



Universidad de
San Andrés

Universidad de San Andrés

Escuela de Administración y Negocios

Licenciatura en Finanzas

**Filtro de observaciones anómalas para la
calibración de estructuras temporales de
tasas. El caso del mercado bursátil argentino.**

Autor: Carlos Ignacio Pinto

Legajo: 31171

Director: Alejandro E. Loizaga

Buenos Aires, Argentina – 20 de septiembre de 2023

Tabla de contenido

Abstract	3
1. Introducción.....	3
2. Marco teórico.....	5
3. Descripción y discusión de la base de datos.....	9
3.1 Selección de activos.....	9
3.2 Precios del mercado.....	11
3.3 Tasa interna de retorno a partir del modelo Nelson y Siegel	12
4. Metodología.....	13
4.1 K-means.....	13
4.2 Medidas de error de ajuste del modelo	14
4.3 Medidas de dispersión de ajuste del modelo.....	16
5. Diseño del filtro y proceso de filtrado.....	17
5.1 Proceso de filtrado.....	18
5.2 Condición 1: presencia del último <i>spread</i> en <i>cluster</i> anómalo en serie temporal individual.	19
5.3 Condición 3: presencia del último <i>spread</i> en <i>cluster</i> anómalo en corte transversal grupal de vecinos de curva.....	21
6. Resultados	23
6.1 Errores modelo-mercado	23
6.2 Datos cualitativos y estadísticas porcentuales del filtrado	28
6.3 Movimiento de curvas en el período analizado	30
7. Conclusión.....	32
8. Bibliografía.....	35
9. Apéndices.....	38
9.1 Apéndice 1. Condición de Hartigan	38
9.2 Apéndice 2. P-valores del test de medias	41
9.3 Apéndice 3. Falencia del filtro de solo condición 1.	42
9.4 Apéndice 4. P-valores del <i>test</i> de Levene.....	44

Índice de gráficos

Gráfico 1. Proceso de filtrado	19
Gráfico 2. Agrupación de <i>spreads</i> del activo MAC2O por <i>clustering</i> para el 10 de abril de 2023.	21
Gráfico 3. <i>Spreads</i> de activos agrupados por <i>clustering</i> para el 10 de abril de 2023.....	23
Gráfico 4. Historial de errores modelo-mercado calculados con todas las entradas de precios.....	26
Gráfico 5. Historial de errores modelo-mercado calculados solo con precios no filtrados.....	27
Gráfico 6. Historial de TIR modelo, TIR entradas y TIR filtradas con cada filtro.	29
Gráfico 7. Cantidades y porcentajes diarios promedio para cada filtro.....	30
Gráfico 8. Historial de curvas de tasas de interés a partir de las diferentes calibraciones.	31
Gráfico 9. Agrupación de <i>spreads</i> del activo TLCDO por <i>clustering</i> para el 23 de mayo de 2023.	43
Gráfico 10. Agrupación de <i>spreads</i> del activo MAC2O por <i>clustering</i> para el 23 de mayo de 2023.	43

Índice de tablas

Tabla 1. Activos renta fija <i>dólar link</i> a utilizar.....	11
Tabla 2. Promedios de errores diarios calculados con todas las entradas de precios.	24
Tabla 3. Promedios de errores diarios calculados solo con precios no filtrados.	25
Tabla 4. Promedios de errores diarios previos al 29 de mayo del 2023 calculados con todas las entradas de precios.....	28
Tabla 5. Desvío estándar diario de parámetros del modelo Nelson y Siegel.	32
Tabla 6. Valores de condición de Hartigan en <i>clustering</i> de primera condición.....	40
Tabla 7. Valores de condición de Hartigan en <i>clustering</i> de segunda condición.....	41
Tabla 8. P-valores de <i>test t</i> de Welch	42
Tabla 9. P-valores de <i>test</i> de Levene	44

Abstract

Las investigaciones sobre filtros para información de mercado como paso previo a la construcción de un modelo de tasas de interés son escasas y no se realizó ninguna utilizando *K-means*. El estudio se realizó para dieciocho activos *dólar link* argentinos teniendo en cuenta la importancia de construir su respectiva estructura de tasas de mercado previo a las elecciones PASO durante abril del 2023, con el objetivo de realizar un filtro utilizando *clustering* que su aplicación genere un mejor ajuste. Para la elaboración del filtro, se establecieron dos condiciones que permiten distinguir entre valores atípicos y datos normales, con el fin de minimizar la probabilidad de filtrar información relevante.

La eficacia del filtro se evaluó mediante la comparación de métricas de error entre el modelo y el mercado, así como es desvío estándar de los parámetros del modelo de Nelson y Siegel (1987), se empleó el *test* de Levene para comparar la volatilidad entre los parámetros de la curva sin filtro y los generados a través del modelo de filtrado y no se encontró diferencias estadísticamente significativas entre ellas. Los resultados indican que la aplicación del filtro propuesto no reduce significativamente el error entre modelo-mercado cuando se calculan las métricas con todos los datos de precios y, de hecho, una volatilidad ligeramente mayor en las curvas. Sin embargo, al calcular las métricas con precios no filtrados, se observa una disminución del error modelo-mercado, aunque sin diferencias estadísticamente significativas en ningún caso. Además, se identificó que la condición 2 del filtro, cuando se aplica sola, ofrece un mejor desempeño, mientras que la condición 1, cuando se aplica por sí sola, empeora los resultados al eliminar el 44% de los datos en promedio.

1. Introducción

Es innumerable la información que contiene implícitamente una curva de tasas de interés del mercado bursátil. A lo largo de cincuenta años, numerosos investigadores en el área de las finanzas han dedicado un amplio conjunto de investigaciones al minucioso análisis de esta estructura. La importancia de comprender esta estructura radica en que, ya sea como inversores o como estudiosos del campo financiero, nos proporciona una valiosa visión de las tasas que prevalecen en un mercado específico, así como información sobre la economía subyacente y las expectativas del mercado al que se refiere.

En este contexto, la propia construcción de la curva de tasas de interés se posiciona como una de las secciones más fundamentales de estudio. Esto se debe a que la diversidad de condiciones y contextos en los mercados financieros dieron lugar a la creación de una variedad de métodos de estimación. Sin embargo, la falta de un enfoque único para determinar estas tasas ha resultado en múltiples representaciones de las curvas de tasas, lo que complica aún más la situación cuando la información necesaria para aplicar estos métodos es escasa, inexistente o no representa adecuadamente el mercado en cuestión. Este desafío es particularmente relevante y recurrente en mercados emergentes, y por ello, la búsqueda de soluciones es de una gran relevancia.

Dentro de las curvas de tasas de interés se encuentran las expectativas del mercado, y en este contexto, en un momento tan crucial como las elecciones presidenciales de 2023 en Argentina y la posibilidad de un cambio en las políticas económicas, la curva de activos emitidos en dólares estadounidenses y pagaderos en pesos argentinos adquiere un interés especial para los inversionistas, ya que podría ser vista como una forma de cobertura. Sin embargo, la presencia de *outliers* en los datos del mercado presenta un desafío al intentar obtener una calibración precisa de la curva de estos activos *dólar link*. Este es el problema central que aborda esta investigación.

A primera vista, podría parecer que la utilización de todos los precios de mercado de estos activos para calibrar una curva con el modelo de Nelson y Siegel (1987) resultaría en una estructura de tasas que refleja fielmente el mercado. Sin embargo, la escasez de datos y la presencia de valores atípicos mencionados anteriormente a menudo impiden que esto sea así. No obstante, la literatura que trata de métodos para abordar esta problemática parece ser escasa.

El objetivo principal de este trabajo es evaluar si la aplicación de un filtro que permita detectar y eliminar los valores atípicos (o *outliers*) en los precios de mercado de los activos mencionados puede generar una calibración de las curvas de tasas con un menor error modelo-mercado. Es importante destacar que la definición de *outlier* es endógena al filtro, es decir, se consideran *outliers* a aquellas observaciones que el propio filtro identifique como tales. Esta aclaración se hace para no descartar la posibilidad de que las observaciones eliminadas sean representativas del mercado; simplemente buscamos mejorar la calibración de las curvas eliminando las entradas que afecten negativamente su ajuste.

Para lograr este objetivo, utilizamos el modelo de Nelson y Siegel para la calibración de las curvas y desarrollamos un algoritmo de *clustering* que nos permite construir el filtro necesario para eliminar los valores atípicos de los datos antes de la calibración de la curva de tasas. Además, realizamos pruebas adicionales utilizando otros métodos que incorporan únicamente una parte del filtro propuesto, lo cual nos ha permitido comprender su eficacia y funcionamiento en el contexto general del filtro completo. Finalmente calibramos las curvas para el mismo periodo sin usar ningún filtro para contrastar el ajuste de esas curvas con las filtradas. Cabe aclarar que no es la intención de este trabajo adentrarnos en la calibración de la curva mediante el modelo de Nelson y Siegel (1987), sino sólo en la construcción y funcionamiento del filtro.

La estructura del trabajo es la siguiente: en la sección 2 se realiza una revisión literaria para conocer trabajos previos donde se haya aplicado *clustering* y/o modelos de filtrado para la temática descrita, luego, en la tercera sección se describe el período a analizar, los instrumentos que se utilizan y los criterios utilizados para el armado de la base de datos que utilizamos en nuestro análisis. En la cuarta sección se hace una breve descripción de las metodologías utilizadas para la elaboración del filtro, así como de las funciones de error que se emplean para medir el desempeño de las estimaciones. En la quinta sección se explica el diseño del filtro y el proceso en el que se llevó a cabo el filtrado. Por último, en la sexta sección se exponen los resultados obtenidos para las calibraciones con filtro como las sin filtro, y su respectivo análisis para en la séptima sección sacar conclusiones.

2. Marco teórico

En la literatura del campo de las finanzas se ha escrito sobre diversas metodologías para el armado de estructuras de tasas de interés (ETI). Existen dos variantes de estas curvas: las dinámicas y las estáticas. Las **curvas estáticas** tienen a su vez dos corrientes teóricas: por un lado, los que utilizan **splines polinómicos**, donde su principal exponente es Mc Culloch con sus investigaciones de 1971 y 1975; y, por otro lado, los que emplean **modelos parsimoniosos**, aquí se encuentra el modelo propuesto por Nelson y Siegel de 1987, el cual implicó un gran aporte en las finanzas no solo para entender la relación entre tasas de diferentes instrumentos de renta fija, sino también para poder elaborar una ETI. Sin embargo, la diversidad de mercados financieros en el mundo hizo que exista gran cantidad de trabajos enfocados en la aplicación del modelo de Nelson y Siegel. Con el

propósito de llevar a cabo esta investigación enfocada en mercados ilíquidos como el argentino, resulta de interés analizar literatura al respecto.

En 2002, Arango et al. utilizaron el modelo de Nelson y Siegel para estimar la curva de tasas de los CETES¹ de la Bolsa de Valores de Colombia y analizar la inflación esperada implícita en dichas estructuras de tasas. Respecto al mercado de renta fija peruano, Rodríguez y Villavicencio (2002) volvieron a poner en práctica dicho modelo para construir la curva de rendimientos en nuevos soles para bonos cero cupón, y a partir de ella, analizar la efectividad de la política macroeconómica de Perú. Además, es relevante analizar los trabajos de Cortazar y Riesco (2003) y Cortazar y Espinoza (2005), donde se aplican diferentes métodos de modelado de curvas de tasas para el mercado de renta fija chileno, teniendo en consideración el problema de iliquidez que presenta. Posteriormente, respecto al mercado argentino, es de interés la investigación de Corso y Matarelli (2019) que aplicaron el modelo de Nelson y Siegel en tasas *spot* de las letras del Tesoro del Gobierno Nacional para poder analizar las tasas de inflación implícita en dichas curvas. Por otro lado, Alberti (2019) realiza un armado de una curva de tasas de interés en dólares para el mercado argentino utilizando el modelo de Nelson y Siegel explicado anteriormente. El autor al presentar sus resultados encuentra que en los datos del mercado argentino existe mayor presencia de *outliers* cuando se trata de un período muy volátil. Si bien Cortazar y Riesco (2003) analizan problemas de mercados ilíquidos, como la falta de información, no analizan el efecto de la posible presencia de *outliers* en los datos recabados. Entonces, ya que estas observaciones anómalas pueden tener efecto en el ajuste del modelo es de nuestro interés analizarlas, pero las investigaciones realizadas sobre métodos para su detección y su remoción son pocas. Dacorogna et al. (2001) explica que dicha escasez de investigación se debe a que la etapa de limpieza de datos suele pasar desapercibida, ya que es muy técnica y al formar parte del desarrollo previo al procesamiento de bases de datos, no deriva en los resultados que interesan al público. Sin embargo, dichos autores y posteriormente, Brownless y Gallo (2005) llevaron a cabo las investigaciones más importantes en el rubro sobre la detección de *outliers*. Aunque dichos autores, en sus respectivos trabajos, plantean métodos diferentes, ambas técnicas consisten en seleccionar una ventana temporal de la serie de datos, estandarizarlos de una determinada manera, y luego evaluar el dato para determinar si está lo suficientemente alejado de los anteriores para ser considerado o no un *outlier*. Así, se plantean algoritmos

¹ Certificado de Tesorería de la Federación

de limpieza centrados en remover *ab initio* las observaciones “basura” más recientes y funcionan de manera secuencial e iterativa.

La investigación de Dacorogna et. al. será de gran apoyo para nuestra investigación porque los autores enfocan la detección de anomalías como un proceso de filtrado en el que se somete a una serie de datos a un filtro donde aquellas observaciones que se consideran *outliers* serán sustraídas.

