo, aparte de ser una forma de obtener información de manera proactiva, en lugar de reactiva, mejorando la eficiencia durante el desarrollo. Sin embargo, para ser efectivo, el EDA y el preprocesamiento, requieren de datos voluminosos, tiempo y organización; de lo contrario, pueden resultar contraproducentes e innecesarios.

También es relevante el manejo de conceptos dentro del área de la estadística, específicamente métricas como el *MSE (Mean Square Error)*. Esta se utiliza para evaluar la precisión de un modelo, por ejemplo, una red neuronal. Este corresponde al promedio de la diferencia entre cada observación y su valor predicho al cuadrado. El *MSE*, es una de las formas más utilizadas y estandarizadas para evaluar el ajuste de un modelo, debido a su fácil interpretabilidad y modo de cálculo. Sin embargo, se caracteriza por ser altamente sensible a *outliers*^{xxii}, lo que puede llevar a distorsionar la evaluación de los resultados respecto a otras métricas de precisión.

En proyectos como este, se utilizan diversas técnicas de optimización para obtener parámetros óptimos de modelos, calcular métricas como la TIR y suavizar series de datos. Una de estas técnicas es la minimización del error cuadrático medio (*MSE*), que busca reducir la discrepancia entre el valor estimado y el valor real de una variable. En el contexto de las curvas de Takahashi, esta técnica se emplea para obtener los mejores parámetros mediante la minimización del *MSE* con respecto a los datos reales de los fondos. Se elige esta metodología por su rapidez y eficiencia computacional al trabajar con grandes volúmenes de datos. Sin embargo, es importante considerar que esta técnica es sensible a *outliers*, lo que puede afectar la captura de fenómenos con valores extremos, favoreciendo resultados más cercanos al promedio de los datos.

Ahora bien, por el lado de la ingeniería financiera, también es relevante definir algunos conceptos. Para partir, en esta memoria se habla de dos tipos de modelamiento diferentes, por lo que es relevante hacer la distinción. En la primera parte del proyecto, se profundiza en cómo mejorar el modelado de las *J-curves* correspondientes a cada clase de activo. Esto se refiere a, la utilización del modelo de Takahashi para generar una representación de la vida completa de un fondo de inversión para un *asset class* determinado. Por otro lado, el modelo de Yale también se utiliza para hacer proyecciones a futuro de portafolios existentes. A partir de ahora, lo anterior será conocido como *run-off*. Este proceso consiste en el uso del modelo de Takahashi bajo una serie de supuestos, para poder proyectar el rendimiento futuro de portafolios que ya cuentan con flujos realizados para los fondos que lo componen. Es decir, a partir de valores conocidos para las contribuciones, distribuciones y NAV de los fondos, en conjunto con las *J-curves* para los *asset class* correspondientes, se genera una predicción para los valores de estas mismas variables en los próximos trimestres.

La elección de la metodología y el modelo planteado por los autores Takahashi y Alexander para generar las curvas correspondientes a las diferentes clases de activos y el *run-off*, tiene que ver con su simplicidad y alto margen de configuración a través de parámetros en relación con opciones más complejas. Pero, son estas mismas cualidades lo que hacen mejorable. Por ejemplo, en cuanto a la precisión de estos modelos, ya que los resultados son demasiado sensibles a los valores que toman los parámetros, por lo que, si estos son mal determinados al inicio, se produce un error de arrastre para todos los procesos consecutivos. Por otro lado, al ser un modelo iterativo, puede ser computacionalmente demandante y algo lento.

Pasando al área de la computación, para la segunda parte de este trabajo, es importante tener claras las siguientes definiciones. Una red neuronal es un modelo computacional basado en el cerebro humano, que busca relacionar miles o millones de nodos de procesamiento entre sí. Tienen la particularidad de funcionar con un sistema de pesos entrenables y una función de activación para cada nodo. Lo anterior, le permite a la red “entrenarse” o ajustar sus pesos de modo de minimizar la pérdida de las funciones de activación. Este tipo de sistemas se construye mediante capas de diferentes características que poseen neuronas o nodos. Su modo de uso se basa en definir las capas a utilizar para la estructura y posteriormente entrenarla usando *data* a modo de ajustar los pesos para obtener la mejor predicción posible (Huet, 2023). Además, con el tiempo se han desarrollado diferentes tipos de redes, cada una pensada para un propósito distinto. Por ejemplo, las redes neuronales recurrentes o *RNN*, que tienen la característica de recoger y recordar resultados previos para ser utilizados en cálculos futuros. En particular, en esta memoria se trabajará con un subconjunto de estas redes, que con las *LSTM* o *Long Short-Term Memory Neural Networks*. Estas tienen la particularidad de ser una mejora respecto a las *RNN* convencionales respecto al uso de la memoria, logrando centrarse en la información útil e ignorando información no relevante, todo esto para lograr recoger tendencias pasadas para usos futuros (Coursera, 2024). Este tipo de modelos computacionales son útiles para tareas como la predicción de flujos debido a su capacidad de modelar relaciones complejas entre una gran cantidad de variables. Sin embargo, tienen la desventaja de que funcionan como una *black box*, ya que el proceso de cálculo y optimización es poco configurable por el usuario, además de poco intuitivo.

En esta misma línea, es importante definir algunas prácticas implementadas sobre los datos para obtener todo el potencial de estas poderosas herramientas computacionales. Tal es el caso del *padding*, que corresponde a la imputación de datos con el objetivo de que las entradas del modelo con largo variable queden de igual longitud. Usualmente, se utilizan ceros a modo de relleno, en complemento con una máscara^{xxiii}, permitiendo que el modelo ignore los valores del *padding*, realizando las optimizaciones únicamente con los datos reales (Sanjay, 2023). En el contexto del modelo abarcado en este proyecto, el proceso de *padding* es fundamental, dado que, la

data disponible respecto al desempeño de diferentes fondos no es uniforme y tienen diferentes largos. Una vez realizado este proceso, se obtienen series de igual longitud, lo que compatibiliza los datos con la red, permitiendo así un entrenamiento adecuado de esta.

Por último, para comprender el desarrollo del modelo, es clave definir el enfoque de ventana deslizante. Este concepto hace referencia a la división de los datos en dos partes. Primero, en lo que se conoce como ventana de entrada, que corresponde al conjunto de datos que recibirá el modelo como *input*. Por otro lado, está la ventana de salida, que corresponde a los datos que buscara predecir el modelo a partir de la ventana de entrada. La particularidad se encuentra en que ambas ventanas se van desplazando, esto se realiza a modo de maximizar los datos disponibles para entrenar el modelo y para que la red neuronal aprenda de tendencias temporales implícitas en los datos (Collazos, 2023). A modo de ejemplo, si se tiene data correspondiente a 5 años de historia en formato trimestral para 20 fondos y se define una ventana de entrada de 13 trimestres junto con una ventana de salida de 4 trimestres, entonces el modelo contara para cada fondo, con 3 conjuntos de datos, cada uno dividido en ventana de entrada y de salida. Donde para cada uno de los fondos, el primer conjunto corresponde a los trimestres 1 al 18, el segundo a los trimestres 2 al 19 y el tercer conjunto del *quarter* 3 al *quarter* 20.

7. Metodología

7.1 Escenarios para las Curvas J

Para llevar a cabo el presente trabajo de título se presenta una serie de pasos para cada etapa del proyecto. Partiendo por la generación de escenarios para las *J-Curves* de los distintos *asset class*, la metodología realizada fue la siguiente.

7.1.1 Investigación y Levantamiento de Ideas

Para comenzar, el estudiante se debió familiarizar con el modelo y el programa que actualmente utiliza la empresa para modelar las diferentes clases de activos. Para esto, tuvo que leer el artículo explicativo de los autores Takahashi y Alexander, para entender la teoría detrás del desarrollo actual. También, revisó la documentación y los códigos de *Python* con los cuales se realizan los cálculos de las métricas relevantes. En esta misma línea, también es importante revisar literatura y artículos para conocer que otras prácticas se están realizando en la industria y en la academia. Para aquello, el memorista se dedicó a buscar *papers* en los que se abarquen nuevas versiones del modelo de *Yale*, así como nuevas alternativas que generen valor relacionado al proyecto.

7.1.2 Análisis Exploratorio y Preprocesamiento de Datos

Concluido lo anterior, el enfoque se centró en trabajar sobre los datos disponibles. Al respecto, la empresa cuenta con *data* sobre la evolución de los flujos de caja y valorizaciones a lo largo del tiempo para más de 200.000 fondos, lo que se torna en una dificultad al momento de evaluar la calidad de estos datos, así como también llevar a cabo desarrollos con estos.

Para hacer una evaluación inicial de los datos disponibles, el estudiante realizó un *EDA* usando *Python* sobre una tabla extraída directamente de un proveedor de datos de fondos de inversión. Algunas de las conclusiones levantadas a partir del análisis exploratorio fueron las siguientes. En primer lugar, se encontró que muchos de los fondos seguidos por el proveedor no eran de regiones geográficas o de monedas de interés para el *sponsor*. También, se encontró que muchos de los fondos tenían montos registrados atípicos para las variables de interés. Además, dentro de otras cosas, se encontraron errores en la *data* asociados a los signos de ciertos números, por ejemplo, distribuciones negativas o contribuciones positivas, lo que carece de sentido financiero.

Una vez terminado el *EDA*, se procedió con el preprocesamiento de los datos, enfocado en aliviar la base sobre la cual trabajar, eliminar errores y estandarizar el formato de estos. Algunos de los tratamientos realizados fueron, filtros sobre la moneda, la región y el tamaño del fondo. Se seleccionaron los campos anteriores por sobre otros, como el *manager* o administrador del fondo ya que, se buscó construir una base de datos suficientemente amplia para ser estadísticamente significativa, pero que mantuviese al *vintage* como la variable de estudio. Dicho lo anterior, el efecto de variables como el *GP* de un fondo sobre su desempeño no deja de ser un tópico interesante por estudiar, en particular en el mercado privado, donde las diferencias entre un actor y otro generan más diferencias que en su contraparte pública. Volviendo al preprocesamiento de datos, también se eliminaron contribuciones positivas, distribuciones negativas y *outliers* que correspondan a datos más allá de 3 desviaciones estándar respecto a su media. En la misma línea, se calcularon los flujos acumulados para las contribuciones y distribuciones, así como, también se crearon nuevas variables correspondientes al *TVPI*, *RVPI*, *DPI* y a la *TIR* a lo largo del tiempo para cada fondo respectivamente.

7.1.3 Implementación de Escenarios y Comparación de Curvas

Con la base de datos construida, el estudiante, en conjunto con los gerentes de la empresa y el tutor, realizaron la selección de los *vintage* a utilizar como escenarios pesimistas y optimistas. Luego, para construir las curvas, el estudiante calculó métricas descriptivas como la mediana para cada fecha disponible del universo de los fondos seleccionados. De esta forma, se obtuvo una serie temporal que representa a varios fondos con un *vintage* determinado. Este ejercicio se realizó para cada año seleccionado. Para llevar a cabo lo anterior, el estudiante desarrolló un programa usando *Python* en el

que se realizan dichos cálculos y los resultados se almacenen en un arreglo de datos para ser graficados. Dado que se grafica una serie discreta de datos, se realizaron interpolaciones y aproximaciones utilizando polinomios de diferentes grados sobre los datos, para así obtener curvas suavizadas que, luego, fueron comparadas con las generadas por Takahashi, esto utilizando el mismo programa.

7.2 Run-off de Portfolios

Para esta segunda fase del proyecto de título, se explicará el paso a paso que se siguió para construir el modelo para el *run-off*. La metodología fue la siguiente.

7.2.1 Preprocesamiento de Datos.

De forma análoga a la generación de los escenarios para las *J-curves*, se utilizó la misma base de datos con información para más de 200.000 fondos, levantada por un proveedor de datos financieros, pero tomando en consideración solo una clase de activos. Por lo que, aprovechando el trabajo realizado previamente, el estudiante utilizando *Python*, realizó un preprocesamiento análogo al descrito en la sección anterior, aplicando los filtros respectivos, creando los flujos acumulados, escalando las variables en formato de *ratio*, eliminando *outliers* y errores en la *data*, aparte de filtrar el *asset class*, resultando en información limpia para 413 fondos distintos de *Buyout*.

Adicional a lo anterior, el estudiante agregó la información en formato trimestral, teniendo el cuidado de, para cada fondo, sumar las contribuciones y distribuciones realizadas en un mismo *quarter*, pero tomando la última valorización disponible en el trimestre para la variable del NAV. Luego, dado que se trabajó con redes neuronales, fue necesario tratar los datos nulos, para esto, con ánimos de mantener la mayor cantidad de datos disponibles, los valores nulos o vacíos en las variables de interés se reemplazaron con 0. Por último, se realizó un filtro de densidad por *vintage*. Es decir, para cada *vintage* de la *data*, se determinó el número máximo de registros o periodos que tenía algún fondo y se eliminaron todos los fondos que tuvieran menos registros que un determinado porcentaje del máximo encontrado, por ejemplo 40%.

7.2.2 Padding de Datos

Dado que se trabajó con redes neuronales, fue necesario realizar procesos específicos para tener los datos en un formato que fuera compatible con el modelo implementado y el fin predictivo buscado. Uno de dichos procesos fue el *padding* de la *data*. Para realizarlo, se agruparon los datos por *vintage* y para cada grupo se buscó al fondo con más información, o, en otras palabras, aquel fondo más largo para su grupo. Luego, se creó una función que, para cada grupo, replicara el largo del fondo más largo al resto de fondos. Es decir, que sobre aquellos fondos más cortos que el fondo encontrado previamente, se les agreguen filas al final con valores 0 en las variables de

interés (contribuciones acumuladas, distribuciones acumuladas y NAV). Después, se tomó como supuesto que, todos los fondos de un mismo *vintage* parten en el mismo mes. De esta forma, es factible replicar las fechas de cada fila del fondo más largo en el resto de los fondos, logrando estandarizar el formato de todos los fondos de un mismo grupo.

7.2.3 Implementación de las Redes Neuronales

Una vez concluido lo anterior, a modo de experimentación, se crearon 3 modelos diferentes, todos de tipo red neuronal, pero diferentes en cuanto a las variables utilizadas o el universo de fondos con el que fueron entrenados. Primero, se creó una red que recibe como entrada 20 trimestres de historia para las contribuciones acumuladas, distribuciones acumuladas y NAV, logrando retornar una predicción de 8 trimestres para estas mismas variables. Este primer modelo fue entrenado únicamente con fondos de un *vintage*, por ejemplo 2008. Esto con el objetivo de poder trabajar con una escala más pequeña de datos. Luego, se pasó a desarrollar un segundo modelo análogo al anterior, pero entrenado utilizando fondos con *vintage* entre el año 2000 y 2017. Para lograr esto, hay que aclarar que se agregó el *vintage* como dato de entrada, adicional a las variables de interés. Por último, se creó un tercer modelo, que incorpora a la entrada dos variables macroeconómicas, de ahora en adelante llamadas “macro1” y “macro2”, resultando así en, una red neuronal que recibe 20 trimestres de historia de las variables por predecir, sumado al *vintage* del fondo por predecir y las dos variables macros en los periodos correspondientes, para luego retornar 8 trimestres predichos para las *cumulative contributions*, *cumulative distributions* y NAV del fondo respectivamente.

Para cada uno de estos modelos la implementación fue análoga, teniendo algunas consideraciones. Dicho proceso consistió en los siguientes pasos. Primero, se definió el conjunto de entrenamiento y validación siguiendo un *split*^{xxiv} del 80% de los fondos disponibles para entrenar el modelo y el 20% restante para testear los resultados.

Luego, para ambos conjuntos creados anteriormente, se crearon las ventanas deslizantes. Esto se logró creando funciones para cada modelo, que separaran la *data* de los fondos de la siguiente manera. Las series de cada fondo fueron separadas en una ventana de entrada y una ventana de salida, la primera correspondiente a los primeros 20 trimestres consecutivos y la segunda, correspondiente a los 8 trimestres consecutivos posteriores a la ventana de entrada. Luego, los flujos son desplazados un trimestre hacia adelante, pasando a considerar como ventana de entrada del trimestre 1 al 21 y como ventana de salida del 22 al 29. Este proceso se repite hasta cubrir el largo completo de cada fondo. Estas ventanas, a medida que se van creando, se van anexando en conjuntos separados llamados “X” para las ventanas de entrada e “y” para las ventanas de salida. También, con el fin de obtener *data* limpia en los conjuntos “X” e “y”, se incorporó una función de filtrado, que define como criterio que la suma de cada ventana sea distinta de cero, de esta forma se evita agregar a los conjuntos ventanas que sean exclusivamente resultado del proceso de *padding*. La única distinción entre las funciones

creadas para cada modelo tiene que ver con el número de variables consideradas para las ventanas de entrada, ya que en el modelo 1 se consideraron 3 variables (contribuciones acumuladas, distribuciones acumuladas y NAV), en el modelo 2 se consideraron 4 variables (agregando el *vintage* a las anteriores) y para el modelo 3 se consideraron 6 (incorporando las variables “macro1” y “macro2”).

Con los datos para cada modelo en el formato correcto, se procedió a definir la estructura de las redes. Es importante recalcar que para seleccionar que tipo de capas utilizar y definir la estructura general de la red, se decidió utilizar como base la arquitectura propuesta en el artículo “*Supervised Neural Networks for Illiquid Alternative Asset Cash Flow Forecasting*” (Karatas et al., 2021), debido a que los autores plantean que las siguientes capas son óptimas para el manejo de series de tiempo y van en línea con el resultado que se busca obtener. Dicho eso, como primera estructura se utilizó una capa máscara, que tiene como fin lograr que la red ignore las entradas con valor cero. Esto se hace para que la red pueda generar predicciones en caso de que un fondo que tenga menos de 20 trimestres de historia, lo que es común al realizar el proceso de *run-off*. Luego, se agregó una capa *GRU*, que busca capturar tendencias y dependencias de largo plazo. Después, una capa de *Dropout* que sirve para prevenir el sobreajuste en el conjunto de entrenamiento. También se utilizaron capas *LSTM* que tienen la función de realizar las predicciones teniendo en cuenta todas las tendencias temporales de la misma. Por último, se optó por agregar una última capa con función de activación exponencial, para asegurarnos de obtener valores positivos para las variables, evitando errores como distribuciones con signo negativo.

7.2.4 Generación de Resultados y Gráficos

Para concluir, con cada uno de los tres modelos detallados anteriormente entrenados con sus respectivos conjuntos de entrenamiento, se generaron predicciones para el conjunto de validación. Dado que la red arroja una predicción para cada ventana de entrada recibida, se seleccionó una ventana aleatoria y posteriormente se generaron gráficos a partir de los resultados.

8. Desarrollo y Resultados

En relación con el desarrollo de escenarios para las curvas J generadas mediante el modelo de *Yale*, los resultados han sido satisfactorios, logrando cumplir con el objetivo propuesto anteriormente. Específicamente, se logró la creación de un programa junto con un informe que presenta las *J-Curves* junto a áreas coloreadas que representan diferentes desempeños históricos de una lista de fondos *vintage* seleccionados. El objetivo de este informe es evaluar la construcción y el cálculo actual de las curvas mediante la comparación con datos históricos, así como proporcionar información sobre posibles movimientos de los flujos de caja proyectados en escenarios pesimistas y

optimistas. Este ejercicio se ha llevado a cabo para el flujo de caja neto acumulado, el valor neto de los activos (*NAV*) del fondo y la Tasa Interna de Retorno (*TIR*), diferenciando por clase de activo y estrategia, respectivamente. El programa desarrollado permite generar el análisis y los escenarios para diferentes clases de activos y estrategias. Para esto, se debe tener el cuidado de utilizar una base de datos específica para cada *asset class* que se desee modelar o, en otras palabras, filtrar los datos para solo tener fondos de la misma clase. Este ejercicio, se realizó para las diferentes clases de activo sobre las cuales trabaja la empresa *sponsor* de este proyecto, sin embargo, en esta memoria solo se expondrán los resultados para *buyout*.

En las figuras 5, 6 y 7 se presentan ejemplos de lo mencionado anteriormente. En azul se muestran las métricas (flujo de caja neto acumulado, *NAV* y *TIR*, respectivamente) construidas con el modelo de Takahashi para la estrategia *buyout*. También se incluyen las áreas coloreadas, obtenidas a partir de datos de 618 fondos con un *vintage* entre 2000 y 2017. En rojo se representa un mal desempeño histórico de la curva, correspondiente al área entre el percentil 5 y el percentil 25 de este universo de fondos. Por otro lado, en verde se representa un buen desempeño histórico de los flujos de *buyout*, correspondiente al área entre el percentil 75 y el percentil 95 del universo de fondos seleccionados. Los escenarios intermedios se muestran en naranja y amarillo, correspondientes al área entre el percentil 25 y el percentil 50, así como al área entre el percentil 50 y el percentil 75, respectivamente. Es importante destacar que la Figura 7 muestra la *TIR* como una línea constante, lo que va en línea con la construcción del modelo de Takahashi, en donde se modela la tasa de crecimiento del fondo (parámetro *G*) de esta manera. Además, en esta figura se excluyen los escenarios optimistas debido al ruido excesivo en su comportamiento, lo que dificultaba la generalización. Por último, es relevante tener en cuenta que el desempeño para *buyout* modelado con Takahashi ronda un escenario intermedio.

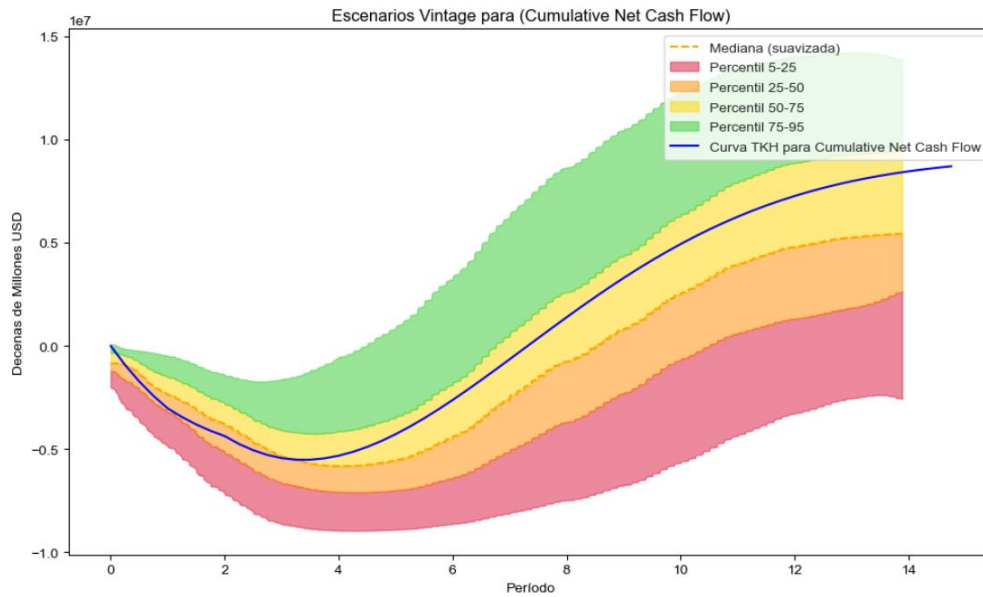


Figura 4: Análisis de escenarios para flujo de caja neto acumulado proyectado con modelo de Yale para buyout.

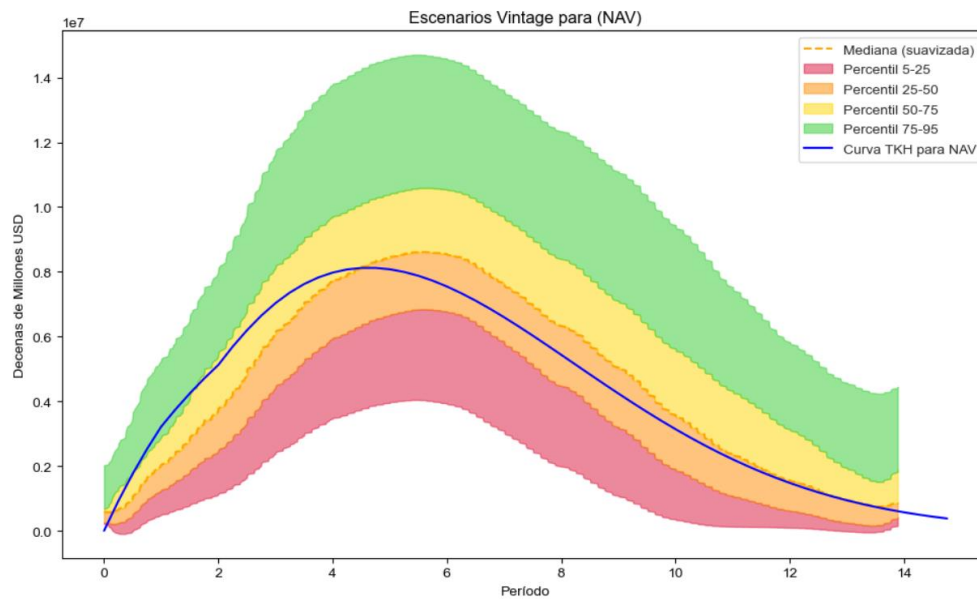


Figura 5: Análisis de escenarios para NAV proyectado con modelo de Yale para buyout.

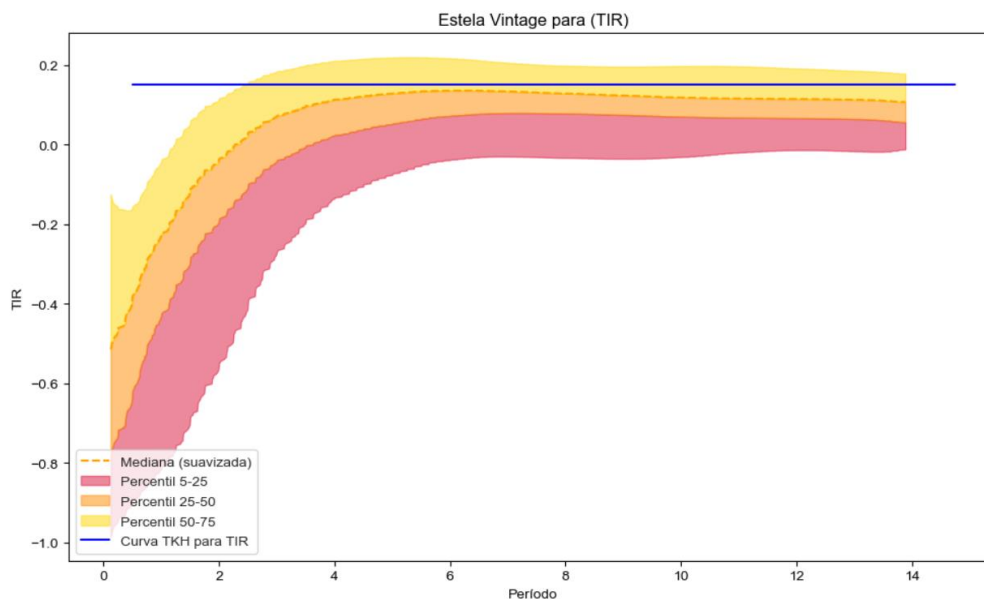


Figura 6: Análisis de escenarios para TIR proyectada con modelo de Yale para buyout.

De igual forma, para la proyección de *portfolios* con flujos ya realizados utilizando redes neuronales, el resultado fue positivo. Logrando cumplir con el objetivo propuesto al principio de este trabajo. En particular, se logró desarrollar un programa e informe en *Python*, en donde junto con preprocesar la *data* de los fondos, definir el formato y la estructura de la red, también se realizó un análisis de los resultados, evaluando la capacidad predictiva dentro de otras cosas.