En su libro *An Introduction to High Frequency Finance* (2001), Dacorogna et. al. proponen filtros que se basan en evaluar la credibilidad de una observación con un valor entre 0 y 1 para calificarla como *outlier* o no según un “umbral” (Dacorogna et. al. usan 0,5). Así, los autores plantean un filtro que va adaptando sus parámetros de clasificación en base a los datos observados en una ventana que ellos denominan *time stamp*. El *time stamp* toma mucha importancia al inicio del filtrado, ya que determina la ventana temporal de los datos con los que se entrena el filtro. El período de entrenamiento lo denominan *built up period*, y el inmediato siguiente, período de filtrado. El proceso de filtrado que plantean tiene tres etapas principales: primero, se estandarizan los datos; luego, se evalúa la credibilidad del dato en análisis; y finalmente, se lo clasifica como *outlier* o no. Conjuntamente, Dacorogna et al. describen dos tipos de filtros: el filtrado individual y el filtrado de pares, la diferencia entre ambos radica en que el primero filtra la observación a evaluar teniendo en cuenta sus vecinos, y la segunda, filtra comparando dos grupos que no son vecinos. Como nuestro filtro es individual, más adelante definimos el término “vecinos” en el contexto de esta investigación.

Por otro lado, con el avance de la tecnología surgen nuevos métodos de *Machine Learning* que corresponden analizar. De esta manera, dentro de esa técnica, encontramos el *Data Mining* que ofrece una posible solución al problema de la limpieza de datos: el método de *clustering* es el que se analiza en esta investigación. Por ese motivo, a continuación, nombramos las investigaciones de ese rubro que tomamos de base para la construcción del filtro.

Cai et. al. (2016) ubica al *clustering* como uno de los principales métodos de *Data Mining* utilizados en el campo de las finanzas. Los autores dividen dicho método en tres categorías:

1. En primer lugar, encontramos los **métodos de partición**: el principal representante de esta categoría es el algoritmo de *k-means*, en el cual el objetivo es dividir las n observaciones en k *clusters*. Cada observación pertenece a un *cluster*. Entonces, como todos los datos a clasificar se tratan con la misma

importancia, se toma una media como centroide de las observaciones de cada grupo o *cluster*.

2. En segundo lugar, se encuentran los **métodos basados en densidad**, donde los más usados son *OPTICS* o *DBSCAN*. Tang et. al. (2022) describen cómo esta técnica es muy útil para predecir los movimientos a corto plazo de los precios de acciones.
3. Por último, ***Data Stream Clustering***, este método agrupa todas las observaciones que pertenecen a un grupo de características predefinidas, las que no cumplen formarían sus propios grupos. Dentro de esta técnica de *clustering* se encuentra el método *On-line evolving approach*, en finanzas, se utiliza para la detección de anomalías en bases de datos.

Esta última categoría es puesta a prueba por Omanovic et. al (2007), quienes prueban la eficacia de dicho método para detectar anomalías en datos contables. Sin embargo, como es un método en donde finalmente la clasificación definitiva de “anomalía” la asigna quien opera el algoritmo, los autores concluyen que a partir de esa técnica de *clustering* se obtienen resultados exitosos, siempre y cuando quien lo opere clasifique bien los *outliers*, ya que de esa forma se entrena el algoritmo.

En otro orden de ideas, Gutiérrez y Gregori (2008) describen otra aplicación del *clustering* para la detección de *outliers* en el campo de las finanzas. Aunque las autoras no utilizan la técnica de *clustering* para filtrar las anomalías, retoman las investigaciones de Dacorogna et. al. (2001) y Brownless y Gallo (2005) con series previamente clasificadas. De esta manera, la novedad de su estudio consiste en que previamente aplicaron métodos de partición sobre veinticinco series para dividir las en grupos en base a sus características, así, van evaluando sobre cada uno de ellos la eficiencia de los métodos de detección de *outliers* descritos anteriormente. De esta manera, pueden ir variando los parámetros por grupos, para finalmente concluir en cuáles son los parámetros ideales para la serie según el instrumento financiero del que provenga.

Llegando a este punto, son numerosos los trabajos escritos sobre la aplicación de *clustering* en las finanzas y día a día son más. Sin embargo, son escasas las investigaciones realizadas sobre dicha técnica aplicada a la detección de *outliers*, y menos aun las que utilizan el método de partición para esa práctica. El propósito de este trabajo será utilizar el método de *clustering k-means* para detectar anomalías en los datos de precios del mercado argentino y removerlas para obtener un mejor ajuste en la calibración de la curva de tasas por medio del modelo de Nelson y Siegel (1987).

3. Descripción y discusión de la base de datos

El período seleccionado para realizar las pruebas del método de filtrado de *outliers* es desde el 27 de marzo de 2023 hasta el 1 de junio de 2023. Aunque podemos realizar el filtro de precios para un solo día de cálculo de una curva de tasas de interés, decidimos realizarlo para cada día hábil de dicho período. De esta manera, contamos con un período lo suficientemente largo como para poder comparar los resultados de la calibración de curva entre precios sin filtrar y precios filtrados. Además, dado que el proceso de filtrado debe ser secuencial e iterativo (Gutiérrez y Gregori, 2008), a medida que se vayan construyendo las curvas de cada día irán formando parte de la base de datos requerida para filtrar los datos y construir el modelo al día siguiente.

3.1 Selección de activos

Los activos que utilizamos para nuestro análisis son instrumentos de renta fija del mercado argentino de similares características, con moneda de denominación en dólares estadounidenses y moneda de pago en pesos, también denominados *dólar link*. Dichos activos utilizan como referencia el tipo de cambio de referencia comunicación "A" 3500 del Banco de la Nación Argentina. De aquí en adelante nos referiremos a ellos como grupo de activos *dólar link*.

Bono	Denominación	Fecha de emisión	Fecha de vencimiento	Monto emitido	ISIN	Descripción	Amortización
DV04A	FF Agro Solutions DVA S4 Clase A	22/12/2022	10/8/2023	US\$2.429.756,00	AREQ UI600 VL3	3% anual, con pago mensual pagaderos por período vencido. Las fechas de pago son el día 10 de cada mes desde abril 2023 hasta el vencimiento.	En 5 cuotas mensuales las mismas fechas de los pagos de intereses.
HB04A	FF HAB IV	3/11/2022	20/5/2024	US\$5.648.827,00	ARRO SR600 ATO	3% anual, en 2 pagos pagaderos por período vencido. Las fechas de pago son: 20/9/2023 y 20/5/2024	En 2 cuotas, la primera el 20/9/2023 y la segunda el 20/5/2024.
ATS10	ON Agrality Serie 1	11/5/2023	11/5/2023	US\$10.000.000,00	ARAG TY560 015	Como la tasa de interés aplicable es de 0%, no se realizan pagos de intereses.	En 3 cuotas trimestrales. La primera del 33% el 11/11/2024, la segunda de 33% el 11/2/2025 y la última en la fecha de vencimiento de 34%
MAC10	ON MOLINOS AGRO C1	18/5/2021	18/5/2023	US\$28.462.533,00	ARMO AG560 030	1.5% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 18 de cada mes desde agosto de 2021 hasta el vencimiento.	Íntegramente al vencimiento

MAC2 O	ON MOLINOS AGRO C2	18/5/2021	18/5/2024	U\$S25.392.110,00	ARMO AG560 048	2.5% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 18 de cada mes desde agosto de 2021 hasta el vencimiento.	Íntegramente al vencimiento
MAC3 O	ON MOLINOS AGRO C3	3/9/2021	3/9/2023	U\$S33.539.700,00	ARMO AG560 055	1% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 3 de cada mes desde diciembre de 2021 hasta el vencimiento.	Íntegramente al vencimiento
MAC4 O	ON MOLINOS AGRO C4	3/9/2021	3/9/2024	U\$S36.460.000,00	ARMO AG560 063	1% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 3 de cada mes desde diciembre de 2021 hasta el vencimiento.	Íntegramente al vencimiento
PN1CO	ON Pan American Energy C1	16/7/2020	16/7/2022	U\$S20.298.845,00	ARAX IO5600 84	Como la tasa de interés aplicable es de 0%, no se realizan pagos de intereses.	Íntegramente al vencimiento
PN5CO	ON Pan American Energy C5	3/9/2020	3/9/2024	U\$S50.000.000,00	ARAX IO5600 B3	3% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 3 de cada mes desde diciembre de 2021 hasta el vencimiento.	Íntegramente al vencimiento
PN7CO	ON Pan American Energy C7	19/11/2020	19/11/2025	U\$S50.000.000,00	ARAX IO5600 E7	4.75% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 19 de cada mes desde febrero de 2021 hasta el vencimiento.	Íntegramente al vencimiento
PN9CO	ON Pan American Energy C9	12/2/2021	12/2/2024	U\$S30.000.000,00	ARAX IO5600 H0	Como la tasa de interés aplicable es de 0%, no se realizan pagos de intereses.	Íntegramente al vencimiento
PNECO	ON Pan American Energy Clase 13	12/7/2021	12/7/2031	U\$S260.136.729,00	ARAX IO5600 L2	5% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 12 de cada mes desde octubre de 2021 hasta el vencimiento.	En 14 cuotas semestrales los días 12 de cada mes desde enero del 2025 hasta el vencimiento. Los porcentajes de amortización serán de 5% desde la cuota 1 hasta la 8 incluida y desde la 9 a la última del 10% del capital.
PN1CO	ON Pan American Energy C17	7/2/2022	7/2/2032	U\$S80.000.000,00	ARAX IO5600 Q1	4.75% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 7 de cada mes desde mayo de 2022 hasta el vencimiento.	En 14 cuotas semestrales los días 7 del mes desde agosto del 2025 hasta el vencimiento. Los porcentajes de amortización serán de 5% desde la cuota 1 hasta la 8 incluida y desde la 9 a la última del 10% del capital.
PNJCO	ON Pan American Energy C18	7/2/2022	7/2/2027	U\$S25.097.123,00	ARAX IO5600 P3	1.25% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 7 de cada	Íntegramente al vencimiento

						mes desde mayo de 2022 hasta el vencimiento.	
TLC9O	ON Telecom Argentina C9	7/6/2021	7/6/2024	US\$91.831.939,00	ARTE CO560 0B6	2.75% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 7 de cada mes desde septiembre de 2021 hasta el vencimiento.	Íntegramente al vencimiento
TLCDO	ON Telecom Argentina C12	9/3/2022	9/3/2027	US\$97.722.400,00	ARTE CO560 0E0	1% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 9 de cada mes desde junio de 2022 hasta el vencimiento.	Íntegramente al vencimiento
TLCFO	ON Telecom Argentina C14	10/2/2023	10/2/2028	US\$62.390.111,00	ARTE CO560 0G5	1% anual, realiza un pago trimestral a periodo vencido. Las fechas de pago son el día 10 de cada mes desde mayo de 2023 hasta el vencimiento.	Íntegramente al vencimiento
TLCGO	ON Telecom Argentina C15	2/6/2023	2/6/2026	US\$87.403.386,00	ARTE CO560 0H3	Como la tasa de interés aplicable es de 0%, no se realizan pagos de intereses.	Íntegramente al vencimiento

Tabla 1. Activos renta fija *dólar link* a utilizar

3.2 Precios del mercado

Para nuestra investigación, la base de datos que contiene los precios resultantes de operaciones de mercado está compuesta de la información proveniente de tres mercados bursátiles argentinos: MAE (Mercado Argentino Electrónico) y BYMA (Bolsa y Mercados Argentinos).

Debemos tener en cuenta que puede no existir información de los precios de algunos activos del grupo para todo el período (Riesco y Cortázar, 2003). Entonces, para esos activos, les calcularemos un “precio modelo” con la TIR modelo resultante de la calibración de ese día a través del método de flujos descontados descrito por Fabozzi F. (2007).

Ahora bien, con la información de precios del grupo de activos de la Tabla que recopilamos de los mercados descritos anteriormente, armamos una base de datos que contiene todos los precios necesarios para el cálculo de las curvas de tasas. A continuación, se describen las categorías de la tabla elaborada:

Fecha: fecha del precio del activo.

Mercado: mercado del que proviene el precio del activo. Puede ser “MAE” o “BYMA.

Ticker: corresponde al código de bolsa con el que se identifica el activo (coincide con el mencionado en la Tabla 1).

Plazo: corresponde al plazo de liquidación con el que se operó el activo. Puede contener las palabras: “contado”, “24 horas”, “48 horas” o “72 horas”.

Precio de cierre: es el último precio con el que se concretó una operación en la fecha correspondiente para el activo. En caso de ser 0 significa que no hubo operación registrada.

Cantidad: indica la cantidad de valores nominales en dicha operación del activo.

Precio de compra: precio de demanda denominado “Bid”. Si no se dispone de dicho valor, se asigna 0.

Precio de venta: precio de oferta denominado “Ask”. Si no se dispone de dicho valor, se asigna 0.

3.3 Tasa interna de retorno a partir del modelo Nelson y Siegel

Se analiza para cada uno de los bonos descritos en la Tabla 1 la tasa interna de retorno (TIR):

$$\text{Precio bono} = \sum_{i=1}^n \frac{FF_i}{(1 + TIR)^i}$$

Asimismo, la TIR de un activo particular obtenida por cálculo del modelo Nelson y Siegel (1987), en adelante se denomina “TIR modelo” y la proveniente de su precio de mercado “TIR de entrada”.

Para empezar el período de prueba necesitamos un *build up period* de 21 días, es decir que necesitamos contar con la información de las TIR modelo sin filtrar de 21 días hábiles anteriores al primer día del período de filtrado. Para ello, calibramos el modelo y recopilamos la información de las TIR modelo obtenidas para el grupo de activos en análisis. Esta base de datos es de suma importancia, ya que nos permite empezar a aplicar el filtro en el período de interés descrito anteriormente.

Además, también necesitamos calibrar el modelo para el período de análisis de la misma forma que antes (sin filtro), porque con esas tasas calculamos los precios modelo que luego se comparan con los precios del modelo en que se aplicó el filtro. De este modo, ordenamos la información de ambos períodos en una base de datos dividida en las siguientes categorías:

Fecha de cálculo: es el día que se realizó el cálculo de las TIR de los activos.