Al respecto, para el primer modelo desarrollado, es decir, aquel que recibe como entrada 20 trimestres de historia sobre los flujos acumulados y el valor del fondo, los resultados fueron los siguientes. Primero, detallar que el entrenamiento de la red se realizó con 30 fondos correspondientes a fondos de *buyout* con *vintage* del 2008. Además, tras la creación de ventanas, de los 30 fondos se lograron extraer 469 ventanas de entrada y salida para el conjunto de entrenamiento y 96 ventanas de entrada y salida para el conjunto de validación. Usando los datos anteriores, al término de 150 épocas, se obtuvo un *MSE* de entrenamiento de 0,0286 y un *MSE* de validación de 0,02972 aproximadamente. Si bien estos errores son superiores a los que consiguieron llegar los autores del artículo referencia de este trabajo, no dejan de ser valores relativamente bajos. Respecto a las predicciones, se generaron diferentes gráficos en donde se selecciona aleatoriamente una ventana de entrada del conjunto de validación, y se compara la ventana de salida predicha con la ventana de salida real. La figura 7 muestra la ventana de entrada seleccionada en color azul y tanto la ventana de salida predicha como la ventana de salida real en color rojo y verde respectivamente. Este ejercicio se realizó para las tres variables de interés y en el anexo 1, se encuentra la versión más detallada para cada variable.

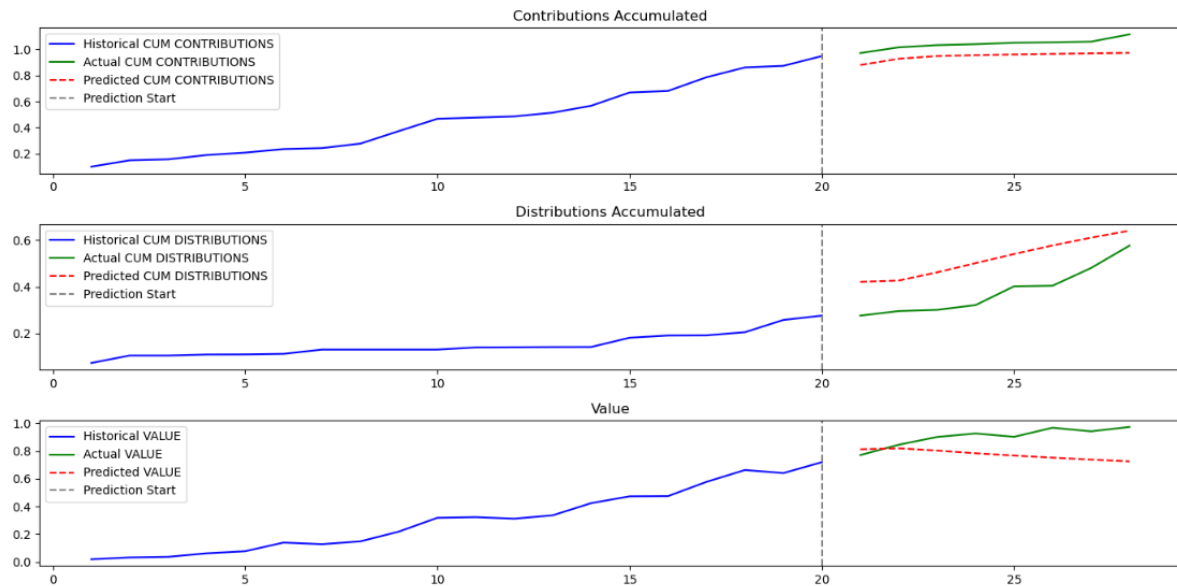


Figura 7: Resultados primer modelo para cumulative contributions, cumulative distributions y NAV.

Sobre la figura 7, el eje “Y” corresponde a la magnitud de la variable observada, que va entre 0 y 1, debido a que se está trabajando con *ratios* respecto a un nocional dado. Para el eje “X”, se tiene el tiempo en trimestres. Al respecto, se aprecia que el modelo logra capturar la tendencia tanto para las contribuciones como para las distribuciones, pero no así para el NAV del fondo. Un elemento relevante para tener en cuenta es que, se están estimando las variables por separado en contraste con el modelo de *Yale*, que relaciona mediante ecuaciones y parámetros a las tres variables. Para poder evaluar el efecto de este factor, se realizó también un gráfico que corresponde al *TVPI* calculado para los datos predichos y los datos reales. Esto, con el objetivo de lograr apreciar como interactúan entre sí las variables estimadas en relación con las reales. El resultado se puede observar en la siguiente figura.

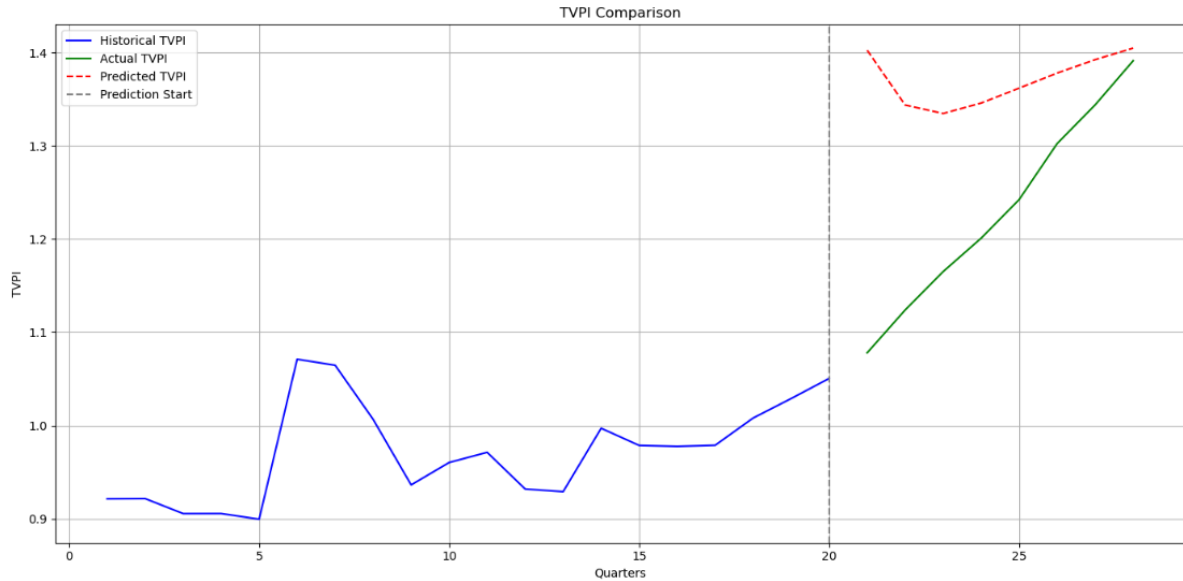


Figura 8: Resultados primer modelo para TVPI.

En la figura 8, se hace evidente el efecto de estimar las variables por separado, eliminando las relaciones existentes entre estas y generando así, una agregación o suma de los errores de cálculo para cada variable en el resultado global. Esto se aprecia fácilmente ya que, la línea roja no ajusta de ninguna manera sobre el TVPI real en verde.

De forma adicional, en el anexo 2 se muestra el desempeño del modelo ante casos de fondos con menos de 20 trimestres de historia, en particular, se fabrica un fondo que contiene valores para 10 trimestres de contribuciones acumuladas, distribuciones acumuladas y NAV. Luego, sobre este fondo experimental se realiza *padding*, agregando ceros al final de las series existentes, haciendo que el fondo quede con un largo de 20 trimestres, siendo compatible con el formato de entrada de la red. Sobre esto, se puede afirmar que la red logra estimar a partir del último dato diferente de cero, por lo que se concluye que la máscara implementada funciona de forma correcta. También, es evidente que, si a la red se le entrega como entrada menos datos, el resultado de las predicciones será peor, pese a lograr predecir.

Ahora, pasando al segundo modelo desarrollado, es importante recordar que este modelo incorpora una nueva variable de entrada que corresponde al vintage del fondo por predecir. Sobre la cantidad de datos utilizados para entrenar el modelo, para este modelo se consideró un universo de fondos de *buyout* con *vintage* entre el año 2000 y 2017, correspondiente a 413 fondos distintos. De estos fondos, se extrajeron, 3262 ventanas de entrada y salida para el conjunto de entrenamiento y 972 ventanas de entrada y salida para el conjunto de validación. Con estos datos, tras 50 épocas, el MSE obtenido en entrenamiento fue de 0,0113 y en validación fue de 0,0160. De esta forma,

el error tanto en entrenamiento como en validación fue bajo respecto al modelo anterior, posiblemente debido a que se contó con más datos para entrenar el modelo. Ahora bien, sobre las predicciones, de forma análoga al modelo anterior, en la figura 9 se observa el contraste entre la ventana de salida predicha y la ventana de salida real para una ventana de entrada seleccionada aleatoriamente dentro del conjunto de validación. Además, en el anexo 3, se encuentra el detalle del gráfico sobre las diferencias entre los valores predichos y valores reales para las *cumulative contributions*, *cumulative distributions* y *NAV*.

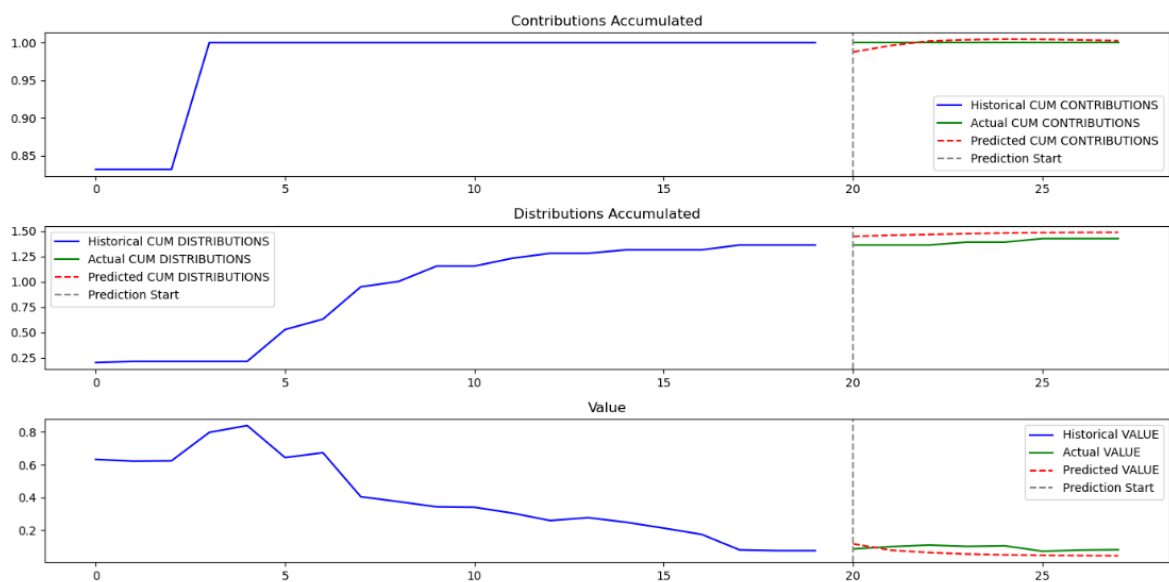


Figura 9: Resultados segundo modelo para cumulative contributions, cumulative distributions y NAV.

De la figura 9 se extraen algunas conclusiones interesantes. Para partir, con relación al modelo anterior, se aprecia un ajuste mayor de la predicción sobre los datos reales, esto para las 3 variables de interés. También, los resultados predichos tienen más sentido financiero que los del modelo anterior. Esto se puede afirmar, ya que por ejemplo las contribuciones logran llegar al valor 1, lo que se traduce en que el capital comprometido fue efectivamente llamado en su totalidad, esto tiene sentido para un fondo con 7 años de historia y es lo que se espera de un modelamiento correcto de las contribuciones de un fondo. En línea con lo anterior, también se generó un gráfico sobre el múltiplo o *TVPI* del fondo. El resultado se aprecia en la siguiente figura.

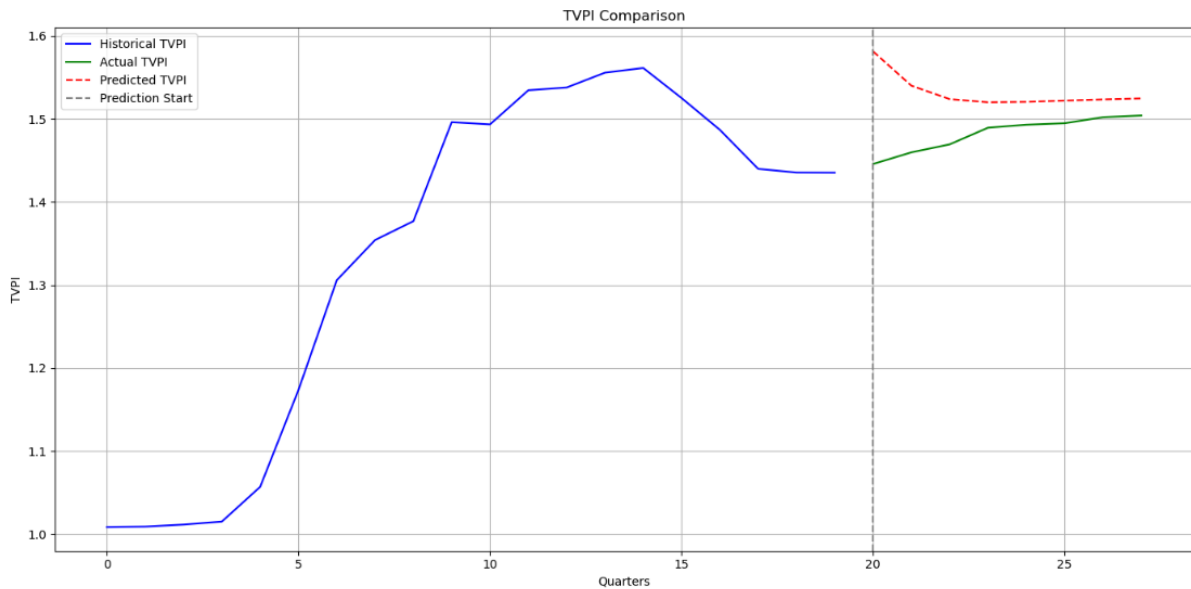


Figura 10: Resultados segundo modelo para TVPI.

En la figura 10, se aprecia una mejora respecto al modelo anterior al relacionar las tres variables estimadas al calcular el *TVPI*. Si bien la predicción no es perfecta, si logra, de forma relativa, converger hacia el mismo valor que el múltiplo real. Lo anterior se explica con una reducción en el error de predicción de las 3 variables de interés.

Por último, se repitió el ejercicio ahora para el tercer modelo, en donde a las variables consideradas por el modelo 2 se les adiciona dos variables más que tienen origen macroeconómico, que por interés del *sponsor* serán conocidas como “macro1” y “macro2”, resultando así en 6 variables de entrada. Con los mismos 413 fondos utilizados en el modelo anterior, se crearon 3262 ventanas de entrada y salida para el conjunto de entrenamiento y 972 ventanas de entrada y salida para el conjunto de validación. Con dichos datos, tras 50 épocas, se obtuvo un *MSE* en entrenamiento de 0,0086 y de 0,0146 aproximadamente para el conjunto de validación. Luego, se generaron predicciones para una ventana de entrada seleccionada aleatoriamente dentro del conjunto de validación y los resultados se graficaron junto a las predicciones realizadas por el modelo anterior para la misma ventana de entrada. Este gráfico se puede observar en la figura a continuación. El detalle de los resultados se encuentra en el anexo 4.

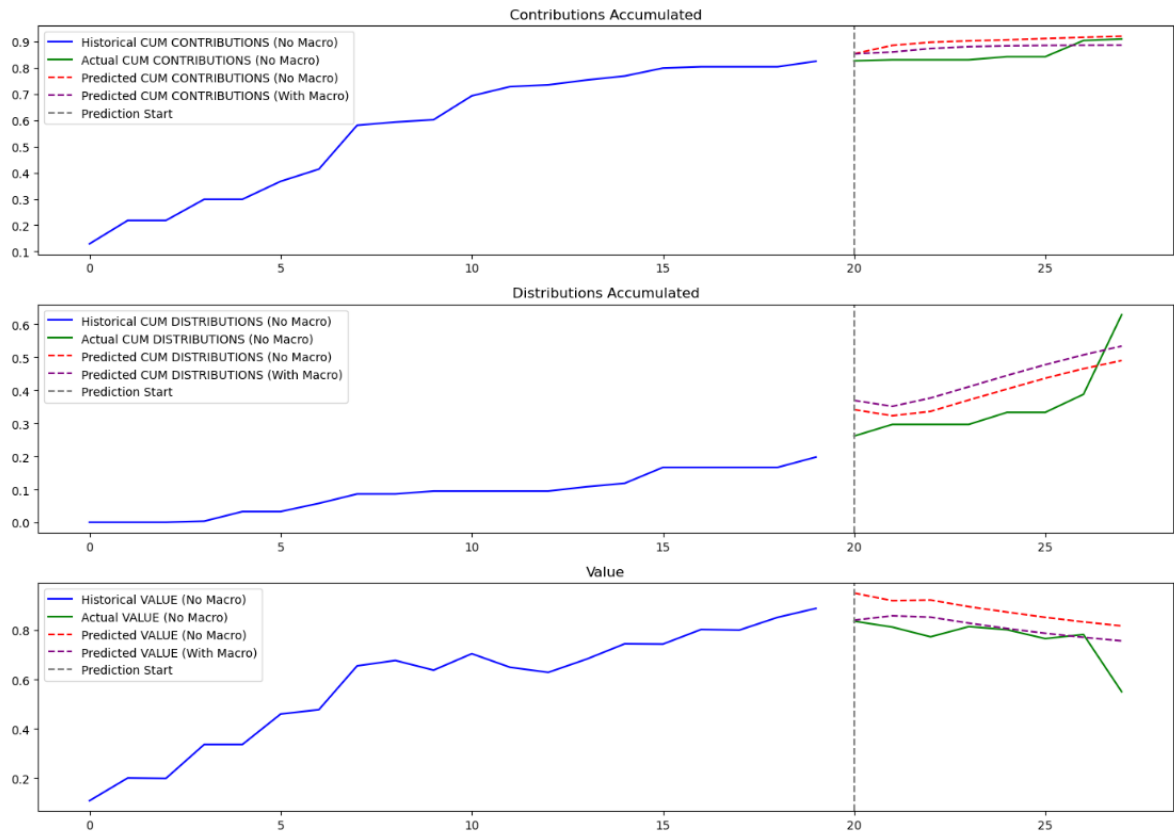


Figura 11: Resultados tercer modelo para cumulative contributions, cumulative distributions y NAV.

Respecto a la figura 11 y el tercer modelo, en línea punteada morada se distingue el modelo que incorpora variables macroeconómicas, en línea roja punteada el modelo revisado anteriormente y en línea continua verde la ventana de salida real. Los resultados no fueron los esperados, en el sentido que, al incorporar las variables macroeconómicas, se buscaba marcar una diferencia considerable en los resultados respecto al modelo 2, sin embargo, la puesta en práctica termino mostrando una leve mejora en la predicción. Esto en parte puede ser producto a la elección de las variables utilizadas o incluso el preprocesamiento realizado sobre estas, por ejemplo, que la escala utilizada para las variables difiere a las de las otras.

De forma análoga a los modelos anteriores, también se realizó el ejercicio de calcular el *TVPI* usando las variables predichas, para luego graficar el resultado. Esto se puede apreciar en la siguiente figura.

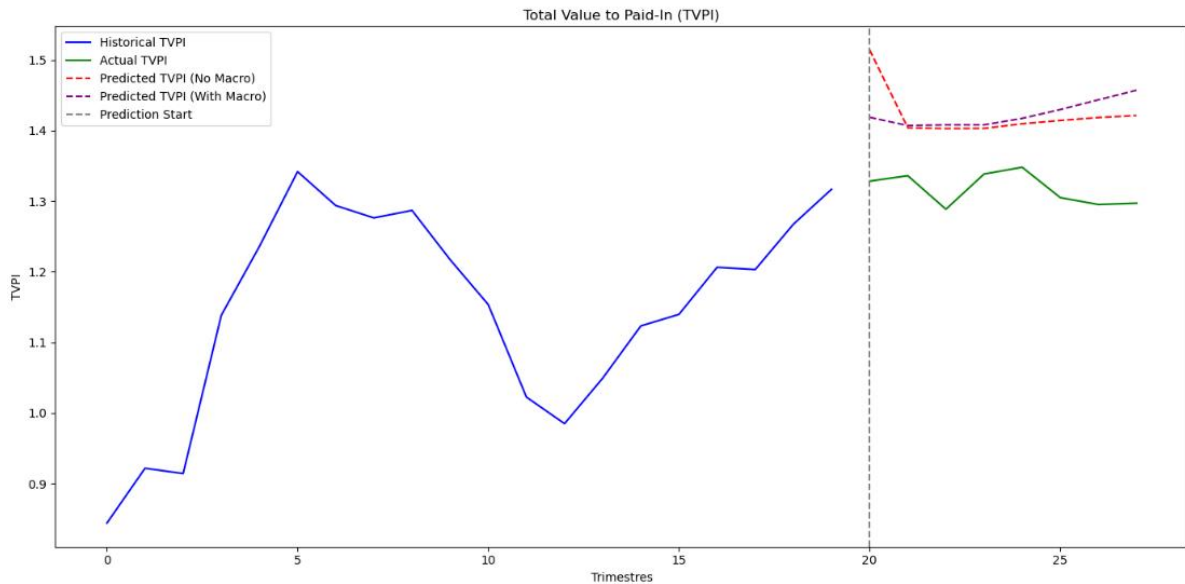


Figura 12: Resultados primer modelo para TVPI.

En la figura 12, es evidente que tanto el segundo modelo como el tercer modelo, con sus predicciones, no logran ajustar de buena forma respecto al múltiplo real. Si se comparan ambos modelos, se puede afirmar que la inclusión de variables macroeconómicas suaviza los resultados, alineándolos con el sentido financiero buscado y reduciendo el ruido en la predicción, por ejemplo, eliminando picos abruptos en los resultados.

A modo de resumir las características fundamentales de cada modelo y facilitar la comparación de los resultados de estos, se elaboró una tabla comparativa que detalla para cada modelo, las variables de entrada, salida, cantidad de datos utilizados para el entrenamiento y el error de validación entre otras cosas. Dicho cuadro se encuentra a continuación.

Característica	Modelo 1	Modelo 2	Modelo 3
Variables de entrada	Contribuciones acumuladas, distribuciones acumuladas y NAV (3)	Contribuciones acumuladas, distribuciones acumuladas, NAV y <i>vintage</i> (4)	Contribuciones acumuladas, distribuciones acumuladas, NAV y <i>vintage</i> , macro1 y macro2 (6)
Variables de salida	Contribuciones acumuladas, distribuciones acumuladas y NAV (3)	Contribuciones acumuladas, distribuciones acumuladas y NAV (3)	Contribuciones acumuladas, distribuciones acumuladas y NAV (3)
Número de fondos utilizados	30 fondos (todos <i>vintage</i> 2008)	413 fondos (<i>vintage</i> entre 2000 y 2017)	413 fondos (<i>vintage</i> entre 2000 y 2017)
Cantidad de ventanas para entrenamiento	469	3262	3262
Cantidad de ventanas para validación	96	972	972
Error de entrenamiento	0,0286	0,0113	0,0086
Error de validación	0,02972	0,0160	0,0146

Tabla 1: Comparación de los modelos tipo red neuronal y sus características.

Recapitulando, en cuanto a la primera fase del proyecto, podemos afirmar que esta se encuentra en línea con los objetivos planteados al inicio del semestre. Se logró generar una comparación del resultado que arroja el modelo de Takahashi con datos históricos, lo que resulta útil para analizar las clases de activos y evaluar o identificar, de forma visual y rápida, posibles riesgos asociados a malos periodos de inversión. Tras conversar con la contraparte, se identificaron algunas mejoras en los filtros utilizados para seleccionar los fondos, relajando algunas de las restricciones actuales. Además, se sugirió incorporar al análisis una mejor representación de las formas que tienen las curvas de los percentiles bajos en comparación con los altos. Por ejemplo, un fondo en el percentil 95 debería distribuir más rápido que uno en un percentil bajo, resultando en una curva J más corta. En esta misma línea, un buen fondo debiese tener una curva J más alta que un mal fondo, esto debido a que sus distribuciones son más grandes. Estas diferencias en las formas de las curvas no son completamente capturadas por la versión actual del programa desarrollado.

Por otro lado, la segunda fase del proyecto también cumplió con los objetivos, ya que, con este ejercicio se buscaba explorar alternativas al proceso de *run-off* que actualmente se realiza usando el modelo de *Yale* y se logró generar un programa capaz de realizar predicciones con un error relativamente bajo (*MSE* en torno a 0,02) y a una velocidad considerablemente superior al modelo que actualmente se utiliza para realizar las proyecciones. Al respecto, el modelo que mejor desempeño fue el tercero, que incluye variables macroeconómicas aparte de los flujos realizados. Sin embargo, no se logró llegar al nivel de precisión que obtuvieron los autores del artículo “*Supervised Neural Netowrks for Illiquid Alternative Asset Cash Flow Forecasting*” (Karatat et al., 2021).

Por último, a modo de contrastar la alternativa explorada con la utilizada actualmente, se realizó el ejercicio de escoger una ventana de entrada de forma aleatoria dentro del conjunto de validación y luego a partir de esta se realizaron predicciones con el tercer modelo y con el modelo de Takahashi. Para realizar las proyecciones utilizando el modelo de *Yale*, se usaron los parámetros que la empresa usa para modelar la curva de *buyout*. Los resultados se muestran a continuación.

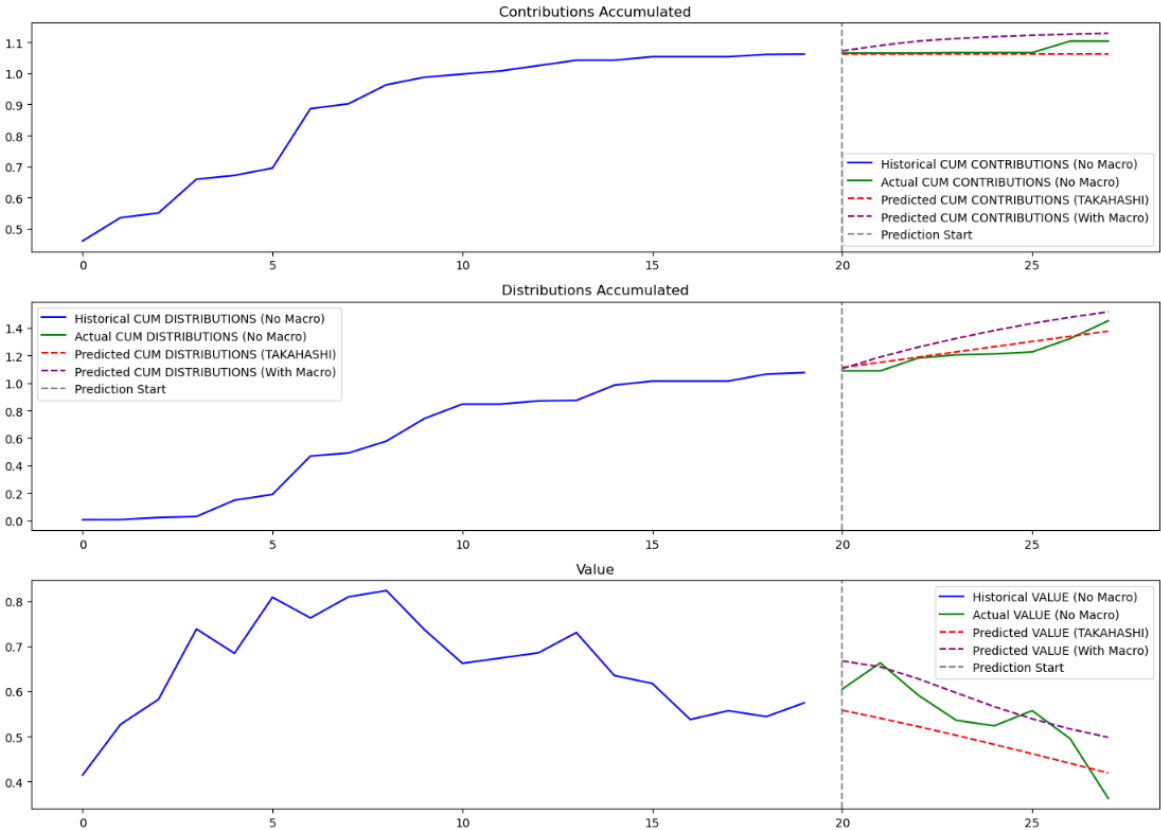


Figura 13: Comparación modelo 3 con Takahashi.