Código: corresponde al código de bolsa con el que se identifica el bono.

Precio de entrada: es la ponderación por volumen de los precios registrados en los diferentes mercados descritos anteriormente

TIR de entrada: es la tasa interna de retorno calculada en base al Precio de entrada de instrumentos *dólar link* de la tabla nombrada anteriormente.

TIR modelo: es la tasa proveniente de la curva calibrada mediante el modelo Nelson y Siegel con los precios de entrada en base a su duración.

Duración Macaulay: definimos la duración de Macaulay (1938) de la siguiente manera:

$$Duracion_i = \left[\frac{c}{(1+y)^1} + \frac{2 * c}{(1+y)^2} + \dots + \frac{n * c}{(1+y)^n} \right] * \frac{1}{p},$$

siendo $Duración_i$ la duración del activo i , donde c es el cupón del bono, y es la TIR del bono y p es el precio observado.

Precio del modelo: es el precio calculado mediante el modelo de flujos descontados utilizando la TIR modelo.

Finalmente, hay que tener en cuenta que la información obtenida de la calibración de las curvas **sin** el filtro para ese período se recopiló en una base de datos, la cual tiene las mismas categorías descritas anteriormente. De esta manera, comparamos ambos resultados.

4. Metodología

4.1 K-means

En primer lugar, debemos definir el algoritmo de *clustering k-means*. Dicha técnica, como se describió anteriormente, es una categoría de *clustering* perteneciente a los métodos de partición y es la herramienta principal que utilizamos para el funcionamiento del filtro de detección de *outliers*. Podemos describir dicho algoritmo mediante la siguiente fórmula:

$$\min_{\{\mu_1, \dots, \mu_k\}} \sum_{k=1}^C \sum_{j=1}^{C_k} (||x_j^k - \mu_k||)^2$$

donde,

$||x_j^k - \mu_k||$ es la distancia euclídea entre x_j^k y μ_k ,

x_j^k es la “ j ” esima observación del “ k ” esimo *cluster*, proveniente de un conjunto X de n observaciones $X = \{x_1, \dots, x_n\}$,

C_k es la cantidad de observaciones en el “k” esimo *cluster*,

μ_k es el centroide del “k” esimo *cluster*,

C es el número de *clusters* en nuestro análisis será $C = K$,

a su vez, denominamos M al conjunto de centroides $M = \{\mu_1, \dots, \mu_k\}$.

El agrupamiento mediante este algoritmo de *clustering* se divide en las siguientes etapas:

1. Etapa inicial: Debemos seleccionar dos condiciones para la inicialización. La primera es determinar en cuántos subconjuntos agruparemos los datos (es decir, cuál será el valor de K). La segunda es la selección inicial de los centroides, la cual se realizará de manera aleatoria.
2. Etapa de clasificación: donde se realiza el cálculo de distancia $(\|x_i - \mu_k\|)^2$ para cada x_i de X y con cada centroide μ_k de M ; y que se asigna el objeto al centroide del grupo cuya distancia es la menor, $x_i = x_j^k$, si $(\|x_i - \mu_k\|)^2 < (\|x_i - \mu_l\|)^2$, donde $\mu_k \neq \mu_l$ y $(\mu_k, \mu_l) \in M$.
3. Etapa de cálculo de centroides: una vez que se generaron los centroides propios de cada *cluster*, el algoritmo realiza un proceso de iteración recalculando esos centroides buscando un mejor ajuste, es decir una menor suma de distancias de observaciones a sus respectivos centroides
4. Etapa de convergencia: se debe establecer un criterio para que la iteración se detenga, en este caso será cuando se alcance el número de iteraciones K definido previamente. Si la condición de convergencia no se cumple se vuelve a repetir las etapas 2, 3 y 4.

4.2 Medidas de error de ajuste del modelo

Debemos tener en cuenta que contamos con resultados para dos modelos (con y sin filtro). Así, debemos definir la manera en la que mediremos el nivel de ajuste de cada modelo, y así, compararlos y evaluar la eficacia del filtro diseñado. Si bien existe una gran cantidad de medidas de error modelo-mercado, en este trabajo se utilizan las siguientes cuatro:

- Error Absoluto Promedio:

$$MAE = \sum_{i=1}^N \frac{|P_i^* - P_i|}{N}$$

- Error Absoluto Relativo Promedio:

$$MAPE = \sum_{i=1}^N \frac{|P_i^* - P_i|}{P_i} * \frac{100}{N}$$

- Raíz cuadrada del error cuadrático promedio:

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(P_i^* - P_i)^2}{N}}$$

- Raíz cuadrada del error cuadrático relativo promedio:

$$RMSPE = \sqrt{\sum_{i=1}^N \frac{(P_i^* - P_i)^2}{P_i} * \frac{100}{N}}$$

Donde P_i^* es el precio modelo del activo i de un conjunto de N activos y P_i el precio de entrada del activo i .

Estas medidas se calculan para cada día del período seleccionado. El criterio de decisión se basa en la elección del modelo que menor promedio de error contenga.

Adicionalmente, realizamos un *test* de medias entre la serie de errores normales y contra el error obtenido en cada uno de los tres filtro para cada uno de los cuatro tipos de errores descritos anteriormente sin asumir igualdad de varianza, para eso utilizamos el *test t* de Welch (1947). Las hipótesis nula del *test* mencionado es que las media de las dos series de errores evaluados son iguales y la hipótesis alternativa es que son diferentes (ya que realizamos el *test* es de dos colas). La fórmula para calcular el estadístico t en este caso es la siguiente:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \cdot \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \frac{1}{n_1 + \frac{1}{n_2}},$$

donde:

- \bar{X}_1 y \bar{X}_2 son las medias de las dos muestras **a** y **b**, respectivamente.
- s_1^2 y s_2^2 son las varianzas de las dos muestras **a** y **b**, respectivamente.
- n_1 y n_2 son los tamaños de las dos muestras **a** y **b**, respectivamente.

Siempre la muestra **a** corresponde a la serie de errores sin filtro del error modelo-mercado que corresponda en cada caso y la muestra **b** dependerá contra qué tipo de filtro evaluamos la muestra **a**.

A diferencia de la fórmula de la prueba t estándar, aquí se multiplica el estadístico t por un factor adicional que tiene en cuenta las diferencias en las varianzas y los tamaños de las muestras. Este factor se calcula utilizando los grados de libertad ajustados.

4.3 Medidas de dispersión de ajuste del modelo

Finalmente, es necesario llevar a cabo una exhaustiva evaluación de la volatilidad de las tasas que derivan del modelo de estructuras de tasas. El propósito de esta evaluación es determinar la concordancia entre la variabilidad exhibida en las curvas de tasas resultantes del modelo y la variabilidad intrínseca del entorno de mercado. Esta evaluación forma parte de los componentes esenciales para determinar la eficacia y la fidelidad de los modelos empleados en la reproducción de las dinámicas de las tasas de interés. Con el propósito de evaluar dicha fidelidad en relación a las volatilidades, se empleará el *test* estadístico de Levene (1960), una herramienta reconocida en el campo estadístico.

El *test* de Levene representa una herramienta estadística que se utiliza en la evaluación de la igualdad de varianzas entre uno o más grupos de datos. En la presente investigación, se llevará a cabo una aplicación específica del *test* en un enfoque de comparación por pares. Más precisamente, se analizarán los *spreads* de mercado en contraposición a los *spreads* generados por el modelo sin filtrar, así como los *spreads* de mercado en comparación con el modelo filtrado.

Así, en el contexto de la metodología empleada, establecemos como **hipótesis nula** que las varianzas poblacionales de los elementos en ambos grupos evaluados son equivalentes. Por otro lado, la **hipótesis alternativa** postula que existe una discrepancia significativa en las varianzas entre estos grupos. De esta manera, si el p-valor resultante de esta prueba es menor a 0,05 podemos afirmar con un 95% de confianza que las varianzas poblacionales de ambos grupos son diferentes. A continuación, definimos el estadístico del *test* y como se compone.

En primer lugar, definimos las observaciones que utilizamos para el *test*:

$Y_{it} = TIR_t^i - TIR_{t-1}^i$, donde $i \in \{1, 2\}$ sea 1 el grupo de tasas de mercado y 2 el grupo de tasas del modelo que corresponda, además, $t=1, 2, \dots, 19$ debido a que son 19 días para los que calculamos las diferencias de TIR.

Luego se calculan las distancias absolutas de los valores a las medias de sus respectivos grupos:

$Z_{it} = |Y_{it} - \bar{Y}_i|$, donde \bar{Y}_i es la media del “ i ” esimo grupo.

Además, definimos el estadístico de prueba W como:

$$W = \frac{(N - k)}{(k - 1)} \frac{\sum_{i=1}^k N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^k \sum_{t=1}^{N_i} (Z_{it} - Z_{i.})^2},$$

donde

k es el número de grupo a los que pertenecen las muestras, como mencionamos anteriormente, en nuestro caso $k=2$ (mercado y modelo),

N es el número total de valores en los 2 grupos, para nuestra investigación $N=38$ (2 grupos de 19 observaciones),

N_i es el número de observaciones en cada grupo (19 en ambos casos),

$Z_{..}$ es la media de Z_{it} definida como $Z_{..} = \frac{1}{N} \sum_{i=1}^k \sum_{t=1}^{N_i} Z_{it}$,

$Z_{i.}$ es la media de Z_{it} solo para el grupo i definida como $Z_{i.} = \frac{1}{N_i} \sum_{t=1}^{N_i} Z_{it}$.

Finalmente, se prueba la significancia de W contra $F(\alpha, k - 1, N - k)$ donde F es un cuantil de la distribución F de Fisher, con grados de libertad $k - 1$ y $N - k$, y α es el nivel de significación (para nuestro análisis, $\alpha = 0,05$).

5. Diseño del filtro y proceso de filtrado

Los precios de entrada al modelo de cálculo de tasas de interés tienen *outliers*. Entonces, como de ellos depende la precisión de los resultados del modelo, la intención del filtro diseñado es removerlos, sin embargo, la parte compleja recae en su detección. Debemos tener en cuenta la relevancia de un precio de entrada en un mercado ilíquido como el argentino donde no tenemos un precio de operación diario para todos los activos del mercado (Riesco y Cortazar, 2003). De esta manera, el filtro está diseñado para considerar “a priori” cada precio como válido, es decir que representa el valor de mercado del activo. Entonces, si queremos verlo estadísticamente como la clasificación de una observación en una categoría, podemos definir una hipótesis nula y alternativa:

H_0 : El precio ingresado es un precio de mercado

H_A : El precio ingresado es un *outlier*

Ya con ese planteo, para minimizar el error de tipo 1, es decir, rechazar la hipótesis de que el precio de entrada es de mercado y concluir incorrectamente que es un *outlier*, desarrollamos dos condiciones que debe cumplir la observación en análisis, para finalmente ser catalogada de *outlier* y removida de los datos con los que se arma la curva,

en otras palabras, ser filtrado. Eso sí, recordemos que el método de cálculo de curvas es el propuesto por Nelson y Siegel (1987). Las dos condiciones son las siguientes:

1. Presencia del último *spread* en *cluster* anómalo en serie temporal individual.
2. Presencia del último *spread* en *cluster* anómalo en corte transversal grupal de vecinos de curva.

No obstante, con el objetivo de evaluar la eficiencia de cada una de las condiciones y su desempeño dentro del filtro de dos condiciones diseñado, desarrollamos dos filtros adicionales: uno que incorpora **exclusivamente la primera condición** y otro que se basa **únicamente en la segunda condición**. Los resultados de los tres tipos de filtro (con dos condiciones, solo condición 1 y solo condición 2) se presentan detalladamente en la sección de resultados. Solo se filtran los precios que cumplan las dos condiciones, y, por lo tanto, no es necesario que un activo pase por ambas, ya que en cuanto no cumpla una de ellas no se evaluará en la siguiente, pues no se considera un *outlier* y se utiliza para el cálculo de la curva.

5.1 Proceso de filtrado

El proceso de filtrado se lleva a cabo de manera **secuencial** e **iterativa**, lo que implica que cada tipo de filtro se aplica sucesivamente a los datos de cada día del período de análisis, abarcando desde el 27 de marzo hasta el 1 de junio, ambos inclusive. En el procedimiento diario de filtrado, se comienza por aplicar el filtro correspondiente a ese día a los datos de entrada, seguido de la calibración de la curva específica para ese día, generando así la TIR modelo correspondiente.

Es importante destacar que, aunque se realiza la calibración de la curva en la condición 1, estos resultados son endógenos al proceso y se emplean exclusivamente para determinar qué precios avanzan a la condición 2, como se describe más adelante. Finalmente, se evalúan los resultados que cumplen con la condición 2, y con base en estos resultados se determinan los datos que se filtran. Luego de este proceso, se procede a calibrar los resultados de la TIR modelo finales para ese día, excluyendo los datos filtrados. Por esta razón, el proceso de filtrado debe llevarse a cabo de manera secuencial, avanzando al siguiente filtro solo después de completar el filtrado de todo el período utilizando un tipo de filtro específico. De esta manera, se logra calibrar la curva para la totalidad del período seleccionado con un tipo de filtro a la vez, y además sin aplicar ningún filtro. El orden de aplicación de los filtros es el siguiente:

1. Calibración de curvas sin ningún filtro.
2. Calibración de curvas con el filtro de dos condiciones.
3. Calibración de curvas con el filtro basado únicamente en la condición 1.
4. Calibración de curvas con el filtro basado únicamente en la condición 2.

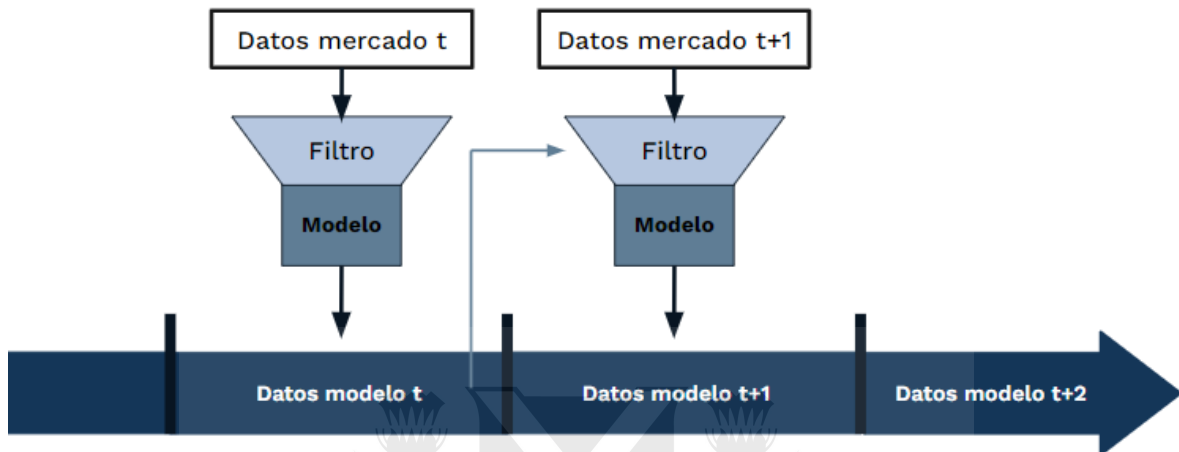


Gráfico 1.Proceso de filtrado

El Gráfico 1 muestra el proceso secuencial de filtrado para obtener los datos modelo (TIR) del día t los cuales luego serán usados en el día $t+1$ por la condición 1 la cual se describe a continuación.

5.2 Condición 1: presencia del último *spread* en *cluster* anómalo en serie temporal individual.

La primera condición se centra en la evaluación del *spread* del día de análisis, que es generado por la tasa modelo de ese día. En particular, determinamos el *spread* t de la siguiente manera:

$$Spread\ modelo_t^k = TIR\ modelo_t^k - TIR\ modelo_{t-1}^k,$$

donde $i= 1, 2, 3, \dots$

Siendo $Spread_t^k$ el *spread* del activo k para el período t .

El *spread* representa la diferencia entre la tasa modelo de ese activo en el día t y la del día anterior. Este indicador nos proporciona una perspectiva del movimiento diario de la tasa de ese activo específico en el contexto de un conjunto de activos comparables que conforman la curva de tasas.

Para cumplir con esta condición, el *spread* generado por el precio de entrada de ese día se evalúa en relación con el comportamiento histórico del activo durante los últimos 20

días. En otras palabras, comparamos el *spread* actual con el comportamiento pasado. Es fundamental destacar que el cálculo del *spread* para el día t requiere la TIR modelo correspondiente a ese día, la cual se obtiene mediante una calibración de curva utilizando el modelo de Nelson y Siegel. Esta calibración se realiza específicamente para esta condición y se utiliza exclusivamente para determinar si el movimiento de la tasa de interés de ese activo en el día t es anómalo en relación con sus movimientos previos en las TIR modelo.

En el proceso de *clustering*, hemos establecido los siguientes parámetros:

- **$k=3$ (número de *clusters*):** utilizamos tres *clusters* para clasificar los movimientos de tasas en categorías positivas, neutrales y negativas. Esto permite capturar las distintas dinámicas que pueden presentar las curvas de tasas, que pueden ser ascendentes, descendentes o mantenerse paralelas (Fabozzi, 2007).
- **$n=20$ (cantidad de *spreads* a *clusterizar*):** tomamos en consideración los *spreads* de los últimos 20 días como parte del estudio previo, basándonos que se a un periodo lo suficientemente grande para realizar un *clustering* apropiado guiándonos por los resultados de la condición de Hartigan (ver Apéndice 1).

El concepto de "*cluster* anómalo" se define como aquel cuyo centroide tiene la mayor distancia euclidiana con respecto al origen.

Este proceso de clasificación se aplica al precio de entrada que generó ese *spread*, y si cumple con esta condición, se procede a evaluarlo en la siguiente etapa del análisis.

Esta condición nos proporciona un primer filtro importante para identificar movimientos anómalos en las tasas de interés propias de un activo específico, lo que es fundamental para luego evaluar ese último movimiento en conjunto con sus comparables en la siguiente condición.

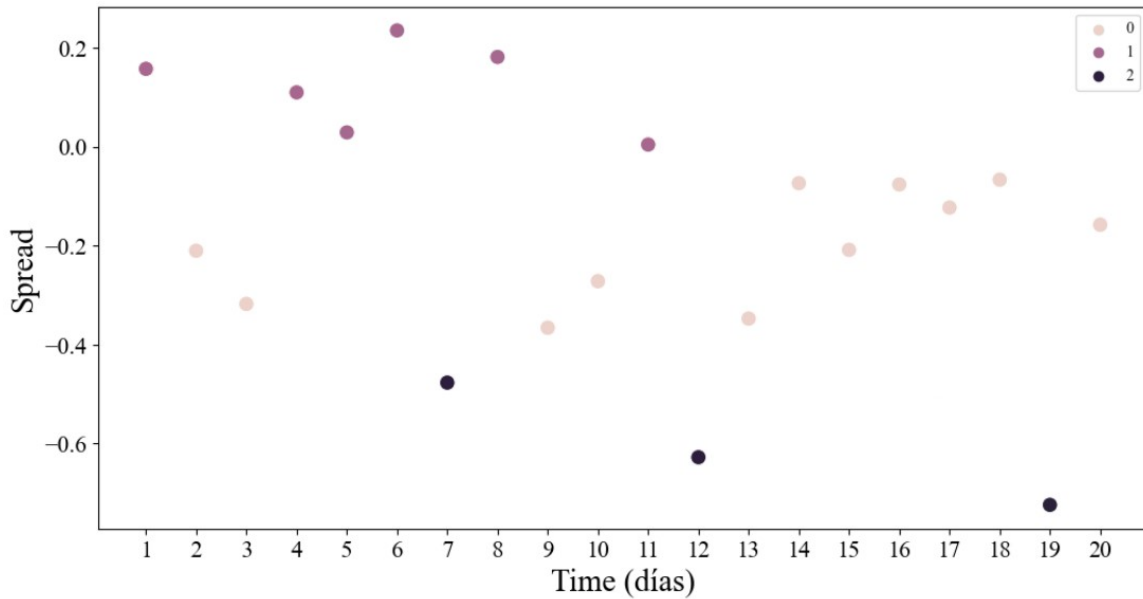


Gráfico 2. Agrupación de *spreads* del activo MAC2O por *clustering* para el 10 de abril de 2023.

En el Gráfico 2, que muestra los 20 *spreads* de un activo *dólar link* para el 10 de abril de 2023 en distintos colores, se observa que el *spread* más reciente (*spread t*), se encuentra en el cluster 0, pero el centroide de ese *cluster* no es el más lejos del origen, y, por lo tanto, no pasa la condición 1. Para cumplir con esta condición y avanzar a la siguiente evaluación, el *spread t* debería estar en el *cluster* 2, que es el más alejado del cero.

5.3 Condición 3: presencia del último *spread* en *cluster* anómalo en corte transversal grupal de vecinos de curva

En esta fase, consideramos todas las entradas de precios de mercado de activos “vecinos” en la misma curva de tasas del *dólar link*. Esta etapa de filtrado tiene como objetivo asegurarse de que la observación previamente identificada como anómala en su serie histórica de *spreads* no esté en concordancia con los movimientos de tasas de otros activos en el mismo mercado. Para lograr esto, primero definimos el concepto de “*spread* de mercado *t*” de la siguiente manera:

$$Spread\ mercado_t^k = TIR\ entrada_t^k - TIR\ modelo_{t-1}^k,$$

El $Spread\ mercado_t^k$ mide cómo se han movido las TIR de entrada actual en comparación con las tasas del activo k el día t respecto al día anterior. Esta medida nos proporciona información sobre el movimiento de la tasa de interés para ese activo en todos los días en los que se genera una TIR de mercado. Es importante destacar que la

TIR de mercado no está disponible todos los días debido a la falta de liquidez en el mercado (Riesco y Cortázar, 2003).

El proceso de *clustering* en esta fase se basa en dos especificaciones fundamentales. En primer lugar, los parámetros utilizados son los siguientes:

- ***n***: la cantidad de *spreads* de mercado que se *clusterizan*, y esta cantidad depende de las entradas de precios del día. Durante el período analizado, se observó un promedio de 7.8 entradas de precio, lo que permitió un *clustering* adecuado basado en la condición de Hartigan (ver Apéndice 1).
- ***k=2* (número de *clusters*)**: utilizamos dos *clusters*. Hemos elegido dos *clusters* para representar movimientos normales y anómalos. Dada la limitada disponibilidad de datos y la necesidad de mantener la precisión, consideramos que dos *clusters* son apropiados para el mercado argentino durante el período analizado.

Un aspecto esencial en la determinación de la "anomalía de un *cluster*" se basa en la distancia del origen y la cantidad de activos en el *cluster*. En otras palabras, un *cluster* que se aleje más del origen se considera más anómalo. Sin embargo, si todos los activos se mueven de la misma manera, excepto uno, esto podría indicar que el comportamiento es común en la curva y el activo que no se mueve sería el atípico. Por lo tanto, hemos desarrollado la siguiente medida de "anormalidad" para un *cluster*:

$$\text{Medida de anormalidad para el cluster } k = d_k * \frac{N}{Q_k},$$

donde:

- *k* es el *cluster*,
- d_k es la distancia euclidiana desde el origen hasta el centroide del *cluster* *k*,
- *N* es la cantidad total de *spreads* de mercado del día *t* (es decir, el número total de observaciones que se *clusterizan*), y
- Q_k es la cantidad de observaciones que el *clustering* determina que pertenecen al *cluster* *k*.

Es importante destacar que el *cluster* que se considera anormal es aquel que posee la mayor medida descrita. En este contexto, es relevante tener en cuenta que una mayor distancia desde el centroide y una menor cantidad de observaciones dentro de ese *cluster* tienen un impacto positivo en la medida de anormalidad.

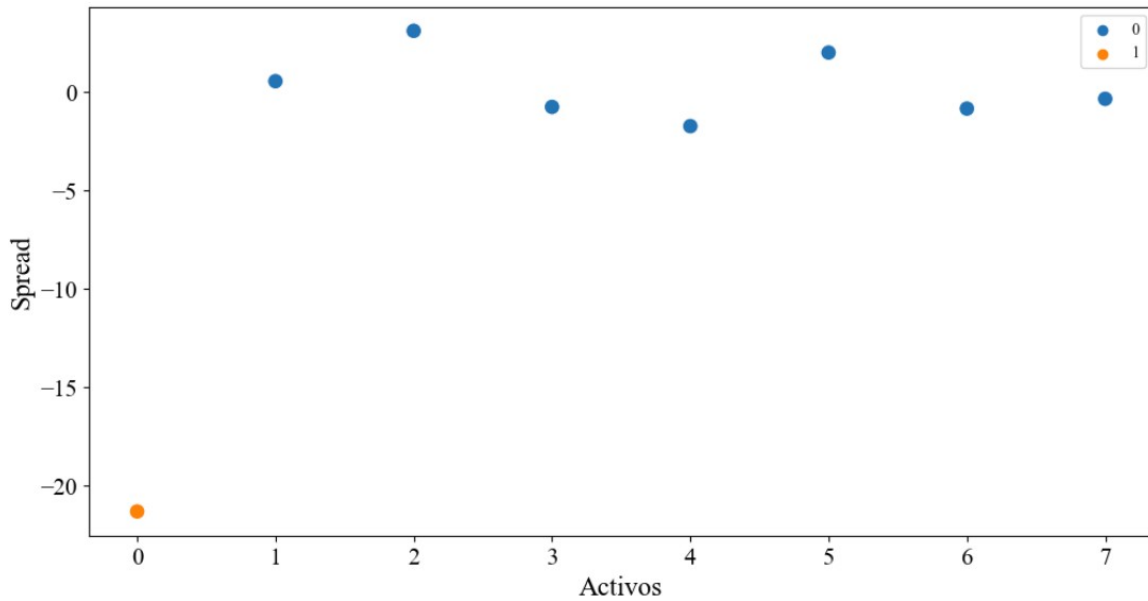


Gráfico 3. *Spreads* de activos agrupados por *clustering* para el 10 de abril de 2023.

En el Gráfico 3, presentamos la agrupación de todos los activos de la curva del *dólar link* que tuvieron entradas para el 10 de abril de 2023. El punto naranja en el gráfico representa un *spread* considerado anómalo debido a su inclusión en el *cluster* con la mayor medida de anormalidad.

En consecuencia, después de esta fase de análisis, el *spread* t del activo k se clasifica finalmente como *outlier* y se excluye de la base de datos de precios de entrada.

6. Resultados

6.1 Errores modelo-mercado

Para evaluar la efectividad del filtro que hemos diseñado, hemos llevado a cabo cuatro series de calibraciones distintas. Estas series incluyen:

1. Sin filtro: en esta serie de calibraciones, no se aplicó ningún tipo de filtro al proceso.
2. Utilizando el filtro con dos condiciones: en esta serie de calibraciones, se utilizó el filtro con ambas condiciones simultáneamente, tal como explicamos en la sección de diseño.
3. Filtrando utilizando solo la primera condición: en esta serie de calibraciones, aplicamos únicamente la primera condición del filtro.
4. Filtrando utilizando solo la segunda condición: en esta serie de calibraciones, aplicamos únicamente la segunda condición del filtro.

También realizamos análisis por separado (serie tres y cuatro) utilizando cada una de las condiciones de manera individual, ya que nos permite evaluar la efectividad de ambas condiciones y comprender cómo funcionan dentro del contexto del filtro.