Sobre la figura 13, se puede afirmar que ambas alternativas muestran resultados similares y en general un buen desempeño. Al respecto, se podría afirmar que ambas alternativas presentan poder predictivo en el corto plazo. Una particularidad que se alcanza a apreciar gráficamente es que, mientras Takahashi estima que para los 8 trimestres predichos no se llamara más capital, la red neuronal sí estima un crecimiento en las contribuciones y, por ende, logra ajustar y capturar el comportamiento real de dicha variable. También, se puede afirmar que la red neuronal logra seguir la tendencia de los flujos reales, llegando en instancias a predecir de buena forma el valor final o la cola de la serie, mientras que el modelo de *Yale* logra mayor ajuste en los valores intermedios de la serie.

9. Discusiones

Para abarcar las discusiones en este informe, se realizará un repaso por las dos etapas principales del proyecto, levantando oportunidades de mejora y tópicos por discutir según corresponda. Partiendo por los escenarios desarrollados para las *J-curves* que modelan los diferentes *asset classes*, el *sponsor* buscó con este desarrollo generar una herramienta que permita a los clientes entender de forma sencilla a que se podrían exponer al invertir en los diferentes tipos de activos alternativos, sin embargo, florecen las siguientes dudas. ¿Es la generación de trayectorias gráficas la mejor forma de explicar de forma sencilla a los clientes el riesgo asociado a invertir en cierto *asset class*? En esta misma línea, ¿Son estos gráficos del interés de un cliente que invierte en este tipo de activos? Al respecto, se puede afirmar que los activos alternativos tienen una naturaleza poco líquida, lo que hace que una correcta evaluación de riesgos sea una necesidad, siendo de utilidad la estadística y datos numéricos para este propósito. Además, se sabe que los escenarios generados en este proyecto no entregan montos cuantificables de pérdidas potenciales. Tampoco aportan probabilidades de ocurrencia para los distintos escenarios. Por lo que, si bien son una herramienta visual que es fácil de comprender, ya que no requiere de conocimientos técnicos, el impacto y valor que entrega a los clientes es limitado. Un posible complemento al desarrollo realizado podría haber sido adaptar el cálculo de métricas como el VaR^{xxv} a los activos alternativos, permitiendo así al inversor conocer el monto o porcentaje de pérdida de una inversión dado cierto nivel de significancia. Métricas de este estilo, agregarían valor a los escenarios desarrollados, sin dejar de lado una interpretación directa y clara.

Otro argumento detrás de la elaboración de los escenarios para las *J-curves* tiene que ver con buscar explicar de forma sencilla los efectos asociados directamente al *vintage* en el que se invierte. Plasmar estos “efectos” es particularmente complicado, ya que no existe forma de conocer que *vintages* serán buenos o malos, aparte de ser poco cuantificable. Sobre este punto, se puede decir que las trayectorias desarrolladas en este trabajo apuntan a generar valor mediante datos históricos de buenos y malos *vintages*,

sin embargo, se podría buscar mayor interpretabilidad en los resultados generados. Ya que, más allá de la explicación gráfica, la información numérica y conclusiones que aportan a los clientes la solución en su estado actual es mejorable. Una buena opción por explorar hubiese sido buscar ajustar a la mediana o al promedio de los fondos de un mal *vintage* nuevas *J-Curves* construidas con Takahashi. De esta forma, mediante el ajuste y selección de los diferentes parámetros del modelo, se puede construir una curva pesimista, pero con conocimiento de los supuestos detrás de dicho escenario. En otras palabras, al conocer los valores que toman los parámetros de la curva pesimista, resulta directo generar un caso de estudio explicado para los clientes, un proceso análogo a un análisis de sensibilidad de evaluación de proyectos, en que se explican los supuestos que se consideraron para crear el peor caso. Al realizar lo anterior, es fácil lograr que los clientes entiendan bajo que supuestos y valores numéricos de los parámetros, los flujos podrían asimilarse a flujos que históricamente fueron malos, como en el caso de una crisis.

Surgen también preguntas como: ¿qué otras variables podrían haber sido factores relevantes en la creación y determinación de escenarios? Aunque las posibilidades son numerosas, algunas de las más interesantes para explorar incluyen el *manager* y el tamaño del fondo. Dado que el rendimiento de los fondos de inversión en el mercado privado depende en gran medida de la estrategia de inversión única de cada firma, este mercado es más heterogéneo que su contraparte pública, donde muchas de las propuestas de inversión se centran en replicar un *benchmark* o índice. Por lo tanto, sería interesante utilizar datos históricos y la estadística para clasificar a los diferentes *managers* de estos fondos y luego construir escenarios que modelen las distintas clases de activos a lo largo de su ciclo de vida, considerando sus fluctuaciones dependiendo del tipo de *manager* o *GP*. De manera similar, un análisis enfocado en la variable del tamaño del fondo podría revelar *insights* valiosos, ya que el tamaño suele estar directamente relacionado con la estrategia de inversión del *manager*, afectando significativamente la naturaleza de los flujos de caja. El uso de estas y otras variables sería de gran interés tanto para las empresas del sector como para los clientes, que buscan constantemente diversificar y protegerse ante la mayor cantidad de riesgos posibles.

De igual manera, no se puede pasar por alto los filtros y el tratamiento realizado sobre los datos para elaborar los escenarios donde, por ejemplo, se eliminaron los *outliers* presentes en la base de datos. Sin embargo, dado que el objetivo de esta primera parte del proyecto consistió en modelar y mostrar de forma gráfica las posibles fluctuaciones en los flujos de fondos de inversión, quizá hubiese sido interesante buscar aquellos datos que se desviaron de la media, ya que muchas veces son esos casos extremos los que realmente les interesan a los inversores. En esa misma línea, la herramienta desarrollada se pensó como una forma de controlar los riesgos asociados a inversiones en fondos de activos alternativos, pero también, tiene potencial para usos enfocados en posibles ganancias, tornándose así en una herramienta comercial interesante para explorar.

Pasando a la segunda fase de este proyecto, el *sponsor* se mostró interesado en buscar una alternativa computacional al *run-off* actual debido a que, este es un proceso lento por su naturaleza iterativa y demanda de poder computacional. Ante estos requerimientos, se podría afirmar que hubiese sido mejor opción idear mejoras para el modelo actual en vez de explorar opciones para un desarrollo completamente nuevo. Ya que, construir un modelo desde cero, que sea innovador y que logre entregar resultados de forma precisa, todo esto demorándose menos tiempo que el modelo actual, es algo sumamente complejo, que requiere de un equipo de desarrollo más grande y más conocimientos técnicos en materia de modelado financiero y computación. Todo esto es algo que termina quedando grande a un practicante que trabaja en modalidad *part-time* que de por sí tiene que invertir tiempo extra en conocer el negocio de la empresa, familiarizarse con las herramientas que utilizan y aclimatarse a la nueva etapa, haciendo que los 6 meses del proyecto se reduzcan aún más. Al respecto, para mejorar los tiempos del *run-off* actual, se podría haber indagado en una forma de utilizar metodologías de vanguardia sobre el modelo de *Yale*, como podría ser la programación en paralelo para llevar a cabo cálculos independientes de forma simultánea. Esto, sumado a una revisión y optimización del código, quizá hubiesen permitido hacer más eficiente el proceso de cálculo que se utiliza para el *run-off* y haberse logrado satisfactoriamente a lo largo del semestre por el alumno en práctica.

Bajo esta misma línea, podemos destacar la motivación de la compañía patrocinadora de este proyecto en invertir tiempo buscando desarrollos y metodologías innovadoras en la literatura académica. Ya que ante una industria que constantemente está sometida a cambios y el hecho de que la empresa busca diferenciarse de su competencia, es relevante desafiar el *statu quo* y visitar los modelos utilizados, planteando mejoras y cambios. Dicho lo anterior, se pueden afirmar dos cosas. Por un lado, se demostró con el trabajo realizado durante este semestre, que modelos computacionales, como las redes neuronales, tienen cabida en la industria financiera, en particular para los activos alternativos. Por lo que, ante la búsqueda de generar herramientas más complejas y que sean útiles para las empresas, sería interesante indagar en como generalizar el poder predictivo a corto plazo que demostró la red neuronal programada, a un poder predictivo de largo plazo. Es relevante entender que dicha exploración no se abarcó en esta memoria principalmente debido a temas de tiempo, ya que como se mencionó en un punto anterior, un desarrollo grande requiere de más recursos de los que se dispuso. Sin embargo, se alcanzaron resultados prometedores que en un futuro podrían servir como base para un segundo proyecto. En particular, sería interesante explorar la utilización de redes neuronales como las utilizadas en este trabajo para modelar etapas específicas de la vida de los fondos de inversión. En otras palabras, dado que un fondo de este tipo tiende a tener una duración de unos 15 años, se podría modelar cada fondo en dos fases, una de crecimiento y otra de madurez. Luego, generar redes neuronales análogas a las discutidas anteriormente específicas a

cada etapa. De esta forma, se podría lograr realizar un *run-off* para la vida completa de un fondo, independiente de cuanta historia se tenga de este.

Por otro lado, tras la investigación realizada a inicios de semestre, es claro que existe interés y áreas de investigación respecto al modelamiento de los activos alternativos. Por lo que, sería interesante indagar formas en la en las que se puedan modelar los activos alternativos que no se basen en el modelo de *Yale*. Por ejemplo, alternativas estocásticas o el uso de modelos computacionales que logren capturar tendencias sobre *data* histórica sobre fondos de inversión, idealmente buscando obtener mejores resultados o más grados de libertad que el conocido modelo de Takahashi.

Por último, es importante hablar sobre la metodología que se siguió para la segunda etapa del proyecto. Ya que como se mencionó anteriormente, el trabajo se realizó utilizando como base teórica el artículo “*Supervised Neural Networks for Illiquid Alternative Asset Cash Flow Forecasting*” (Karatat et al., 2021). En dicho *paper*, los autores abarcan dos metodologías para la predicción de flujos. En primer lugar, se propone un modelo que reciba series de tiempo para las contribuciones, distribuciones y NAV de los distintos fondos, para retornar una serie de parámetros de Takahashi que logren modelar los flujos futuros a dos años. A este método, los autores le llaman modelo de predicción indirecta. Por otro lado, en el artículo también se plantea un modelo de predicción directa, que recibe series de tiempo para los flujos y NAV del fondo, para retornar predicciones a dos años para estas mismas variables. En este trabajo, se optó por seguir la segunda opción, dado que la implementación requería menos pasos previos que el modelo de predicción indirecta, como, por ejemplo, calcular para cada fondo y cada periodo los parámetros de Takahashi. Sin embargo, dado que los desarrollos que tiene la empresa fueron construidos sobre la base teórica del modelo de *Yale*, haber experimentado con el modelo de predicción indirecta hubiese sido sumamente valioso para el *sponsor*, ya que el desarrollo hubiese sido compatible con el ecosistema actual, minimizando el tiempo y los recursos invertidos en relacionar las diferentes herramientas que dispone la compañía. Por lo que, queda propuesto el modelo de predicción indirecta para investigaciones o proyectos futuros.

10. Conclusiones

En conclusión, se considera el trabajo realizado como un caso exitoso pero mejorable. Partiendo por la generación de escenarios, en donde se logró cumplir con el objetivo propuesto, generando una herramienta en *Python* que complementa el modelado de las *J-curves* asociadas a las diferentes clases de activos. Esto aportó a la empresa una herramienta dirigida al equipo, permitiendo a los analistas hacer una inspección gráfica de las curvas construidas con el modelo de *Yale* y compararla con *data* histórica de forma sencilla y rápida. También, esta herramienta es particularmente útil para los

clientes de la compañía. En efecto, dentro de los clientes de la empresa se encuentran *family offices* o clientes particulares que no tienen el nivel de tecnicismo de un inversor institucional, pero aun así demandan información respecto a los riesgos asociados a los activos alternativos. Es por esto por lo que, los escenarios son una forma sencilla de explicar y describir como se espera que se comporten los flujos de los diferentes *asset class*.

Sin embargo, la herramienta se puede mejorar, siempre apuntando a que sea fácil de entender y aporte información concreta y relevante para los inversores. Una forma de lograr esto, sería mediante la incorporación de métricas estándar para cuantificar el riesgo de inversión, como el *VaR*, que aporta a los inversores una idea del monto que se podría perder dado cierto nivel de significancia. Otra forma de lograr el fin buscado sería mediante un análisis de sensibilidad para los flujos esperados de las diferentes clases de activos. En particular, este se podría construir modificando los parámetros de la curva de Takahashi, a modo de ajustarla al desempeño histórico de un mal *vintage*. Luego, el significado de dichos parámetros y sus valores podrían ser presentados a los inversores a modo de explicar que se podría esperar de un escenario pesimista. Es relevante también, pensar en que otras variables podrían utilizar para construir y modelar escenarios, ya que existe evidencia del efecto y relevancia que cobra el *manager* y tamaño de un fondo en su desempeño, lo que abre un abanico de soluciones por explorar siempre pensando en agregar valor a los clientes y a la empresa.