En primer lugar, nuestro enfoque se centra en el análisis de los resultados relacionados con los errores del modelo en relación con el mercado, ya que esta métrica es determinante en nuestro análisis para evaluar el ajuste de las calibraciones. Con el propósito de cuantificar el impacto de las observaciones que han sido sometidas al proceso de filtrado, hemos llevado a cabo dos cálculos que arrojan el promedio de los errores modelo-mercado.

En primer lugar, el cálculo presentado en la Tabla 3 considera todos los precios de entrada diarios que se habrían obtenido si no se hubiera aplicado el filtro. En esta tabla, comparamos los precios de entrada de la serie 1 con los modelos correspondientes de cada serie.

Por otro lado, el segundo cálculo se encuentra en la Tabla 4 y se centra en los precios de entrada de cada serie que no han sido sometidos al proceso de filtrado. En este caso, comparamos los precios de entrada no filtrados de cada serie con los modelos obtenidos en esa misma serie.

En ambas tablas se muestra el promedio del cálculo de los errores diarios para el período de análisis (44 días).

Modelo de Nelson y Siegel

Modelo	MAE	MAPE	RMSE	RMSPE
Normal	10,5791	4,0833	13,0899	8,1456
Utilizando 2 condiciones	10,6382	4,1482	13,5890	8,5290
Utilizando solo condición 1	18,0764	7,2920	21,3661	13,6127
Utilizando solo condición 2	10,4316	4,0974	13,7416	8,6614

Tabla 2. Promedios de errores diarios calculados con todas las entradas de precios.

Modelo de Nelson y Siegel

Modelo	MAE	MAPE	RMSE	RMSPE
Normal	10,5791	4,0832	13,0899	8,1456
Utilizando 2 condiciones	9,8203	3,7711	12,1865	7,5600
Utilizando solo condición 1	12,4219	4,4830	14,2961	8,5881
Utilizando solo condición 2	9,0126	3,4503	11,7605	7,2899

Tabla 3. Promedios de errores diarios calculados solo con precios no filtrados.

Al analizar los datos presentados en la Tabla 3, podemos inferir que al calcular los errores teniendo en cuenta todas las entradas de precios, no se evidencia una mejora significativa en el ajuste del modelo en términos de errores modelo-mercado al aplicar el filtro con dos condiciones. Aunque las diferencias no son considerablemente marcadas, el uso del filtro de dos condiciones no parece conducir a una mejora sustancial en comparación con la ausencia de filtro.

Además, se observa que el uso exclusivo de la condición 1 resulta en un ajuste notablemente inferior, con errores significativamente mayores. En contraste, la exclusiva aplicación de la condición 2 apenas difiere en el promedio de errores obtenido al emplear ambas condiciones simultáneamente. Específicamente, en métricas como MAE y MAPE, se registran valores más bajos al usar solo la condición 2 en comparación con el uso de ambas condiciones, e incluso en algunos casos, los valores son inferiores a los obtenidos sin aplicar ninguna condición.

Cabe mencionar que, durante el análisis, sometimos las series de errores a pruebas de significación de medias con un nivel de confianza del 5%. En base a ese *test* establecimos que el filtro de solo la condición 1 en relación al caso sin filtro es significativamente mayor en todos los errores. Para el resto de los filtros (2 condiciones y solo condición 2), no hemos logrado establecer que la utilización de al menos una de las condiciones de filtrado conduzca a series de errores cuyas medias sean estadísticamente distintas de aquellas que resultan de no aplicar ningún filtro. Los detalles y p-valores de estas pruebas se encuentran en el Apéndice 2.

No obstante, es relevante destacar una observación interesante que se desprende del análisis visual. Al examinar el Gráfico 4 y 5, se aprecian períodos en los cuales, ya sea al considerar el error calculado con todos los precios o únicamente con aquellos que no fueron objeto de filtrado, se registra una disminución notoria en los errores. Esta observación implica que, si bien las pruebas de significación no revelan diferencias estadísticas significativas, hay momentos en los cuales la aplicación del filtro parece generar un impacto positivo en la precisión de los resultados.

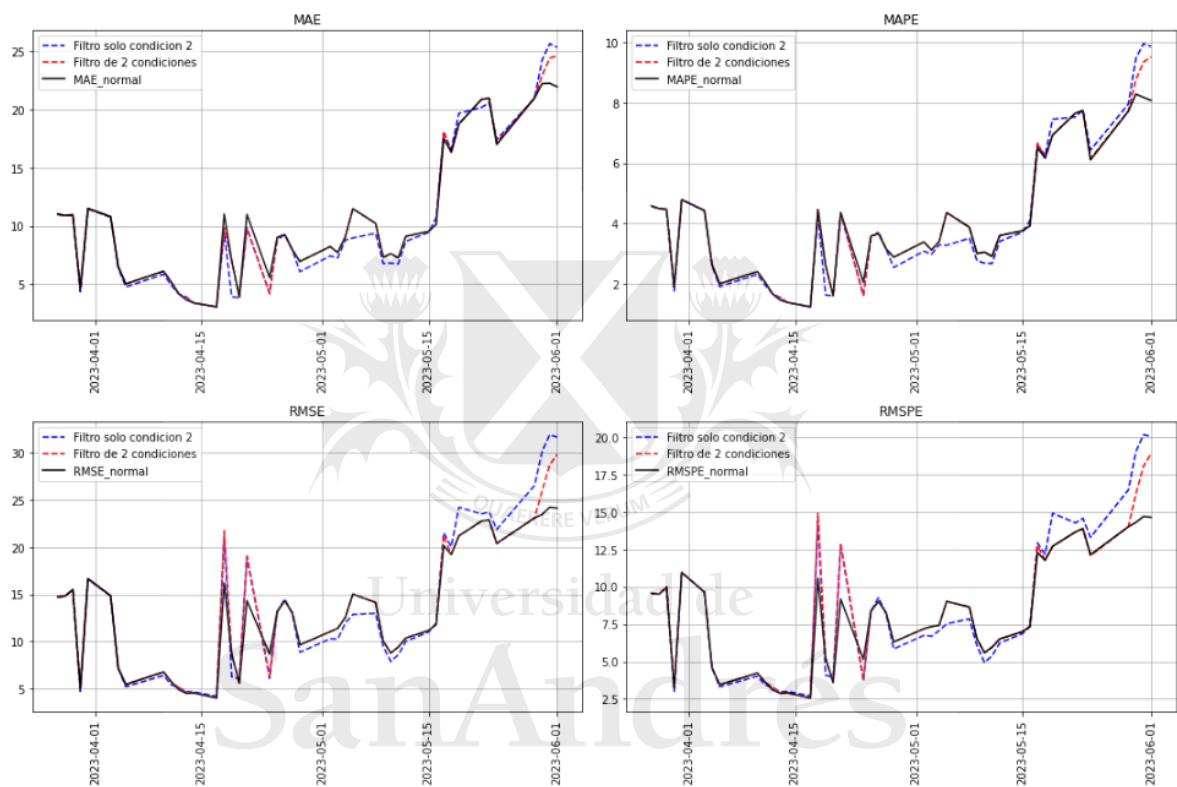


Gráfico 4. Historial de errores modelo-mercado calculados con todas las entradas de precios.

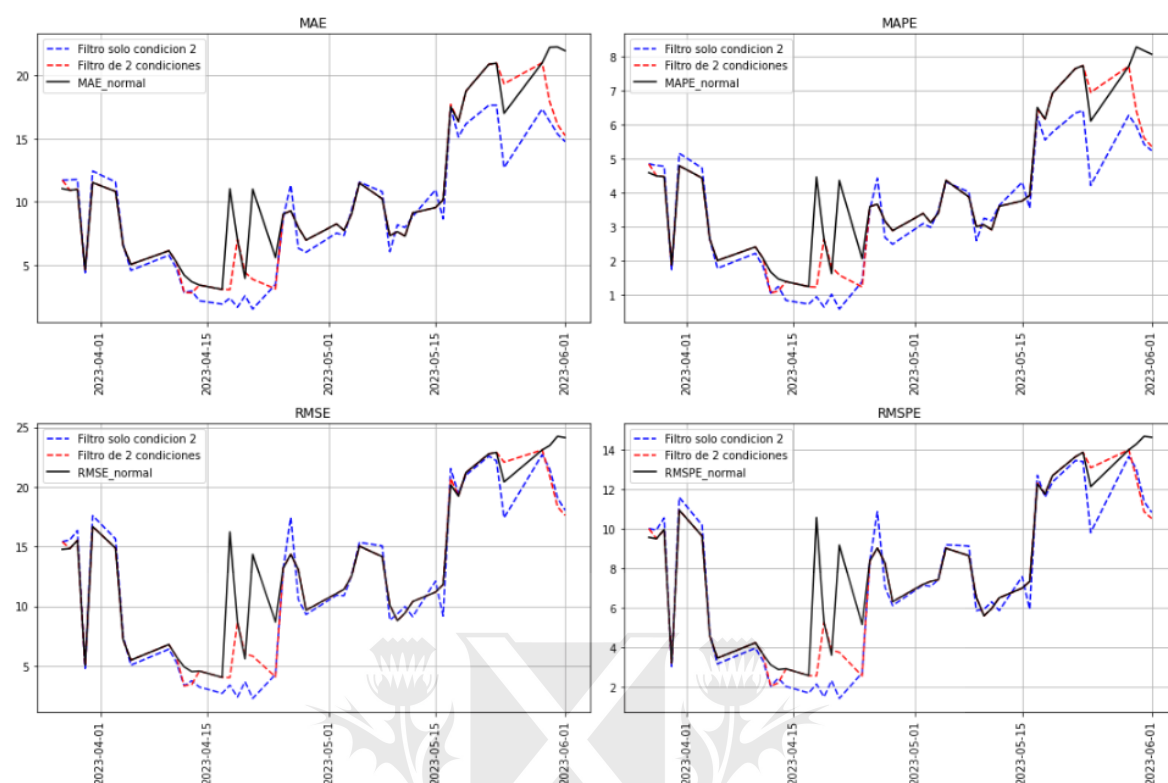


Gráfico 5. Historial de errores modelo-mercado calculados solo con precios no filtrados.

Al examinar detenidamente el Gráfico 5, notamos que los errores modelo-mercado asociados al filtro de dos condiciones siguen un patrón de movimientos que se asemeja al escenario en el cual no se aplica ningún tipo de filtro. Este fenómeno se explica en parte debido a la escasa cantidad de datos que se eliminan mediante el proceso de filtrado, los cuales presentamos en el Gráfico 7. Sin embargo, resulta de particular interés que, a partir del 29 de mayo, se evidencia un aumento considerable en los errores, lo que sugiere un posible desajuste en las calibraciones en ese período específico.

En paralelo, hemos evaluado en el mismo Gráfico 4 el rendimiento del filtro aplicando únicamente la condición 2. En la mayoría de los períodos analizados, se observa que los errores obtenidos son menores en comparación con el uso del filtro de dos condiciones y sin ningún filtro. No obstante, es relevante destacar que, de manera similar al filtro de dos condiciones, también en el 29 de mayo se registran medidas de errores notablemente superiores en comparación con las calibraciones previas del filtro, e incluso superiores al filtro de dos condiciones.

Estos hallazgos subrayan la importancia de analizar el aumento de errores observado en el 29 de mayo, ya que podría ser fundamental para comprender mejor el impacto del filtro la calibración de curvas. Con el objetivo de abordar esta cuestión de manera más

detallada, hemos procedido a recalcular los mismos valores que se presentaron en la Tabla 2, pero esta vez limitando nuestros datos a aquellos previos al 29 de mayo. Los resultados de esta evaluación se exponen en la Tabla 4 a continuación.

Modelo de Nelson y Siegel

Modelo	MAE	MAPE	RMSE	RMSPE
Normal	9,4507	3,6854	12,0257	7,5190
Utilizando 2 condiciones	9,3775	3,6793	12,2573	7,7030
Utilizando solo condición 1	17,6977	7,2150	21,1295	13,5328
Utilizando solo condición 2	9,0686	3,5763	12,1098	7,6350

Tabla 4. Promedios de errores diarios previos al 29 de mayo del 2023 calculados con todas las entradas de precios.

La Tabla 4 revela que, en general, los errores disminuyen en comparación con la Tabla 2. Es importante destacar que el MAE y el MAPE son más bajos cuando se utiliza el filtro de dos condiciones que cuando no se usa ningún filtro, y son aún menores cuando se aplica solo la condición dos. Por otro lado, el RMSE y el RMSPE siguen siendo ligeramente más bajos en el escenario sin filtro, pero son apenas inferiores al utilizar solo la condición dos del filtro. Estos resultados respaldan la utilidad del filtrado, especialmente la condición dos, para mejorar la precisión del modelo en términos de MAE y MAPE.

6.2 Datos cualitativos y estadísticas porcentuales del filtrado

Ahora bien, para obtener una comprensión más profunda de los resultados relativos a los errores, es esencial analizar las TIR de entrada y las TIR modelo resultantes después de la calibración de la curva. Para llevar a cabo este análisis, presentamos los siguientes cuatro gráficos:

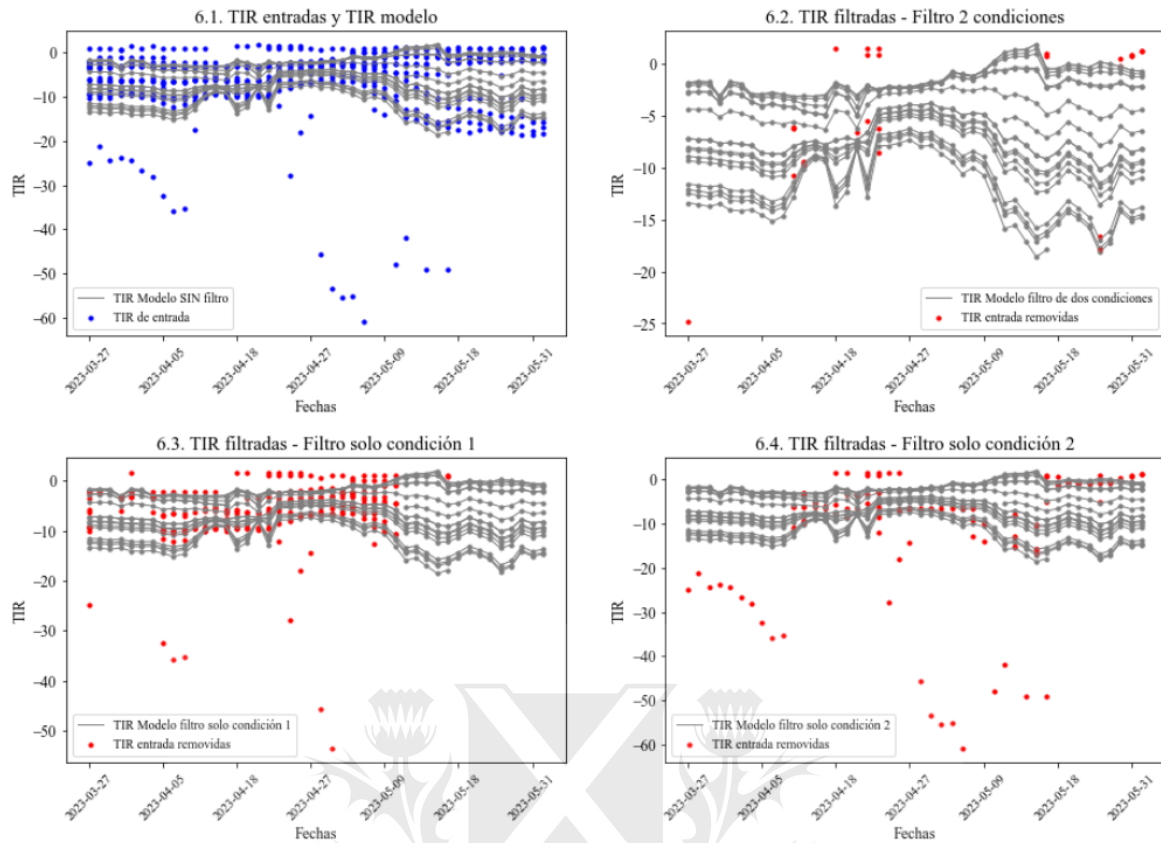


Gráfico 6. Historial de TIR modelo, TIR entradas y TIR filtradas con cada filtro.

Los cuatro subgráficos que componen el Gráfico 6 muestran las TIR modelo en gris cuando no se aplica filtrado a los precios de entrada. En el Gráfico 6.1, se representan en azul todas las TIR de entrada utilizadas en la calibración sin filtro. Luego, en los Gráficos 6.2, 6.3 y 6.4, en rojo, resaltamos las TIR de entrada específicas (tomadas del Gráfico 6.1) que fueron eliminadas por los diferentes filtros: el filtro de dos condiciones en el Gráfico 6.2, el filtro basado en la condición 1 en el Gráfico 6.3 y el filtro de solo condición 2 en el Gráfico 6.4. Estos gráficos ofrecen una representación visual concisa de cuántas y cuáles entradas se incluyen o excluyen en la calibración de las curvas. Al observar los resultados de manera gráfica, se puede apreciar que el filtro de dos condiciones tiene un impacto relativamente limitado en la reducción de datos. Es notorio que la condición 1 es la principal influencia en el proceso de filtrado, seguida por la condición 2. Este fenómeno intermedio se confirma a través del análisis visual presentado en el Gráfico 7, donde se representa el promedio diario de las cantidades filtradas en cada uno de los escenarios evaluados.

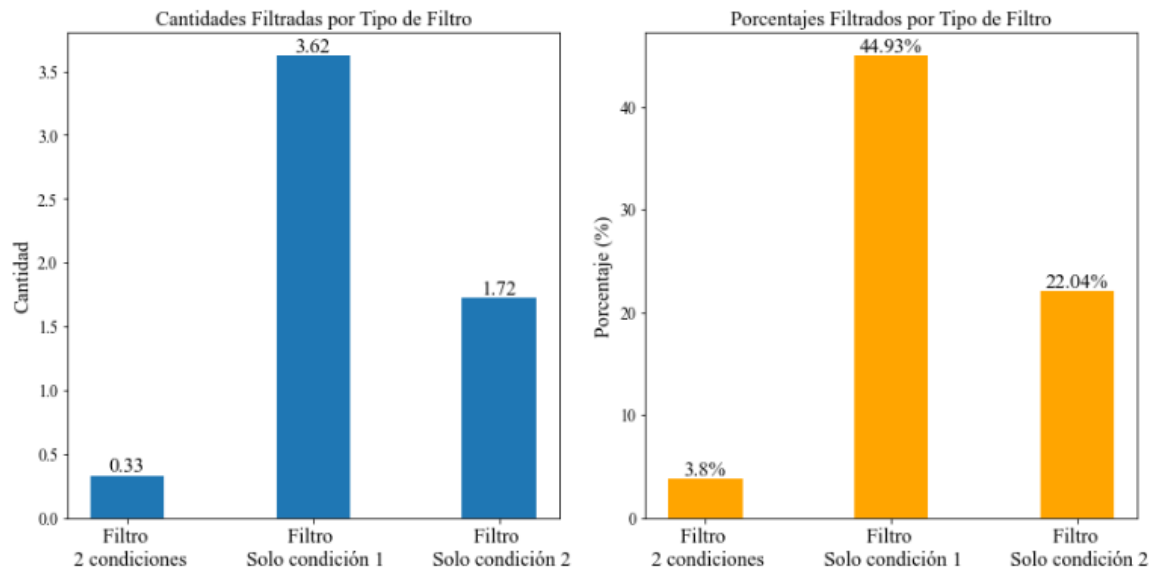


Gráfico 7. Cantidades y porcentajes diarios promedio para cada filtro.

Una vez que analizamos la los errores del modelo en cada filtro y las cantidades entendemos que en caso que consideremos el cálculo de errores utilizando todas las entradas podemos entender mejor el funcionamiento del filtro. En cuanto al filtro basado únicamente en la condición 1, se nota que esta condición, debido a su naturaleza en el proceso de *clustering*, tiende a generar un agrupamiento significativo alrededor del origen, clasificando todo lo que se aleja de este punto como anómalo y, por lo tanto, sujeto a filtrado. Se proporciona un análisis más detallado en el Apéndice 3, lo que justifica nuestra decisión de no continuar evaluando este escenario de filtrado en particular.

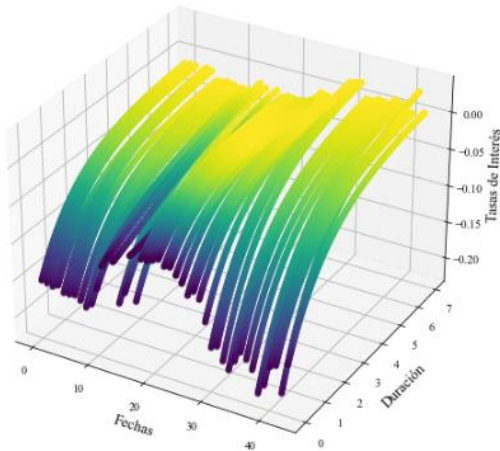
En relación al filtro basado únicamente en la condición 2, aplicando el mismo razonamiento en cuanto al cálculo del error, no se observa una mejora en el error en comparación con el caso sin filtrado. Sin embargo, los resultados sugieren que este filtro podría preservar de manera más efectiva la información relevante.

Cuando evaluamos el error modelo-mercado utilizando únicamente los valores que no fueron filtrados ni clasificados como valores anómalos, se encuentra que el mejor ajuste se obtiene con el filtro basado únicamente en la condición 2.

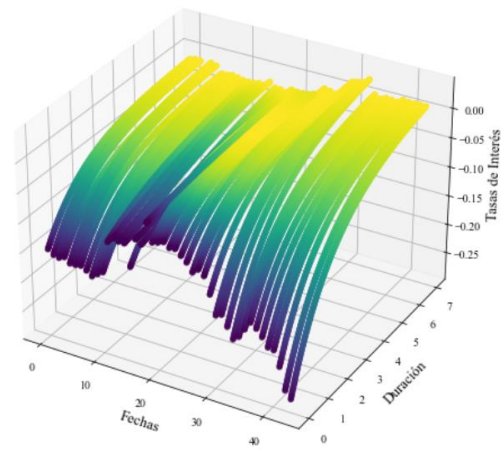
6.3 Movimiento de curvas en el período analizado

Además, es esencial llevar a cabo un análisis de la modificación en la forma de la curva resultante tras la implementación del filtro. Para abordar esta evaluación, realizamos el Gráfico 8 el cual se divide en 4 subgráficos. Estos gráficos permiten observar la evolución temporal de las curvas teóricas bajo la influencia de cada tipo de filtro, a partir de las cuales se derivan las TIR modelo para cada activo.

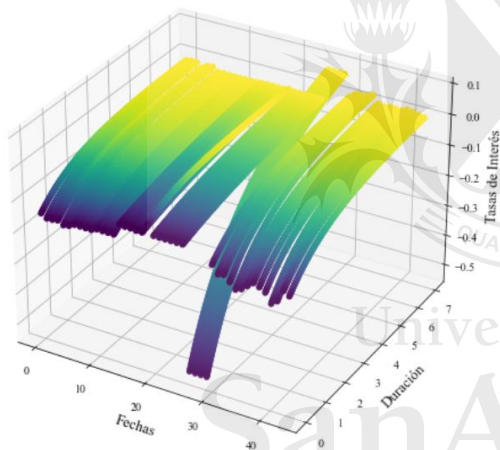
8.1 Curvas sin filtro



8.2 Curvas filtro 2 condiciones



8.3 Curvas filtro solo condición 1



8.4 Curvas filtro solo condición 2

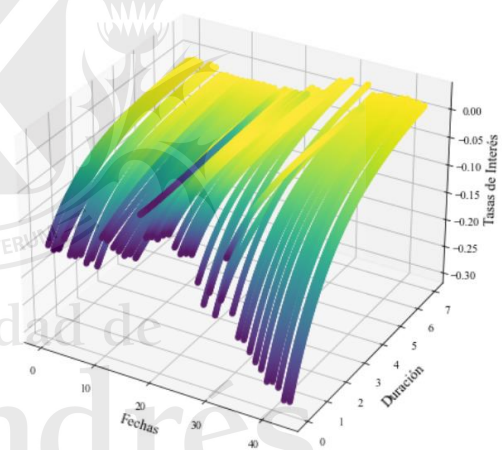


Gráfico 8. Historial de curvas de tasas de interés a partir de las diferentes calibraciones.

En el Gráfico 8, se observa que las curvas resultantes de la calibración del modelo Nelson y Siegel, bajo el filtro de dos condiciones en comparación con la calibración normal, son ligeramente más planas, pero no muestran cambios significativos. El filtro basado en la condición 1 selecciona un nivel de curva cercano a cero y filtra desviaciones que no sean acordes a ese valor. Esto resulta en una curva mayormente plana en la mayoría de los períodos, que no se ajusta adecuadamente a las demás tasas de interés, excepto en un período donde se encuentra una agrupación "normal" con una curva más empinada.

Por último, al examinar la curva generada por el filtro basado únicamente en la condición 2, se observa que es más volátil en comparación con la calibración normal sin filtro. Además, esta curva tiende a posicionar las partes de menor duración en niveles más altos.

Esta mayor volatilidad se confirma en la Tabla 5, donde se calcula el desviación estándar diaria de los parámetros diarios utilizados para el cálculo de las curvas.

Volatilidad de los parámetros

Modelo	Beta 0	Beta 1	Beta 2	Tau
Normal	2,55%	5,96%	5,94%	Fijo
Utilizando 2 condiciones	2,74%	7,21%	7,30%	Fijo
Utilizando solo condición 1	6,67%	17,33%	7,48%	Fijo
Utilizando solo condición 2	3,24%	9,52%	6,47%	Fijo

Tabla 5. Desvío estándar diario de parámetros del modelo Nelson y Siegel.

Acorde a lo expuesto en el Gráfico 8, podemos observar en la Tabla 5 que el menor desvío estándar se registra en el caso de las calibraciones normales sin filtro, y aunque hay un ligero aumento en el desvío estándar con el filtro de dos condiciones, este cambio no es significativo. En contraste, al aplicar el filtro basado únicamente en la condición 1, se produce un aumento sustancial en la volatilidad debido a los cambios abruptos en las curvas durante un período específico. Asimismo, el filtro basado únicamente en la condición 2 también muestra una mayor volatilidad en comparación con la ausencia de filtro.

Sin embargo, es importante destacar que, al realizar el *test* de Levene, descrito en la sección de metodología, para comparar las volatilidades de los parámetros entre los diferentes casos de filtro y el caso sin filtro, no se encontró evidencia estadística a un nivel de significación del 5% que respalde la diferencia en las series diarias de cada parámetro tienen una volatilidad diferente, con la excepción del caso del coeficiente beta 2 en el filtro de solo condición 1 (véase Apéndice 4).

7. Conclusión

Tras examinar detenidamente las entradas de precios y las calibraciones de las curvas mediante el modelo de Nelson y Siegel, hemos podido identificar visualmente ciertas observaciones que podrían considerarse como valores atípicos (*outliers*), los cuales deterioran la calidad de la calibración del filtro. En respuesta a esta problemática,

propusimos un filtro diseñado con dos condiciones que determinan de manera endógena cuáles activos deben ser considerados como *outliers* y, por lo tanto, filtrados. Estas dos condiciones consisten en un análisis de *clustering* sobre un recorte temporal de los *spreads* de un activo en particular (condición 1) y un enfoque transversal que evalúa los movimientos de tasas entre las entradas de precios y el modelo del día anterior (condición 2).

El filtro que se propone en este trabajo fue sometido a pruebas durante el período seleccionado, y también se comparó su desempeño con otros dos filtros: uno únicamente con la condición 1 y otro únicamente con la condición 2. Estas pruebas no solo nos permitieron evaluar la eficiencia del filtro diseñado, sino que también arrojaron luz sobre su funcionamiento.