Respecto al *run-off* con redes neuronales, al igual que la otra etapa, se cumplió con el objetivo propuesto al inicio de este proyecto, generando un modelo de tipo red neuronal en *Python* que es capaz de realizar predicciones de los flujos de caja acumulados y el valor de un fondo a dos años utilizando cinco años de historia, todo con una velocidad de cálculo superior al modelo de *run-off* actual, siendo capaz de entregar resultados prácticamente instantáneamente y con errores de predicción bajos (*MSE* inferior a 0.03). De esta forma, se entrega a la empresa un trabajo investigativo y un análisis completo sobre alternativas sofisticadas al modelo de Takahashi para tareas predictivas. La herramienta desarrollada servirá como segunda opinión al realizar *run-off* a mediano y corto plazo, complementando los resultados arrojados por el modelo de *Yale*. También, con el trabajo realizado la compañía obtiene su primer desarrollo en materia de redes neuronales, que puede servir como punto de partida para desarrollos futuros y que logra potenciar a la empresa desde el punto de vista comercial, al dar indicios de inversión en modelos computacionales de vanguardia.

Sin embargo, tal como se detalló en los alcances, la métrica de éxito para esta etapa era presentar los avances y resultados en un comité de desarrollo y recibir la aprobación para su uso de parte del equipo. En dicho comité, se valoró el trabajo realizado como un excelente *research*, pero no como una herramienta lista para ser usada. Esto principalmente por las restricciones del modelo respecto a los datos que

recibe de entrada y al hecho de solo poder predecir una ventana dos años de largo como máximo. Con los antecedentes anteriores, se concluye que, para el patrocinador de este proyecto, el modelo de Takahashi sigue siendo la mejor alternativa para realizar predicciones de flujos para *portfolios* de activos alternativos con historia. Ya que, este modelo por su naturaleza iterativa permite proyectar independiente de la etapa en la que se encuentre la vida del fondo. También, es sumamente flexible y configurable mediante el ajuste de los parámetros, función que no está presente en un modelo como el explorado en este trabajo. Por último, todo lo anterior lo realiza con un buen ajuste a los datos, con errores de predicción similares al de un modelo computacional avanzado como es una red neuronal, tal como se demostró en la figura 13.

Dicho lo anterior, no se puede omitir el potencial que tienen las redes neuronales en materia de predicción de flujos. Ya que, cuentan con la capacidad de capturar tendencias a largo plazo de forma sencilla, además de ser capaces de modelar relaciones complejas y poco evidentes entre variables. Siendo capaces de obtener información que los modelos actuales no consideran, por ejemplo, capturando relaciones de datos macroeconómicos o datos del mercado público, entre otras cosas. Por lo que, con el objetivo de potenciar la herramienta predictiva construida en este trabajo, se propone modelar la vida de un fondo en 2 etapas y entrenar un modelo específico para cada una. De dicha forma, se logra sacar provecho a las capacidades predictivas a corto plazo de las redes neuronales y también, se logra una usabilidad real al cubrir la vida completa de un fondo.

Por último, es importante tener en consideración que el proyecto realizado se logró exitosamente en el tiempo disponible durante la práctica en modalidad media jornada, aportando a la empresa con herramientas y una investigación respecto a nuevas metodologías. En esa misma línea, el estudiante logro sacar provecho a dicha instancia, aprendiendo e interesándose en los activos alternativos. También hay que mencionar que las propuestas no lograron ser abarcadas en esta instancia por tiempo y conocimientos que en el momento dado no se tenían.

11. Bibliografía

AFP Habitat (s.f). ¿Qué son los MULTIFONDOS?

<https://www.afphabitat.cl/multifondos/#:~:text=%C2%BFQu%C3%A9%20son%20los%20Multifondos%3F,para%20cada%20uno%20de%20ellos>.

BlackRock (s.f.). *Outsourced Chief Investment Officer* [OCIO].

<https://www.blackrock.com/institutions/en-us/solutions/outsourced-chief-investment-officer-ocio#:~:text=Outsourced%20Chief%20Investment%20Officer%2C%20or,a%20third%20party%20investment%20partner>.

Carrasco, M. (2022). *Análisis de múltiples financieros del portafolio de activos alternativos de una AFP versus el mercado público* [Título profesional, Universidad de Chile] Repositorio Académico de la Universidad de Chile.

<https://repositorio.uchile.cl/>

CFA Institute. (s.f.) (1). *Introduction to Alternative Investments*.

<https://www.cfainstitute.org/en/membership/professional-development/refresher-readings/introduction-alternative-investments>

CFA Institute. (s.f.) (2). *Private Equity Investments*.

<https://www.cfainstitute.org/en/membership/professional-development/refresher-readings/private-equity-investments>

Corporate Finance Institute. (s.f) *Value at Risk (VaR)*

<https://corporatefinanceinstitute.com/resources/career-map/sell-side/risk-management/value-at-risk-var/>

Chambers, D., Anson, M., Black, K., & Kazemi, H. (2015). Chapter 1. What Is an Alternative Investment. *Alternative Investments* (pp. 51-61).

https://caia.org/sites/default/files/caia_level_i_chapter_1.pdf

Chen, J. (20 de junio de 2022). *Agency Problem: Definition, Examples, and Ways To Minimize Risks*. Investopedia.

<https://www.investopedia.com/terms/a/agencyproblem.asp#:~:text=An%20agency%20problem%20is%20a,management%20and%20the%20company's%20stockholders>.

Chladek, N. (7 de mayo de 2020). *7 Types of alternative investments everyone should know*. Harvard Business School Online.

<https://online.hbs.edu/blog/post/types-of-alternative-investments>

Coursera. (10 de Abril de 2024). *What Is an LSTM Neural Network*.

<https://www.coursera.org/articles/lstm-neural-network>

Collazos, C. (23 de abril de 2023). *Rolling Window Method for Time Series Forecasting using ML*. LinkedIn.

<https://www.linkedin.com/pulse/rolling-window-method-time-series-forecasting-using-cristian-camilo/>

Finance Strategists (s.f.). *Natural resources investments*. Recuperado el 29 de marzo de 2024.

<https://www.financestrategists.com/wealth-management/alternative-investment/natural-resources-investments/#:~:text=Natural%20resources%20investments%20refer%20to,appreciation%2C%20dividends%2C%20or%20royalties>.

GeeksforGeeks. (6 de Mayo de 2023). *Data Preprocessing in Data Mining*.

<https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>

Hardesty, L. (14 de abril de 2017). Explained: Neural networks. *MIT News*.

<https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>

Huet, P. (13 de abril de 2023). *Que son las redes neuronales y sus aplicaciones*. *OpenWebinars*.

<https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/>

IBM (s.f). *What is exploratory data analysis (EDA)?*

<https://www.ibm.com/topics/exploratory-data-analysis>

ILPA (s.f). *Unfunded Commitment*. Recuperado el 24 de abril de 2024, de

<https://ilpa.org/glossary/unfunded-commitment/>

Infante, M. (15 de abril de 2024). Banco Central eleva límites de inversión de AFP en activos alternativos e industria prevé mayor rentabilidad de los fondos. *Diario Financiero*.

<https://www.df.cl/mercados/pensiones/banco-central-decide-elevar-los-limites-de-inversion-en-activos>

Karatas, T., Klinkert, F., & Hirsra, A. (2021). Supervised Neural Networks for Illiquid Alternative Asset Cash Flow Forecasting.

<https://arxiv.org/abs/2108.02853>

Lehman, R (1). (5 de diciembre de 2023). *Internal Rate of Return*. *Moonfare*. Recuperado el 27 de marzo de 2024.

<https://www.moonfare.com/glossary/internal-rate-of-return-irr>

Lehman, R (2). (5 de diciembre de 2023). *Total Value to Paid-In Capital*. *Moonfare*. Recuperado el 27 de marzo de 2024.

[https://www.moonfare.com/glossary/total-value-to-paid-in-capital-tvpi#:~:text=Total%20Value%20to%20Paid%20In%20Capital%20\(also%20known%20as%20the,into%20the%20fund%20to%20date](https://www.moonfare.com/glossary/total-value-to-paid-in-capital-tvpi#:~:text=Total%20Value%20to%20Paid%20In%20Capital%20(also%20known%20as%20the,into%20the%20fund%20to%20date)

[https://www.moonfare.com/glossary/total-value-to-paid-in-capital-tvpi#:~:text=Total%20Value%20to%20Paid%20In%20Capital%20\(also%20known%20as%20the,into%20the%20fund%20to%20date](https://www.moonfare.com/glossary/total-value-to-paid-in-capital-tvpi#:~:text=Total%20Value%20to%20Paid%20In%20Capital%20(also%20known%20as%20the,into%20the%20fund%20to%20date)

Peiro, A. (30 de septiembre de 2020). *Deuda con dificultades - Distressed debt*. Recuperado el 29 de marzo de 2024.

<https://economipedia.com/definiciones/distressed-debt-deuda-con-dificultades.html>

Prequin. (s.f.). *What Are Alternative Assets?*

<https://www.prequin.com/academy/lesson-1-alternative-assets/what-are-alternative-assets>

Rakaworski, A. (23 de septiembre de 2022). *Infrastructure as an alternative*. *IPE*.

<https://www.ipe.com/infrastructure-as-an-alternative/15611.article>

- Sanjay, A. (13 de marzo de 2023). Padding in Neural Networks: Why and How? *Medium*.
<https://blog.gopenai.com/padding-in-neural-networks-why-and-how-b076ab0a4fc2>
- Takahashi, D. & Alexander, S. (2001). *Illiquid Alternative Asset Fund Modeling*.
<https://doi.org/10.3905/jpm.2002.319836>
- Tensorflow (s.f). *Understanding masking & padding*.
https://www.tensorflow.org/guide/keras/understanding_masking_and_padding
- ULL, B. (2024). *Propuesta de modelo de negocios de Santiago Advisors como administradora de fondos con gestión de portafolio automatizado* [Título profesional, Universidad de Chile] Repositorio Académico de la Universidad de Chile.
<https://repositorio.uchile.cl/>
- Wegener, K. (2020). *Desarrollo de un método de construcción de portafolio de activos alternativos para una administradora de fondos de pensión* [Título profesional, Universidad de Chile] Repositorio Académico de la Universidad de Chile.
<https://repositorio.uchile.cl/>

Anexos

Anexo A: Detalle Contribuciones, Distribuciones y NAV para el Primer Modelo

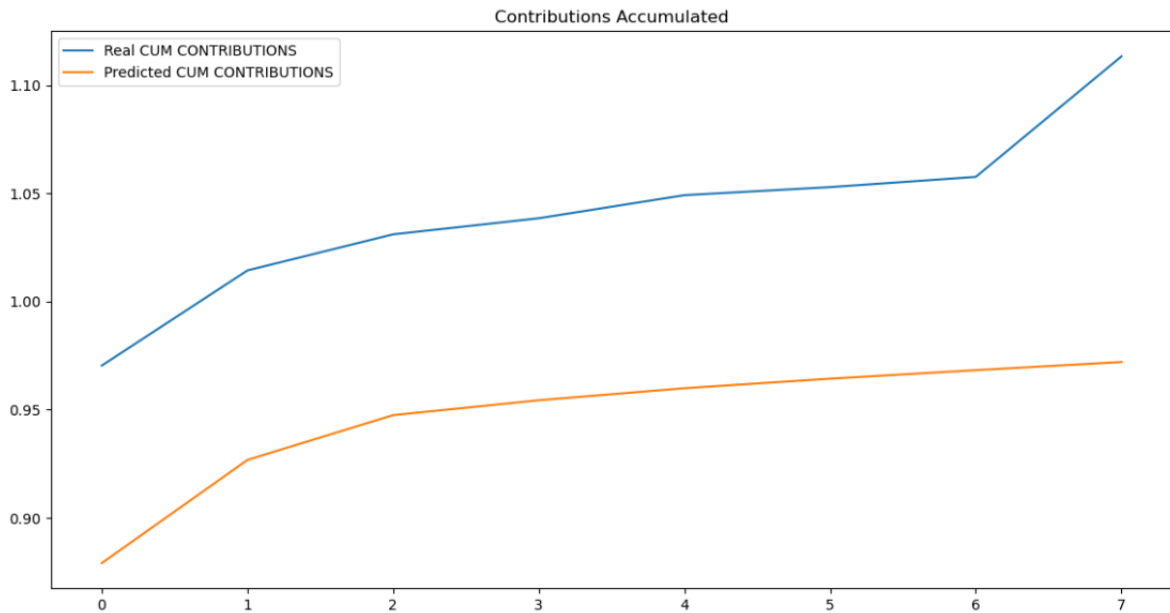


Figura 14: Contribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 1).

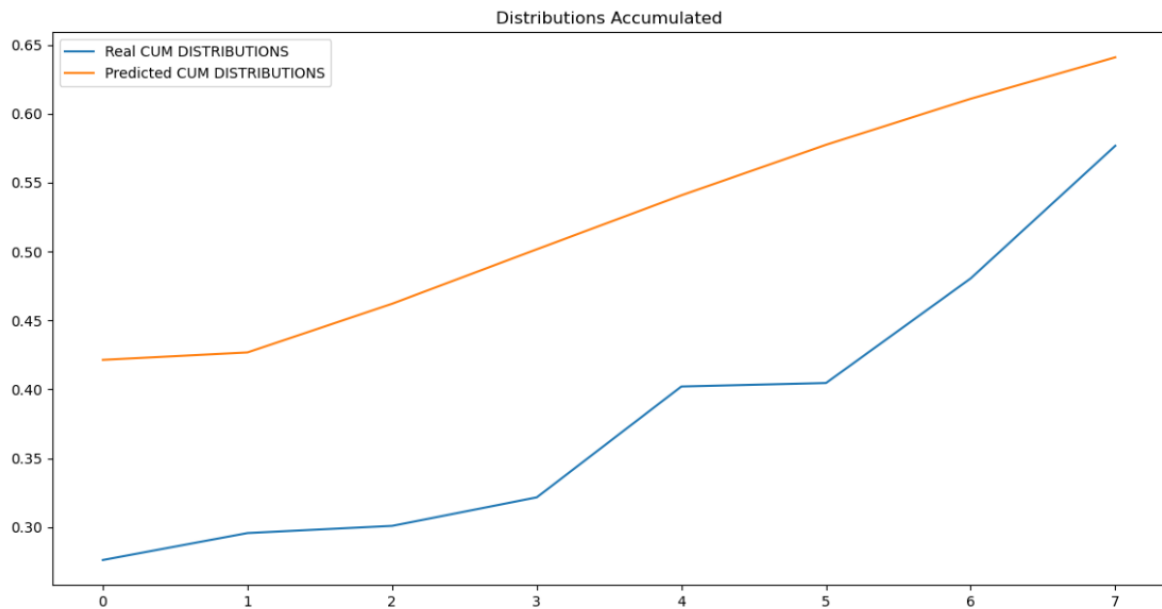


Figura 15: Distribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 1).