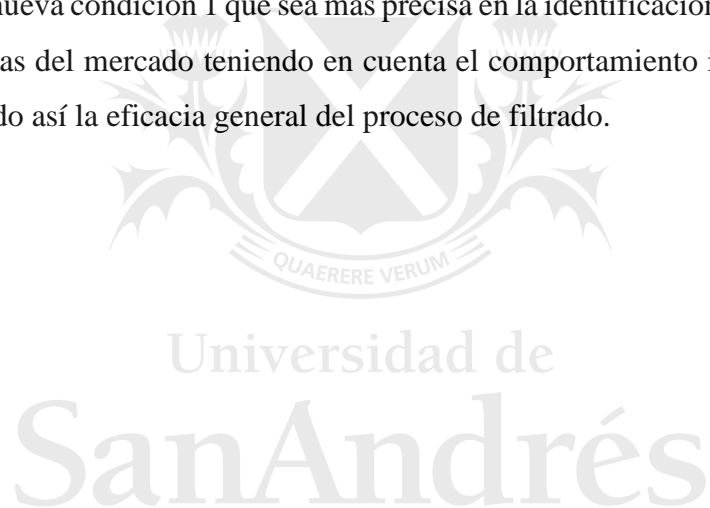
Los resultados obtenidos fueron sumamente esclarecedores. En primer lugar, observamos que el filtro de dos condiciones propuesto inicialmente no generó cambios significativos en la calibración de las curvas, y además aumentó ligeramente la volatilidad (aunque no significativamente estadístico). Esto indica que el filtro concebido a partir de ambas condiciones, en su diseño actual, no resulta suficiente para depurar las entradas de precios que tienen un impacto adverso en la calibración. Por otro lado, el filtro que se fundamenta únicamente en la condición 1 conduce a una calibración de curva notablemente menos precisa al excluir todos los activos que presentan desviaciones significativas respecto al valor cero. Aunque este último filtro ya no se considera una opción viable y no será implementado, su análisis ayudó a comprender que la condición 1, por sí sola, no conduce a mejoras en la calibración de las curvas de tasas de interés.

En segundo lugar, el filtro basado únicamente en la condición 2 mostró un desempeño superior al filtro de dos condiciones, pero aún inferior a no utilizar filtro alguno. Sin embargo, al calcular los errores de ajuste solo con las entradas consideradas como precios de mercado, la situación cambió significativamente. En este contexto, el filtro basado únicamente en la condición 2 demostró ser la mejor opción y el de dos condiciones mejor al caso sin filtro.

Es importante destacar que la utilidad de este filtro dependerá de las entradas de precios disponibles en un momento dado, como lo ilustra la reducción de errores de ajuste al calcularlos para el periodo anterior al 29 de mayo. Esto sugiere que la mejor alternativa es el filtro de dos condiciones, especialmente si consideramos que algunos precios no representativos del mercado son clasificados como *outliers* por el filtro.

No obstante, es fundamental reconocer una limitación importante de este enfoque: su tendencia a realizar clasificaciones incorrectas. Debido a su naturaleza, el filtro siempre agrupa las observaciones en un grupo normal y otro anómalo, sin considerar la posibilidad de que todas las entradas de precios sean normales. Esto puede generar un ajuste deficiente, como se observó en los datos posteriores al 29 de mayo, donde las entradas de precios parecían estar en gran medida en concordancia con los valores de mercado.

Como posible línea de investigación futura, se plantea la exploración de métodos paramétricos de filtrado, como la aplicación de medidas basadas en desviaciones estándar. Esto podría mejorar la detección de *outliers*, especialmente cuando son excepcionalmente prominentes y tienen un impacto significativo en el algoritmo de filtrado, lo que podría resultar en la omisión de precios anómalos. Además, en vista de las limitaciones identificadas en la primera condición del filtro, se sugiere investigar y desarrollar una nueva condición 1 que sea más precisa en la identificación de tasas modelo no representativas del mercado teniendo en cuenta el comportamiento individual de ese activo, mejorando así la eficacia general del proceso de filtrado.



8. Bibliografía

- Alberti, S. (2019).** *Estimación de la estructura temporal de tasas de interés argentina mediante el modelo de Nelson y Siegel* [Tesis de Maestría en Finanzas] Universidad de San Andrés.
- Cai, F., Le-Khac, N. A., & Kechadi, T. (2016).** “Clustering approaches for financial data analysis: a survey”. *arXiv preprint arXiv:1609.08520*.
- Corso, E. A., & Matarrelli, C. (2019).** *Nota Técnica N 3/2019*. Expectativas de Inflación Implícitas en la Curva de Rendimientos. Argentina 2017-2018.
- De Battista, G. (2018).** *Estimación de la Estructura Temporal de Tasas de Interés Argentina: Curva Soberana Ley Nueva York* [Tesis de Maestría en Finanzas] Universidad de San Andrés.
- Espinosa, J. L. M. (2005).** *Valorización de instrumentos financieros en mercados con pocas transacciones: análisis de una metodología basada en un modelo dinámico para la tasa cero real en Chile* [Doctoral dissertation]. Pontificia Universidad Católica de Chile.
- Fabozzi, F. (2007).** *Fixed Income Analysis*. John Wiley & Sons Inc.
- García Riesco, J. (2003).** *Estructuras de tasas de interés nominales y reales en Chile: estimación de modelos estáticos y de modelos dinámicos mediante filtro de Kalman aplicado sobre paneles de datos incompletos* [Tesis de Magíster en Ciencias de la Ingeniería] Pontificia Universidad Católica de Chile.
- Gençay, R., Dacorogna, M., Muller, U. A., Pictet, O., y Olsen, R. (2001).** *An introduction to high-frequency finance*. Elsevier.
- Hartigan, J.A. (1985).** Statistical theory in clustering. *Journal of Classification*, 2, 63–76. <https://doi.org/10.1007/BF01908064>

- Levene, H. (1960).** Robust tests for equality of variances. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Stanford University Press*, 278-292.
- Macaulay, F. (1938).** *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856.* National Bureau of Economic Research.
- Mc Culloch, H. M. (1971).** Measuring the Term Structure of Interest Rate. *Journal of Business*, 44(1), 19-31. <http://dx.doi.org/10.1086/295329>
- Mc Culloch, H. M. (1975).** The Tax-Adjusted Yield Curve. *The Journal of Finance*, 30(3), 811-830. <https://doi.org/10.2307/2326860>
- Melo-Velandia, L. F., Arango, L. y Vásquez, D. (2003).** Estimación de la estructura a plazo de las tasas de interés de Colombia. *Coyuntura económica*, 33(1), 51-76. <http://hdl.handle.net/11445/2095>
- Nelson, C. y Siegel, A. (1987).** Parsimonious Modeling of Yield Curves. *Journal of Business*, 60(4), 473-489. <http://dx.doi.org/10.1086/296409>
- Omanovic, S., Avdagic, Z. y Konjicija, S. (2009).** On-line evolving clustering for financial statements' anomalies detection. *2009 XXII International Symposium on Information, Communication and Automation Technologies*, 1-4. <https://doi.org/10.1109/ICAT.2009.5348416>
- Pérez, J., Hidalgo, M., Castro, N., Pazos, R., Díaz, O., Olivares, V., y Almanza, N. (2018).** Una heurística eficiente aplicada al algoritmo K-means para el agrupamiento de grandes instancias altamente agrupadas. *Computación y Sistemas*, 22(2), 607-619. <https://doi.org/10.13053/cys-22-2-2546>
- Pérez, J., Pires, C., Balby, L., Mexicano, A. y Hidalgo, M. (2013).** Early classification: A new heuristic to improve the classification step of k-means. *Journal of*

Information and Data Management, 4(2), 94-94.

<https://doi.org/10.5753/jidm.2013.1482>

Puigvert Gutiérrez, J. M., y Fortiana Gregori, J. (2008). Clustering techniques

applied to outlier detection of financial market series using a moving window

filtering algorithm. *SSRN Electronic Journal*.

<https://doi.org/10.2139/ssrn.1275842>

Rodríguez , A. y Villavicencio, J. (2002). La formación de la curva de rendimientos en

nuevos soles de Perú. *Revista del Departamento de Economía Pontificia*

Universidad Católica del Perú, 25 (50), 173-204.

<https://doi.org/10.18800/economia.200202.005>

Svensson, L. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-

1994. *National Bureau of Economic Research, Working Paper N° 4871*.

Tang, G., Tian, R. y Wu, B. (2022). An Overview of Clustering Methods in The

Financial World. *Proceedings of the 2022 7th International Conference on*

Financial Innovation and Economic Development (ICFIED 2022), 22, 524-529.

[10.2991/aebmr.k.220307.084](https://doi.org/10.2991/aebmr.k.220307.084)

Welch, B. L. (1947). The generalization of "Student's" problem when several different

population variances are involved. *Biometrika*, 34 (1-2), 28-

35. doi:10.1093/biomet/34.1-2.28. MR 0019277. PMID 20287819.

9. Apéndices

9.1 Apéndice 1. Condición de Hartigan

Podemos aplicar la teoría de Hartigan para seleccionar un número de clúster evaluando la suma de los errores cuadráticos con la siguiente fórmula:

$$\left(\frac{\sum_{i=1}^k ESS}{\sum_{i=1}^{k+1} ESS} - 1 \right) * (n - k - 1) > 10$$

donde k es la cantidad de *clusters* (tres en nuestro caso), ESS representa la suma de los errores cuadráticos (ya sea con k o $k+1$ *clusters*), finalmente, n es la cantidad de datos, en nuestro caso utilizamos $n=17$. De este modo, calculamos esta condición que relaciona la cantidad de *clusters* (en nuestro caso tres y cuatro) y cantidad observaciones para todos los activos que se evaluaron en la condición dos, ya que en ella se utiliza en algoritmo de *k-means*. Los resultados obtenidos se muestran a continuación en la Tabla 6.

Fecha	Valores de condición de Hartigan para cada activo											Promedio
27-Mar-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCFO	TLCDO	TLCFO	20,34
	16,38	13,85	19,86	25,75	25,71	28,25	21,28	16,14	22,35	19,83	18,17	
28-Mar-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PN9CO	PNICO	PNJCO	TLC90	TLCFO	20,34
	22,04	14,39	22,08	22,27	22,22	25,01	9,53	31,09	21,96	15,00	18,17	
29-Mar-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCFO			23,17
	25,38	15,55	24,00	28,16	28,11	28,28	25,12	15,75	18,19			
30-Mar-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90				21,33
	25,35	14,49	21,41	26,32	26,34	15,15	25,12	16,49				
31-Mar-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCFO			32,17
	45,28	17,71	37,46	24,63	24,50	27,03	37,72	22,47	52,72			
3-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCFO			32,43
	42,77	18,64	33,77	29,68	29,59	26,45	41,31	22,27	47,41			
4-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCDO			24,95
	37,01	22,21	28,14	26,46	26,44	19,94	23,34	25,53	15,45			
5-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90				22,27
	27,61	21,83	26,55	17,98	17,97	20,44	21,50	24,28				
10-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90				24,95
	33,96	19,71	27,92	24,00	24,02	22,03	23,08	24,89				
11-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90				21,94
	30,58	17,43	24,13	19,27	19,25	20,88	25,30	18,64				
12-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90				21,30
	18,25	16,28	12,91	30,27	30,07	16,85	29,37	16,38				
13-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90				20,65
	23,24	17,93	13,77	23,93	23,77	23,49	23,68	15,42				
14-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90				18,24

	17,61	15,59	15,70	20,43	20,32	22,03	20,98	13,27			
17-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90		19,02	
	17,99	14,93	15,92	19,48	19,38	30,86	21,10	12,51			
18-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCFO	20,27	
	35,92	16,02	41,38	13,70	13,69	19,89	17,79	14,38	9,65		
19-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90		23,11	
	48,96	16,78	46,01	13,68	13,51	19,62	10,33	15,98			
20-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90		21,01	
	43,96	15,68	34,01	16,75	16,54	17,53	8,09	15,47			
21-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCDO	14,11	
	14,67	15,77	15,37	15,39	15,18	13,45	8,87	17,73	10,56		
24-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90		23,36	
	42,91	27,09	12,35	24,83	24,77	11,95	14,13	28,86			
25-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCDO	21,01	
	32,35	25,09	26,11	22,41	22,30	9,47	13,14	28,07	10,15		
26-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	TLC90	TLCDO	18,61	
	31,60	18,77	25,22	18,87	18,66	10,63	12,98	21,01	9,73		
27-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PN7CO	PNICO	PNJCO	TLC90	TLCDO	17,36
	29,79	19,68	24,06	18,65	18,42	9,77	15,24	9,33	18,79	9,85	
28-Apr-23	MAC10	MAC20	MAC30	MAC40	PN5CO	PNICO	TLC90	TLCDO		21,41	
	38,56	21,02	31,00	18,64	18,41	14,47	18,24	10,91			
2-May-23	MAC10	MAC30	MAC40	PN5CO	PNICO	TLC90	TLCDO			22,34	
	34,97	30,91	17,54	17,55	17,15	21,53	16,71				
3-May-23	MAC10	MAC30	MAC40	PN5CO	PNICO	TLC90	TLCDO			27,50	
	35,54	32,62	22,20	22,23	26,81	30,98	22,08				
4-May-23	MAC10	MAC30	MAC40	PN5CO	PNICO	TLC90	TLCDO	TLCFO		27,46	
	45,81	41,95	23,77	23,58	43,77	18,76	29,75	12,88	6,88		
5-May-23	MAC10	MAC30	PN5CO	PNICO	TLC90	TLCDO				32,07	
	44,56	33,55	34,07	22,65	42,96	14,61					
8-May-23	MAC10	MAC30	PN5CO	PNICO	TLC90	TLCDO				29,39	
	48,21	31,65	26,36	23,49	30,63	16,01					
9-May-23	MAC10	MAC30	PN5CO	PN7CO	TLC90	TLCDO				28,68	
	49,64	32,62	25,88	9,66	32,19	22,11					
10-May-23	MAC10	MAC30	PN5CO	PN7CO	PNICO	TLC90	TLCDO	TLCFO		26,05	
	42,89	31,70	30,28	9,22	21,71	26,21	29,08	17,30			
11-May-23	MAC10	MAC30	PN5CO	PN7CO	PNICO	TLC90	TLCDO	TLCFO		21,14	
	38,39	25,25	19,08	9,76	14,84	15,40	27,25	19,14			
12-May-23	MAC20	MAC30	PN5CO	PN7CO	TLC90	TLCDO				22,56	
	25,98	28,28	18,08	10,96	22,66	29,42					
15-May-23	MAC10	MAC20	MAC30	PN5CO	PN7CO	TLC90	TLCDO	TLCFO		20,44	
	38,16	15,95	18,46	12,50	12,00	17,69	29,82	18,92			
16-May-23	MAC20	MAC30	PN5CO	PNICO	TLC90	TLCDO				17,65	
	12,78	17,37	9,84	22,86	14,19	28,84					
17-May-23	MAC20	MAC30	PN5CO	PNICO	TLC90	TLCDO				18,35	
	13,29	21,14	12,88	21,43	13,57	27,78					
18-May-23	MAC20	MAC30	PN5CO	PN9CO	PNICO	TLC90	TLCDO	TLCFO		18,48	