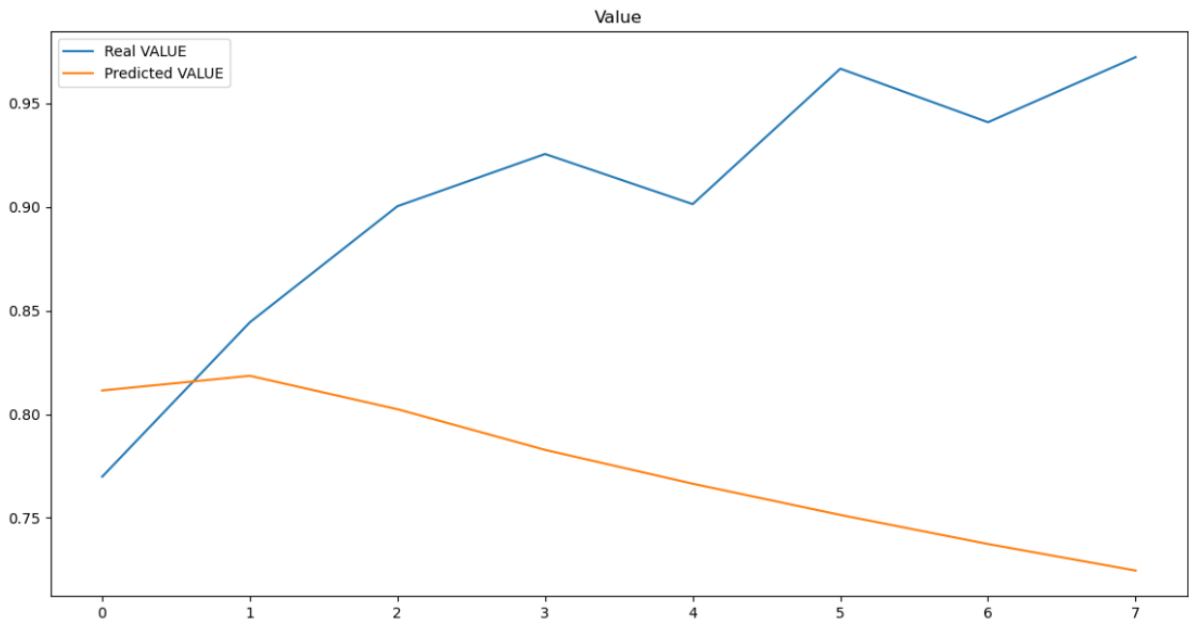


Figura 16: NAV predicho para 8 trimestres a partir de 20 trimestres de historia (modelo 1).

Anexo B: Ejercicio de Fondo con 10 Trimestres de Historia para el Primer Modelo

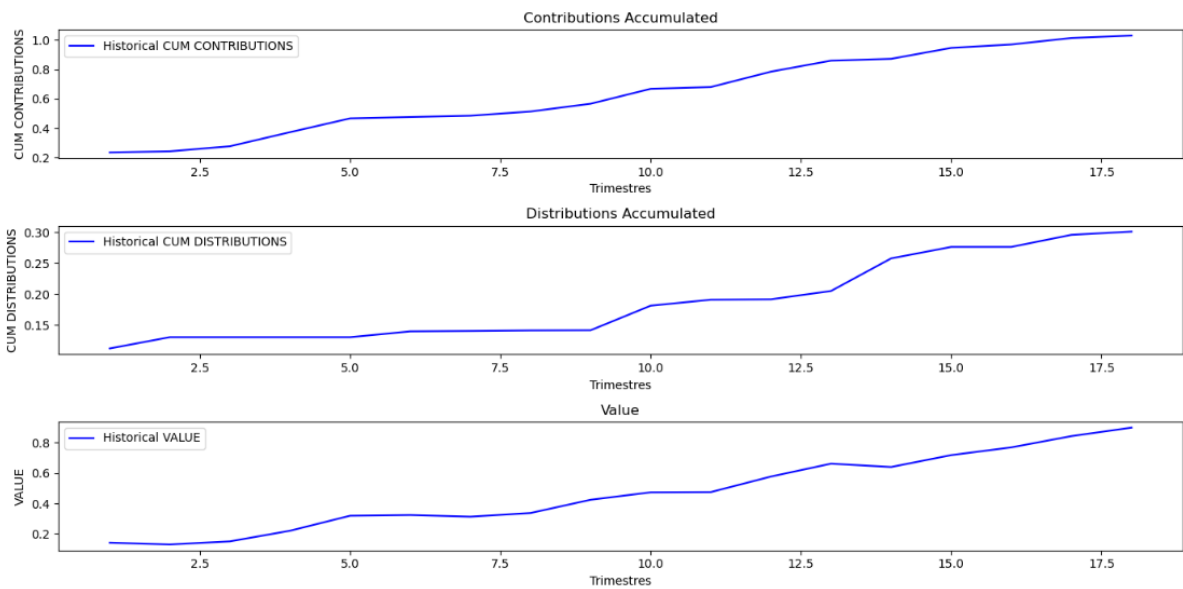


Figura 17: Desempeño real de un fondo (18 trimestres).

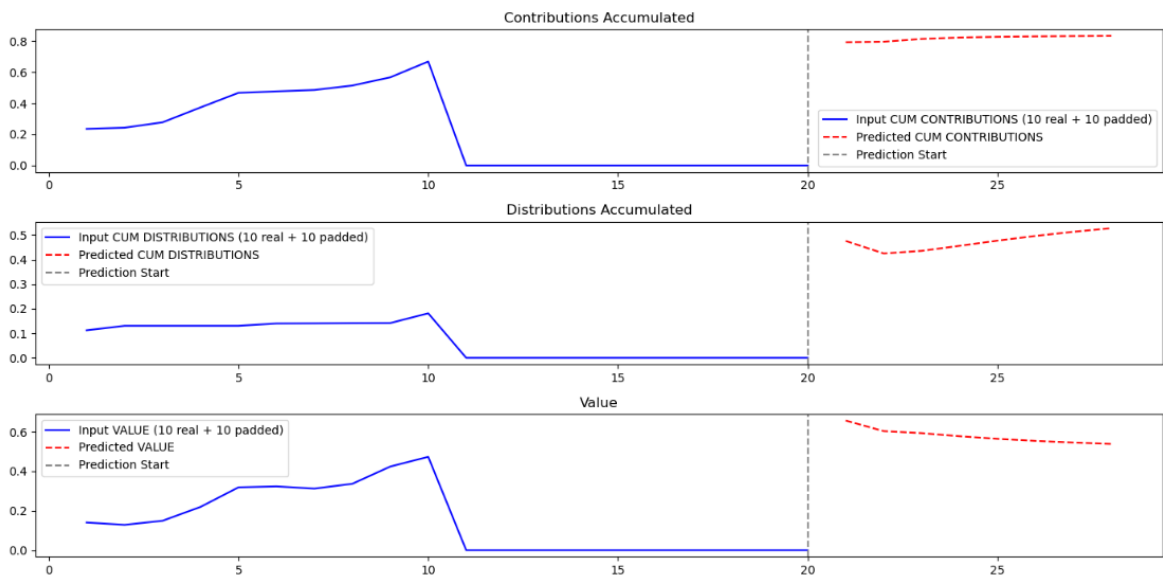


Figura 18: Predicción del modelo a partir de fondo con 10 trimestres de historia y 10 trimestres de padding.

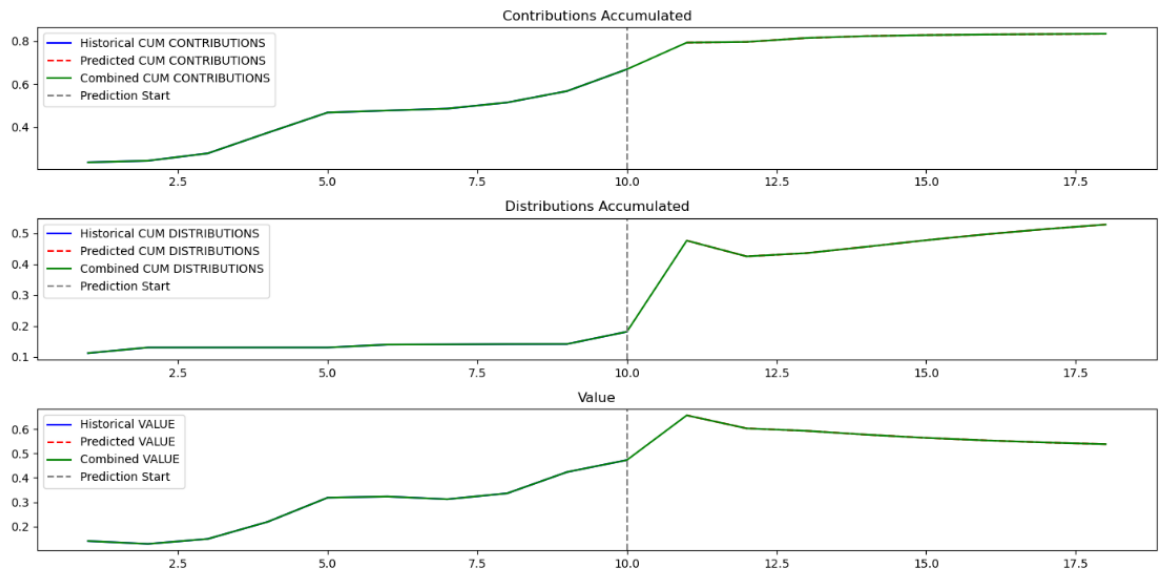


Figura 19: Resultado concatenado de la predicción a partir de los 10 trimestres de historia real.

Anexo C: Detalle Contribuciones, Distribuciones y NAV para el Segundo modelo

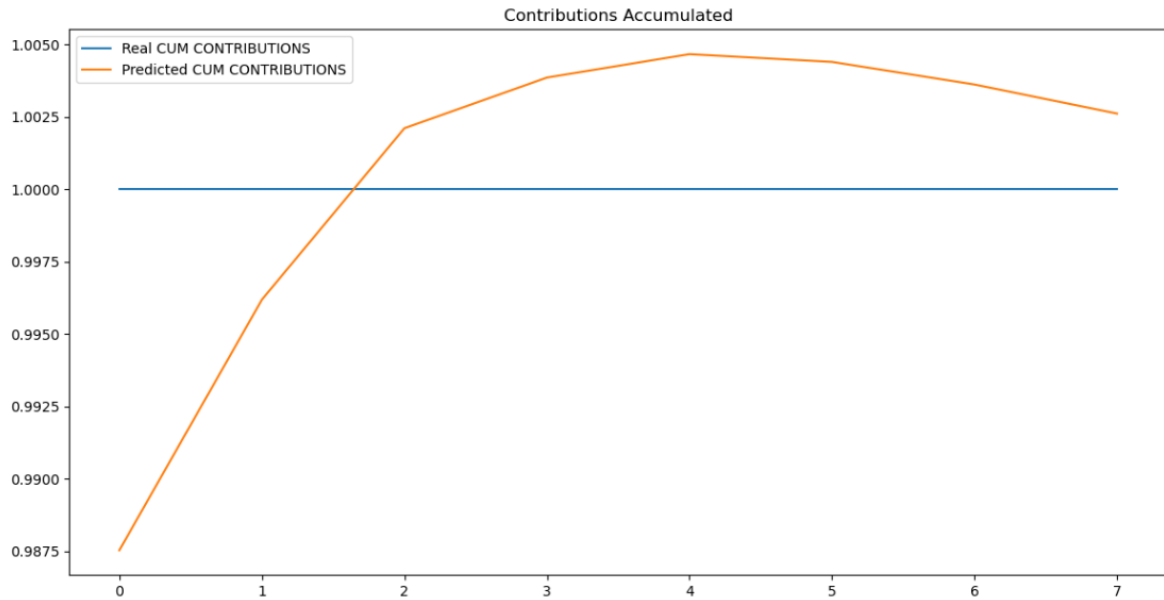


Figura 20: Contribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2).

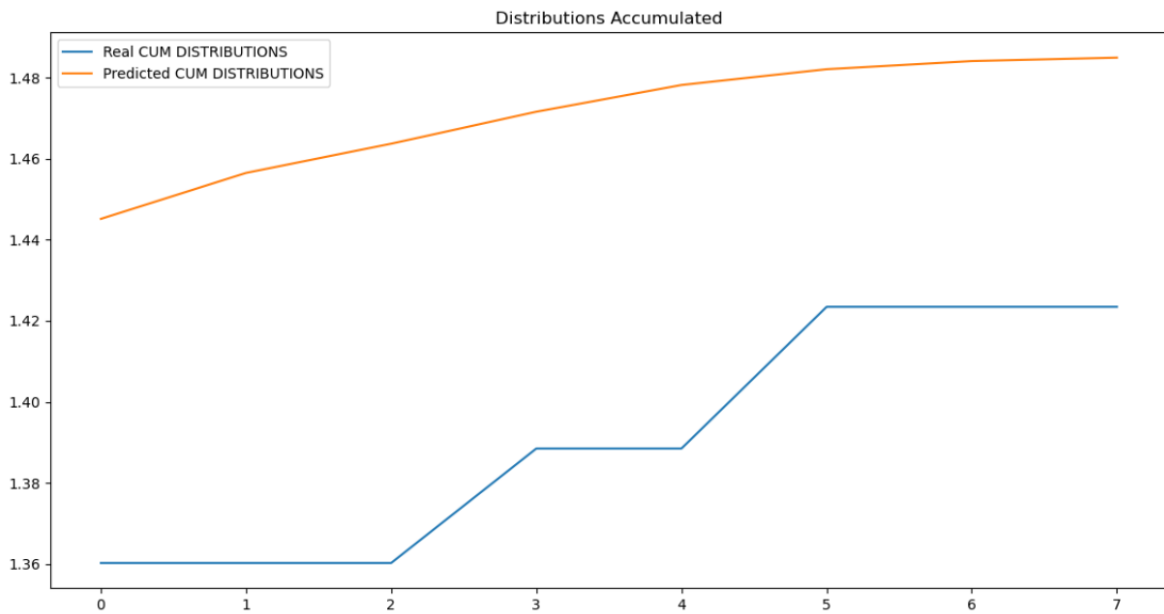


Figura 21: Distribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2).