	15,05	20,35	13,28	16,62	16,27	12,96	33,53	19,82		
19-May-23	MAC20	MAC30	MAC40	PN5CO	PNECO	PNICO	TLC90	TLCDO	TLCFO	18,98
	17,90	23,92	14,17	14,23	20,08	17,77	17,78	26,37	18,59	
22-May-23	MAC20	MAC30	PN5CO	PNICO	TLC90	TLCDO				18,01
	15,45	23,06	13,03	18,69	15,74	22,10				
23-May-23	MAC20	MAC30	PN5CO	PNICO	TLC90	TLCDO				17,63
	17,58	14,65	18,64	20,71	17,33	16,88				
24-May-23	MAC20	MAC30	PN5CO	PNECO	PNICO	TLC90	TLCDO	TLCFO		21,88
	22,15	16,06	26,35	23,18	24,09	22,33	13,78	27,09		
29-May-23	MAC20	MAC30	PN5CO	PNICO	TLC90	TLCDO				17,57
	17,26	13,07	17,31	27,21	15,68	14,91				
30-May-23	MAC20	MAC30	MAC40	PN5CO	PNICO	TLC90	TLCDO			14,75
	9,06	21,08	9,64	9,60	30,22	8,85	14,78			
31-May-23	MAC20	MAC30	MAC40	PN5CO	PNICO	TLC90	TLCDO			15,68
	13,98	21,37	10,09	10,14	26,74	14,31	13,13			
1-Jun-23	MAC20	MAC30	MAC40	PN5CO	PNICO	TLC90	TLCDO			19,71
	16,92	33,65	17,30	17,29	27,72	16,49	8,62			

Tabla 6. Valores de condición de Hartigan en *clustering* de primera condición.

En la Tabla 6 se muestran el valor condicional de Hartigan que se calculó para cada vez que se realizó el *clustering* de la condición de filtrado número 2, podemos ver que el promedio diario de dichos valores en todos los casos es mayor de 10, como sugiere Hartigan en su teoría. En base a este análisis determinamos que $n=20$ es una cantidad acertada de observaciones para realizar el algoritmo *k-means*.

Asimismo, para poder analizar la concordancia de la cantidad de datos que presentes en el segundo proceso de *clustering* con la condición de Hartigan, calculamos el valor de la condición mencionada para cada día del período seleccionado y finalmente calculamos el promedio, dichos datos se exponen en la Tabla 7.

Fecha	Condición de Hartigan en segunda condición
27-Mar-23	21,99
28-Mar-23	58,02
29-Mar-23	31,42
30-Mar-23	68,70
31-Mar-23	16,49
3-Apr-23	89,58
4-Apr-23	48,99
5-Apr-23	21,52
10-Apr-23	21,48
11-Apr-23	28,75

12-Apr-23	13,97
13-Apr-23	47,29
14-Apr-23	5,31
17-Apr-23	13,96
18-Apr-23	17,82
19-Apr-23	21,81
20-Apr-23	18,49
21-Apr-23	12,97
24-Apr-23	13,50
25-Apr-23	12,01
26-Apr-23	9,24
27-Apr-23	9,48
28-Apr-23	15,26
2-May-23	13,34
3-May-23	18,89
4-May-23	62,94
5-May-23	19,57
8-May-23	4,52
9-May-23	10,38
10-May-23	27,26
11-May-23	30,58
12-May-23	7,22
15-May-23	14,07
16-May-23	5,42
17-May-23	6,67
18-May-23	22,41
19-May-23	23,07
22-May-23	3,42
23-May-23	3,07
24-May-23	5,95
29-May-23	7,24
30-May-23	12,22
31-May-23	20,71
1-Jun-23	29,40
Promedio	21,89

Tabla 7. Valores de condición de Hartigan en *clustering* de segunda condición

Vemos que en promedio los valores de la condición se posicionan en 21,89 siendo este valor considerablemente mayor al 10 que propone Hartigan.

9.2 Apéndice 2. P-valores del test de medias

A continuación, en la Tabla 8 se muestran los p-valores luego de realizar el test estadístico de medias *t* de Welch (1947).

Test de medias errores modelo-mercado				
	MAE	MAPE	RMSE	RMSPE
Sin filtro contra filtro de 2 condiciones	0,9624	0,8870	0,7215	0,6554
Sin filtro contra Filtro de solo condición 1	0,0020	0,0010	0,0027	0,0019
Sin filtro contra Filtro de solo condición 2	0,9090	0,9763	0,6644	0,6554

Tabla 8. P-valores de test t de Welch

En nuestro análisis utilizamos 5% como nivel de significación. Así, es relevante destacar que el único *test* de significación que arrojó resultados significativos fue el asociado al filtro con solo la condición 1. Este hallazgo es lógico, ya que la media calculada en esta configuración resultó notablemente superior a las demás, debido a que dicho filtro presenta un ajuste deficiente, como se detalla en el Apéndice 3. Por otro lado, el *test* para las otras condiciones concluimos que no hay evidencia estadística suficiente para determinar que en promedio las medias de esas series son diferentes.

Universidad de

9.3 Apéndice 3. Falencia del filtro de solo condición 1.

Como se mencionó previamente, se llevó a cabo un proceso de filtrado donde se aplicó únicamente la primera condición. Al examinar los resultados de esta serie de filtrados, se observó un ajuste notablemente deficiente. Es de suma importancia analizar las razones detrás de este resultado, ya que ello contribuirá a una mejor comprensión del funcionamiento de esta condición. A continuación, los Gráficos 9 y 10 presentan el resultado del proceso de *clustering* aplicado a la primera condición en relación a los *spreads* de dos activos para el día 23 de mayo, que representa casi el final del período de

análisis:

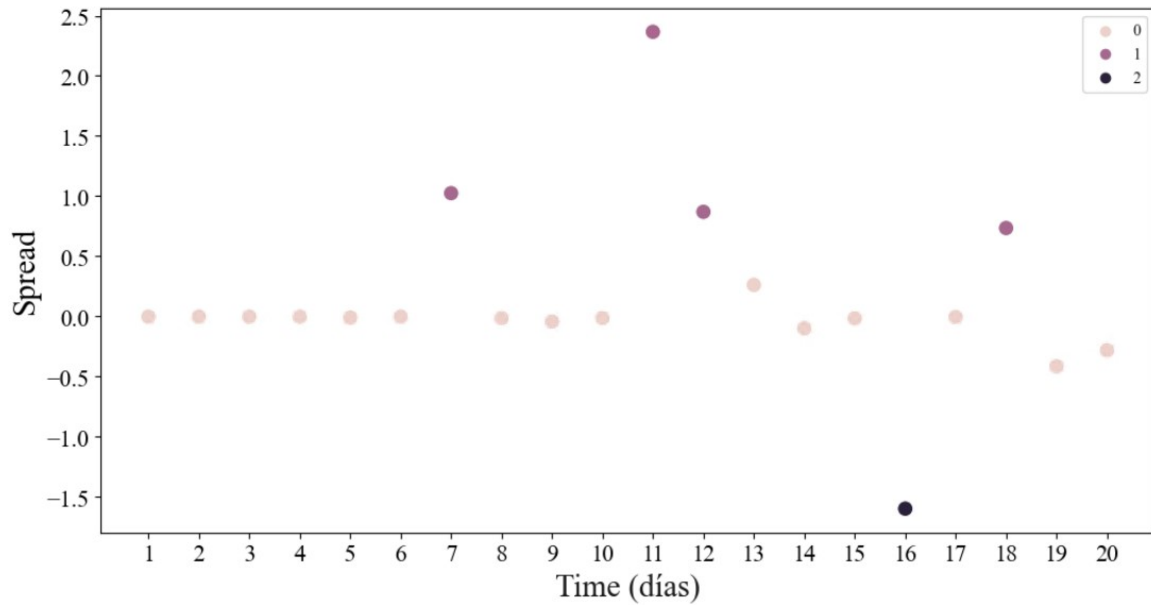


Gráfico 9. Agrupación de *spreads* del activo TLCDO por *clustering* para el 23 de mayo de 2023.

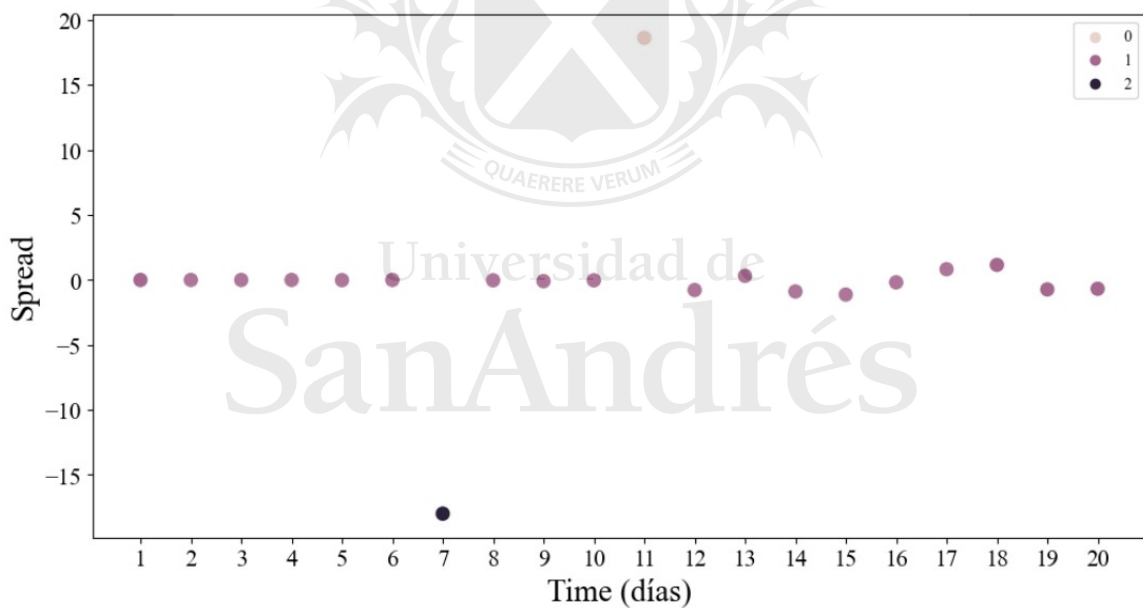


Gráfico 10. Agrupación de *spreads* del activo MAC2O por *clustering* para el 23 de mayo de 2023.

En los gráficos anteriores, se puede apreciar que después de haber aplicado el proceso de filtrado durante un período considerable de tiempo (dado que al 23 de mayo ya se ha acumulado un considerable historial de días filtrados), se evidencia que el filtro se torna altamente restrictivo al aplicar su condición. Con el transcurso del tiempo, este filtro comienza a catalogar como anómalas todas las observaciones que se desvían significativamente de cero, llegando incluso a posicionar los *spreads* claramente en el origen, como se ilustra en los gráficos. Esta tendencia a identificar como anómalos

aquellos movimientos de tasas alejados de cero genera ajustes altamente deficientes para todas las entradas de precios. Por lo tanto, se concluye que utilizar únicamente esta condición como filtro no resulta apropiado, ya que a medida que se extiende su aplicación, tiende a eliminar información relevante y a producir distorsiones en los datos. También es necesario aclarar que el último período en el que esta condición no filtró datos se debió a que, durante ese lapso, no se observaron movimientos significativos alejados de las tendencias históricas.

9.4 Apéndice 4. P-valores del test de Levene

A continuación, en la Tabla 9 se muestran los p-valores resultantes de aplicar el *test* estadístico de Levene (1960) entre la serie de parámetros de las calibraciones sin filtro y con cada filtro.

Test de Levene sobre serie de parámetros			
	Beta 0	Beta 1	Beta 2
Sin filtro contra filtro de 2 condiciones	0,4867	0,3158	0,3856
Sin filtro contra Filtro de solo condición 1	0,1409	0,1382	0,0299
Sin filtro contra Filtro de solo condición 2	0,0823	0,0539	0,2732

Tabla 9. P-valores de test de Levene

En nuestro análisis utilizamos 5% como nivel de significación. Así, es relevante destacar que el único *test* de significación que arrojó resultados significativos para demostrar la varianza diferente contra el caso sin filtro fue el caso del beta 2 con el filtro de solo condición 1. Sin embargo, al ampliar el nivel de significación al 10%, se suman a la lista de resultados significativos los coeficientes beta 0 y beta 1 en el caso del filtro que incorpora solo la condición 2.