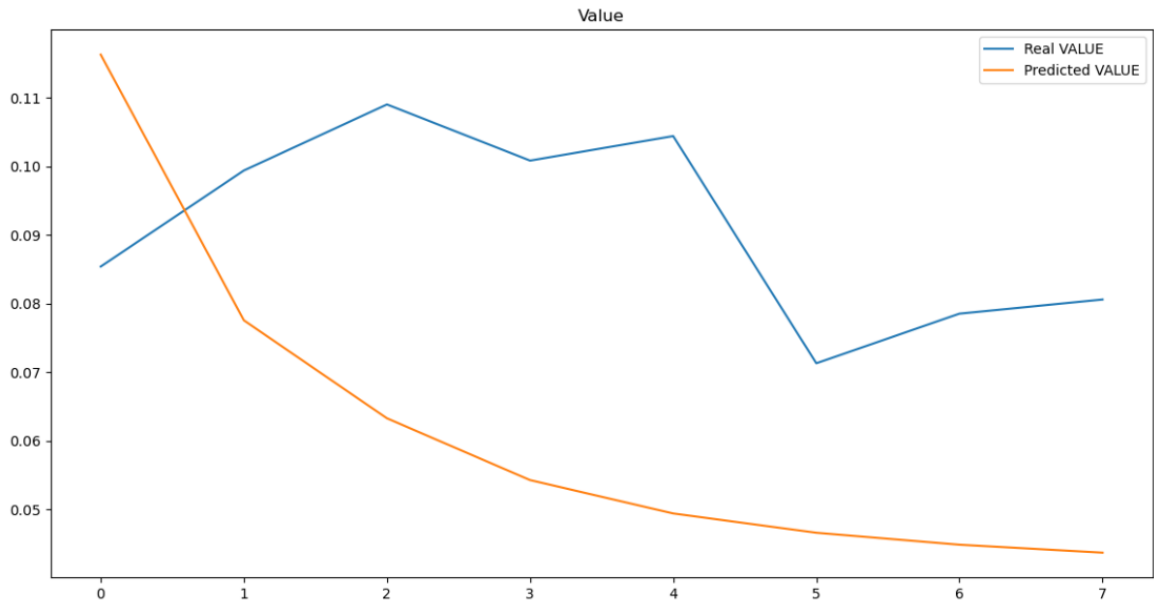


Figura 22: NAV predicho para 8 trimestres a partir de 20 trimestres de historia (modelo 2).

Anexo D: Detalle Contribuciones, Distribuciones y NAV para el Tercer Modelo

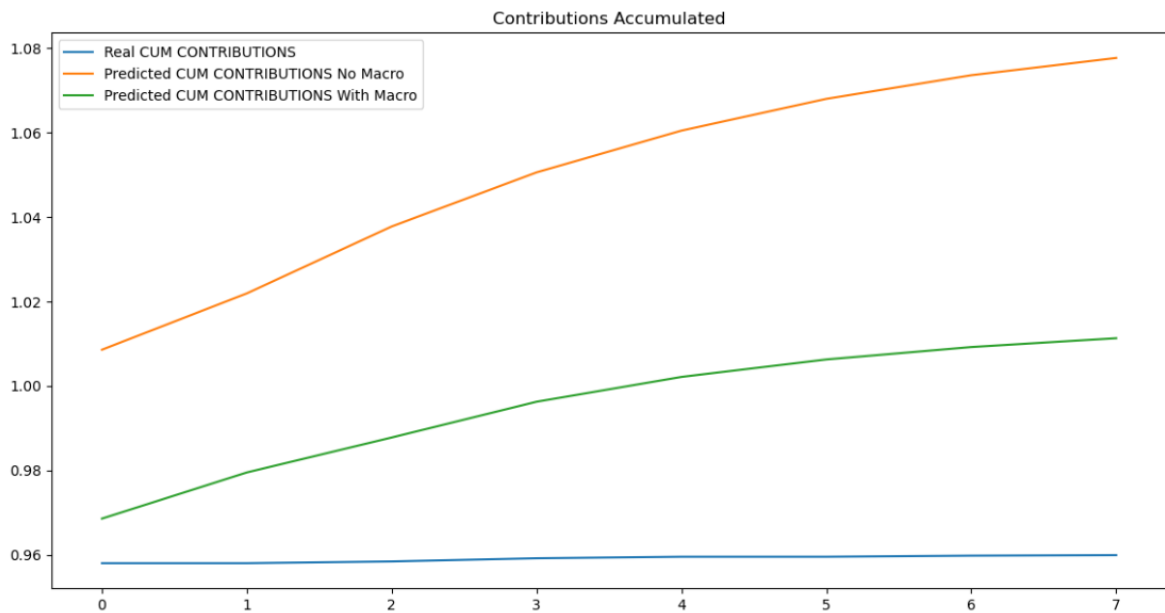


Figura 23: Contribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2 y 3).

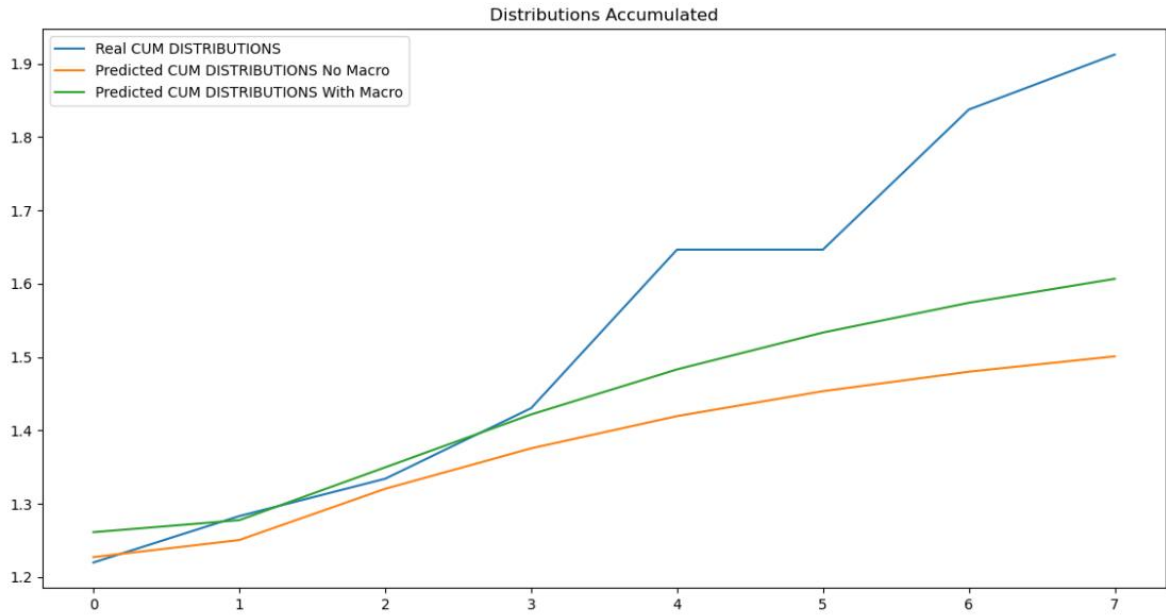


Figura 24: Distribuciones predichas para 8 trimestres a partir de 20 trimestres de historia (modelo 2 y 3).

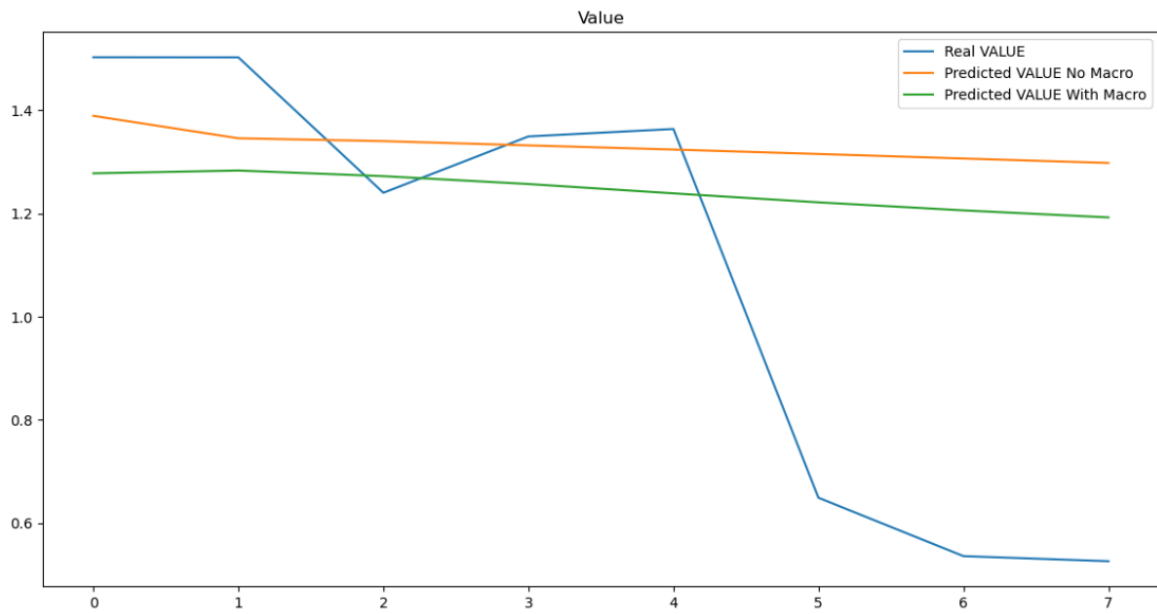


Figura 25: NAV predicho para 8 trimestres a partir de 20 trimestres de historia (modelo 2 y 3).

Anexo E: Glosario

ⁱ *CAIA (Chartered Alternative Investment Analyst)*: Corresponde a una certificación para analistas que trabajan en el mundo de los activos alternativos. La organización encargada de brindar dicha certificación se llama CAIA (Chambers et al., 2023).

ⁱⁱ Recursos naturales: Hace referencia a inversiones que se realizan en activos que derivan de la naturaleza, por ejemplo, agua, minerales, madera o petróleo (Finance Strategyst, s.f).

ⁱⁱⁱ Infraestructura: Hace referencia a inversiones realizadas sobre activos que corresponden a los cimientos de servicios básicos, facilidades o instituciones que dependen del crecimiento y desarrollo de la comunidad. Por ejemplo, redes de telecomunicaciones, redes de energía o carreteras (CFA Institute, s.f (1)).

^{iv} Propiedad intelectual: Intangibles que pertenecen a una compañía y son legalmente protegidos de su uso o implementación sin consentimiento.

^v *Venture Capital*: Forma de *Private Equity* y forma de financiar una *start up*, enfocado en empresas chicas con alto potencial de crecimiento (CFA Institute, s.f (2)).

^{vi} *Buyout*: Adquisición de una porción controladora de una empresa, esto con el objetivo de venderla en un futuro con un margen de ganancia (CFA Institute, s.f (2)).

^{vii} *Mezanine*: Híbrido entre deuda y propiedad, ya sea en formato de garantías, acciones u otros (CFA Institute, s.f (2)).

^{viii} *Distressed Debt*: Deuda emitida a empresas que cuentan con dificultades financieras (CFA Institute, s.f (2))

^{ix} *Family Office*: Oficina privada u organización profesional de una familia que posee un patrimonio importante. Usualmente, enfocados a invertir una proporción del patrimonio de dicha familia.

^x *Manager*: Sinónimo del administrador del fondo. Los *GP* desempeñan la labor de *manager* de un fondo.

^{xi} Distributions: Retornos que reparte el *GP* de un fondo a los LP una vez liquidada cierta inversión.

^{xii} Múltiplo: Se les dice múltiplo al *TVPI*, *RVPI* y *DPI*, debido a su forma matemática, donde representan una proporción o múltiplo respecto al dinero depositado en un fondo.

^{xiii} *OCIO (Outsourced Chief Investment Officer)*: Cuando el dueño de algún activo externaliza algunas o todas las decisiones relacionadas a manejo de portafolio a un tercero (Blackrock, s.f).

^{xiv} *Unfunded*: Capital que ha sido comprometido a una inversión pero que aún no ha sido transferido al *GP* (ILPA, s.f).

^{xv} *Output*: Corresponde a los resultados o variables que arroja un modelo como salida una vez ingresados los datos de entrada.

^{xvi} *Vintage*: Año en que comienza a invertir un fondo.

^{xvii} *Portfolio*: Conjunto de activos financieros pertenecientes a un inversor, ya sea persona natural, sociedad o institución.

^{xviii} *AFP*: Corresponde a la sigla para las administradoras de fondos de pensiones.

^{xix} *Multifondo*: 5 alternativas en las que se pueden elegir invertir tanto los aportes voluntarios como los aportes obligatorios realizados en las AFP. Se denominan alfabéticamente (A, B, C, D, E), cada uno con cierto nivel de riesgo y rentabilidad esperada.

^{xx} *Redes Neuronales*: Modelos computacionales que buscan emular la estructura de un cerebro utilizando nodos y generando conexiones entre estos (Huet, 2023).

^{xxi} *Python*: Lenguaje de programación.

^{xxii} *Outliers*: Datos o valores considerados anomalías o atípicos respecto al resto del conjunto.

^{xxiii} *Mascara*: Corresponde a definir algún criterio o valor para filtrar los datos que entran a un proceso, por ejemplo, un modelo. (Tensorflow, s.f).

^{xxiv} *Split*: Separación de datos en conjuntos.

^{xxv} *VaR (Value at Risk)*: Métrica financiera que estima el riesgo de una inversión. Específicamente, es una técnica estadística usada para medir la pérdida potencial que podría ocurrir en un portafolio dado cierto periodo de tiempo fijo. (Corporate Finance Institute, s.f).